

**Конспект лекций по дисциплине
«Методы интеллектуальной обработки информации»**

Оглавление

Лекция 1 «Основные задачи теории распознавания образов»	2
Лекция 2 «Общая постановка проблемы распознавания объектов и явлений»	14
Лекция 3 «Классификация систем распознавания».....	23
Лекция 4 «Качественное описание задачи распознавания»	31
Лекция 5 «Классификация с помощью решающих функций»	40
Лекция 6 «Нейронные сети и проблемы распознавания».....	61
Лекция 7 «Нейронная сеть в распознавании образов».....	75
Лекция 8 «Статистический подход в теории распознавания образов».....	82

Введение

Распознавание образов – это наука о методах и алгоритмах классификации объектов различной природы.

Можно выделить следующие наиболее важные направления развития интеллектуальных систем (т.е. систем, решающих задачи, традиционно относимые к интеллектуальной сфере), в которых широко используются методы распознавания образов:

- распознавание символов (печатного и рукописного текстов, банковских чеков и денежных купюр и т.д.);
- распознавание изображений, полученных в различных частотных диапазонах (оптическом, инфракрасном, радиочастотном, звуковом) и анализ сцен;
- распознавание речи;
- медицинская диагностика;
- системы безопасности;
- классификация, кластеризация и поиск в базах данных и знаний (в том числе и в Интернет-ресурсах)¹.

Исторически сложилось так, что теория распознавания образов развивалась по двум направлениям: детерминистскому и статистическому, хотя чаще всего строго различить их не удастся. Детерминистский подход включает различные методы: эмпирические, эвристические, в основе которых лежат здравый смысл, более или менее удачное моделирование действий, осуществляемых мозгом человека; математически формализованные, например, основанные на модели порождения объектов (реализаций) того или иного образа. При этом используется различный математический аппарат (математическая логика, теория графов, топология, математическая лингвистика, математическое программирование и др.).

Статистический подход опирается на фундаментальные результаты математической статистики (теория оценок, последовательный анализ, стохастическая аппроксимация, теория информации).

¹ Лепский А.Е., Броневиц А.Г. Математические методы распознавания образов: Курс лекций. – Таганрог: Изд-во ТТИ ЮФУ, 2009. – 155 с.

Многие методы распознавания, появившиеся как детерминистские, получили в дальнейшем статистическое обоснование. Примеры подобного рода рассматриваются в предлагаемом курсе лекций.

В процессе развития теории распознавания различные подходы и применяемый математический аппарат переплелись столь причудливым образом, что классификация различных алгоритмов по используемым методам является условной и неоднозначной. Тем не менее, в данном курсе выделены два раздела: детерминистские методы и статистические методы. Это сделано в основном из педагогических соображений. Детерминистские методы (особенно эмпирические) достаточно наглядны, легче воспринимаются, чем статистические, поэтому методически целесообразно начинать изложение материала с них².

Лекция 1 «Основные задачи теории распознавания образов»

Распознавание образов (объектов, сигналов, ситуаций, явлений или процессов) — едва ли не самая распространенная задача, которую человеку приходится решать практически ежесекундно от первого до последнего дня своего существования. Для решения этой задачи человек использует огромные ресурсы своего мозга, включая одновременно около 10 — 12 млрд нейронов. Именно это дает возможность людям мгновенно узнавать друг друга, с большой скоростью читать печатные и рукописные тексты, безошибочно водить автомобили в сложном потоке уличного движения, осуществлять отбраковку деталей на конвейере, дешифровать аэро- и космические фотоснимки, разгадывать коды, древнюю египетскую клинопись и т. д.

Распознавание представляет собой задачу преобразования входной информации, в качестве которой уместно рассматривать некоторые параметры, признаки распознаваемых образов, в выходную, представляющую собой заключение о том, к какому классу относится распознаваемый образ. Поэтому, учитывая, что кибернетика есть наука об общих законах преобразования информации в сложных системах, распознавание образов есть один из разделов этой науки.

На вход системы распознавания поступает некоторый символ x , написанный на бумажной ленте и который необходимо распознать. Объекты, подлежащие распознаванию, называют образами («pattern»). Техническая система имеет считывающую головку, которая, передвигаясь слева направо, измеряет скорость $x(t)$

² Волошин Г.Я. <http://www.allbest.ru/>

изменения площади зачерненной поверхности в зависимости от времени t (рис. 1.1). Говорят, что функция $x(t)$ является представлением образа-символа x . Можно измерить сигнал в дискретные моменты времени – получим другое представление символа x в виде некоторого вектора X



Рис. 1.1

Будем считать, что входной образ-символ x может принадлежать некоторому классу из множества всех классов $\Omega = \{\omega_1, \dots, \omega_m\}$ – каждый класс соответствует некоторому символу (букве, цифре и т.д.). Предполагается, что классы являются непересекающимися.



Рис. 1.2

Задача распознавания образов состоит в соотнесении исходного образа x одному из классов ω_i . Правила соотнесения образа одному из классов называются классификатором и реализуются в блоке классификации. Если образам соответствуют векторы – элементы метрического пространства, то соотнесение образа классу можно осуществить, например, с помощью вычисления расстояния между вектором и классом. На выходе классификатора должны получить тот класс (номер класса), которому принадлежит входной образ с указанием степени достоверности классификации или получить информацию о том, что входной образ не принадлежит ни одному из классов.

Основные задачи теории распознавания образов:

1. Математическое описание образов.

Наиболее удобным математическим описанием считается векторное описание образов. В этом случае каждому образу x ставится в соответствие некоторый вектор

$\mathbf{x} = (x_1, x_2, \dots)^T$ признаков x_i этого образа – элемент векторного пространства X . Такое векторное пространство называется пространством признаков.

2. Выбор наиболее информативных признаков, описывающих данный образ.

Это одна из основных и важных задач в теории распознавания образов – найти минимальное количество признаков, наиболее информативно описывающих образы в данной системе (или задаче) распознавания. Полный набор выбранных для распознавания признаков называют алфавитом признаков. Минимальный же набор признаков, достаточный для решения данного класса задач распознавания, называют словарем признаков.

3. Описание классов распознаваемых образов.

Эта задача сводится к определению границ классов. Границы классов могут быть заданы явно на этапе разработки системы распознавания или система сама должна их найти в процессе своей работы.

4. **Нахождение оптимальных решающих процедур** (методов классификации), т.е. методов соотнесения вектора признаков образа некоторому классу.

5. **Оценка достоверности классификации образов.** Это необходимо для оценки величины потерь, связанных с неправильной классификацией

Типы характеристик образов:

1. **Физические** характеристики (показания, снятые с различных датчиков)

Физические характеристики могут быть детерминированными и вероятностными. Описываются с помощью векторов и обрабатываются как элементы векторного пространства.

2. **Качественные характеристики** (понятия «темный», «светлый», «высокий» и т.д.) Такие характеристики могут быть описаны с помощью, так называемых лингвистических переменных, методами теории нечетких множеств.

3. **Структурные характеристики.** Эти характеристики употребительны для описания изображений сложных объектов (рис. 1.3) или сцен. При описании некоторый формальный язык (например, теория графов (рис. 1.4)).

4. **Логические характеристики** – это высказывания, о которых имеет смысл говорить истины они или ложны.

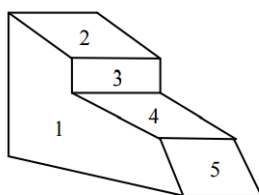


Рис. 1.3

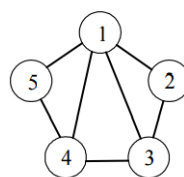


Рис. 1.4

Типы систем распознавания

Зависят от критериев классификации. Существует несколько критериев классификации систем распознавания:

1. по характеру информации о признаках:

- детерминистские;
- вероятностные;
- логические;
- структурные;
- комбинированные.

2. по количеству априорной информации о распознаваемых объектах.

Три основных типа систем распознавания:

1. **Системы без обучения.** Количество априорной информации достаточно для определения алфавита признаков (полного набора признаков), формирования словаря признаков (т.е. определения минимального набора признаков, достаточного для решения задач распознавания) и определения границ классов. Отсутствует блок «обучение».

2. **Системы, основанные на обучении с учителем.** Количества априорной информации достаточно только для выбора алфавита признаков и формирования словаря признаков, но не для определения границ между классами. Системе распознавания предъявляется некоторое множество объектов, которое называется обучающим множеством (обучающей выборкой), с указанием, к каким классам эти объекты принадлежат. Система сама должна настроить параметры правил классификации таким образом, чтобы выполнялось условие минимальности ошибки неправильной классификации.

3. **Системы, основанные на самообучении (на объяснении).** Количества априорной информации недостаточно даже для формирования словаря признаков. В этом случае в систему распознавания образов вводится список правил, объясняющий задачи распознавания образов. Этот список правил вырабатывается, как правило, экспертами – специалистами в данной области знаний, такие системы называют экспертными (интеллектуальными). Система распознавания, исходя из этого набора правил, должна сама сформировать словарь признаков и определить границы классов. При этом, как правило, используются логико-лингвистические методы обработки данных. В такой системе процесс поиска классификационного решения называют логическим выводом или выводом на знаниях.

Системы, основанные на обучении с учителем (рис. 1.5). Системы, основанные на самообучении (на объяснении) - система распознавания для кардиологов (рис. 1.6, 1.7).

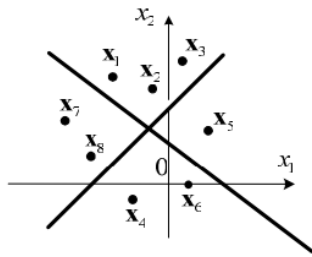


Рис. 1.5

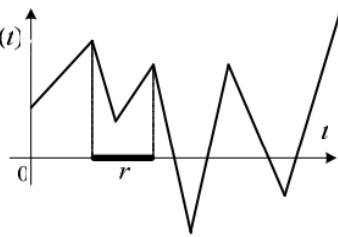


Рис. 1.6

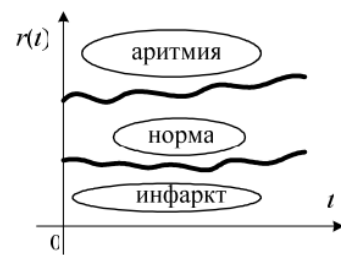
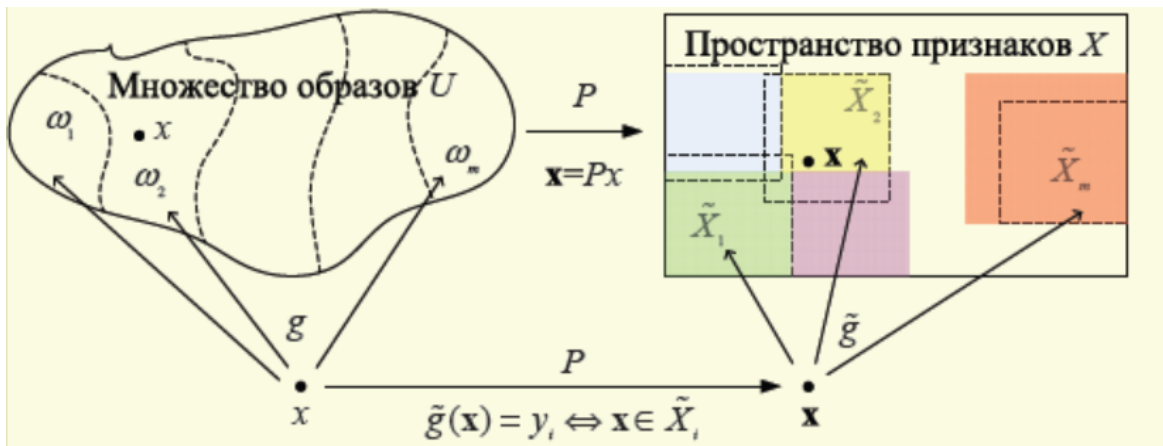


Рис. 1.7

Математическая постановка задачи распознавания



Задача: найти решающую функцию $y = \tilde{g}(x)$ по множеству прецедентов $(\Xi, Y) = \{(x_i, y_i) : i = 1, \dots, N\}$.

«Интеллектуальные методы распознавания информации»

Современные информационные системы и технологии включают в себя большое количество процедур, моделирующих или поддерживающих процесс интеллектуального анализа данных.

Исследованием и разработкой методов, алгоритмов и систем для решения таких задач на ЭВМ занимается дисциплина, получившая название распознавание образов.

В рамках курса рассматриваются основные концепции теории распознавания образов, а также методы и алгоритмы распознавания, применяемые при классификации многомерных количественных данных.

Задачи распознавания относятся к проблематике искусственного интеллекта.

Понятие искусственного интеллекта и интеллектуальной задачи.

Искусственный интеллект (ИИ) (Artificial intelligence, AI) — это научное направление, в рамках которого ставятся и решаются задачи моделирования тех видов человеческой деятельности, которые традиционно считаются интеллектуальными (представление знаний, обучение, общение и т.п.)

Термин «искусственный интеллект» ввел в 1956 Джон Маккарти (р.1927, изобретатель языка логического программирования Лисп).

Особенности интеллектуальных задач

- для интеллектуальной задачи не существует формального алгоритма решения;
- под интеллектуальной задачей подразумевается целый класс задач (н/р, класс геометрических задач, игровых задач и т.д.);
- при решении интеллектуальных задач необходимо использовать некоторую базу знаний.

Примеры:

- 1) «решатели» задач;
- 2) игровые программы.

Критерии интеллектуальности

Тест Тьюринга (статья «Computing Machinery and Intelligence», журнал «Mind», 1950).

Судья (человек) переписывается на естественном языке с двумя собеседниками, один из которых – человек, другой – компьютер. Если судья не может надёжно определить, кто есть кто, считается, что компьютер прошёл тест.

Тьюринг считал, что к 2000 году компьютер с памятью 120 МБ в ходе 5-минутного теста сможет обмануть судей в 30 % случаев.

По состоянию на 2008 год ни одна из существующих компьютерных систем не прошла тест. Ежегодно производится соревнование между разговаривающими программами. В 2008 году победителю соревнования Elbot удалось обмануть троих из двенадцати судей.

«Физическая символическая система имеет необходимые и достаточные средства для произведения базовых интеллектуальных действий, в широком смысле»

Подходы к разработке ИИ

- нисходящий, семиотический — создание символьных систем, имитирующие высокоуровневые психические процессы: мышление, рассуждение, речь, эмоции, творчество и т. д.;

- восходящий, биологический — изучение нейронных сетей и эволюционных вычислений, моделирующих интеллектуальное поведение на основе более мелких «неинтеллектуальных» элементов.

Направления исследований

- моделирование рассуждений (доказательство теорем, принятие решений, теория игр, планирование и диспетчеризация, прогнозирование);

- обработка естественного языка;

- инженерия знаний (машинное обучение, экспертные системы);

- моделирование биологических систем (нейронные сети, генетические алгоритмы);

- распознавание образов;

- интеллектуальные роботы;

- машинное творчество.

Области применения распознавания образов

- распознавание символов (печатного и рукописного текстов, банковских чеков и денежных купюр и т.д.);

- распознавание изображений, полученных в различных частотных диапазонах (оптическом, инфракрасном, радиочастотном, звуковом) и анализ сцен;

- распознавание речи;

- медицинская диагностика;

- системы безопасности;

- классификация, кластеризация и поиск в базах данных и знаний (в том числе и в Интернет-ресурсах).

Основные понятия теории распознавания образов.

Образ - это описание объекта или процесса, позволяющее выделять его из окружающей среды и группировать с другими объектами или процессами для принятия необходимых решений.

Для системы обработки информации образ - это совокупность данных об объекте или явлении, включающая параметры и связи

Целью процедуры распознавания является ответ на вопрос: относится ли объект, описанный заданными характеристиками, к интересующим нас категориям и если относится, то к какой именно.

- Классы – это категории объектов, которые должны быть выделены или на которые необходимо разделить все множество образов в процессе распознавания
- При распознавании изображений существенное влияние на точность оказывают:
- Масштаб. Изображения имеют разный масштаб. Предметы, которые мы воспринимаем как одинаковые, на самом деле занимают разную площадь на разных изображениях.
- Место. Интересующий нас объект может находиться в разных местах изображения.
- Фон и помехи. Предмет, который мы воспринимаем как что-то отдельное, на изображении никак не выделен, и находится на фоне других предметов. Кроме того, изображение не идеально и может быть подвержено всякого рода искажениям и помехам.
- Проекция, вращение и угол обзора. Изображение является лишь двумерной проекцией нашего трехмерного мира. Поэтому поворот объекта и изменение угла обзора кардинальным образом влияют на его двумерную проекцию — изображение. Один и тот же объект может давать совершенно разную картинку, в зависимости от поворота или расстояния до него.
- Анализ образа - отсутствует качественное продвижение в решении таких задач, как *анализ* 3-мерных сцен и перевод с одного языка на другой.

При сравнении зрительного образа с эталоном в зависимости от эталона методы можно разделить на:

- 1) Растровые – в виде матрицы точек, каждая точка обладает яркостью разной силы $(0,1)$. Важно чтобы входное изображение и эталон были приведены к одинаковым размерам и т.д.

Недостатки:

- Необходимо выделять объекты из общего изображения
- Чувствительны к размерам

- 2) Признаковый. На изображении выделяются признаки: размер, цвет, объем и т.д. Выделяется множество признаков, которое можно представить как вектор в пространстве

Недостатки:

- Необходимо выделять объекты из образа
- Выделять признаки априори, что чревато потерей информации из за неполного (не адекватного) представления объекта.
- Кластеризация - разбиение множества объектов на прикластеры по признакам.

3) Структурный метод. На образе выделяются структурные элементы отрезки, дуги, точки. Структурные элементы и распознаются.

Недостатки

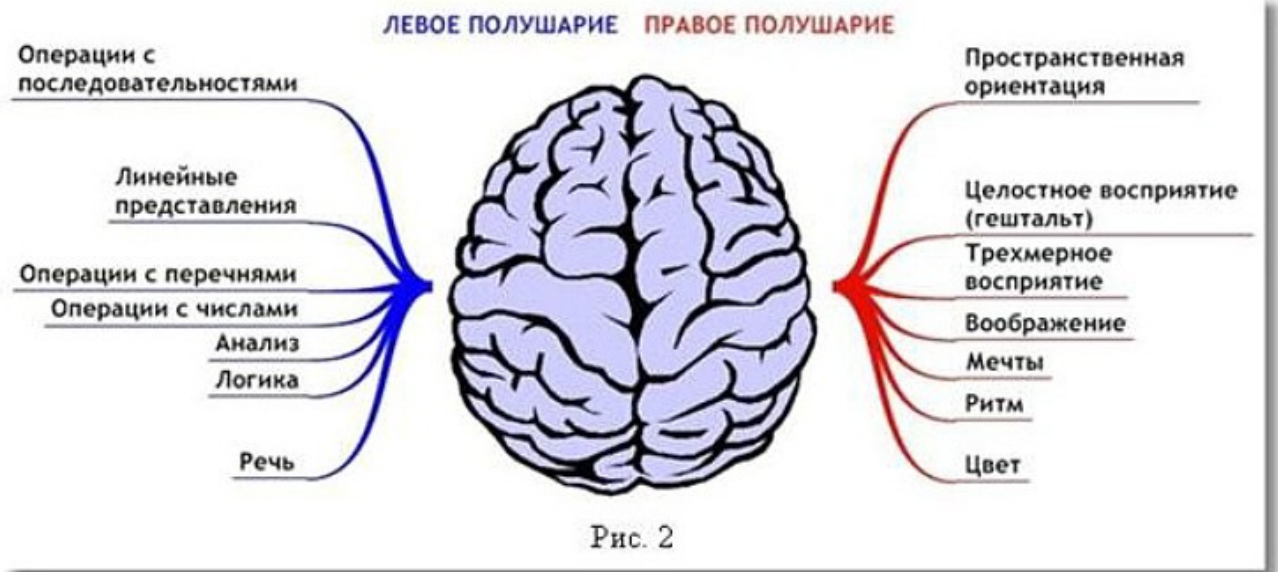
- Необходимость сегментации
- Априорное выделение признаков
- Реальное восприятие (человеческое мышление)
- Использует целостное восприятие среды (картины).
- В мозгу строится целостная модель образа, работают и признаковое и структурное восприятие образа.

Формирование образа у человека.

Сведения об объекте поступают по конечному числу сенсоров (анализируемых каналов) в мозг, и каждому сенсору можно сопоставить соответствующую характеристику объекта.

Помимо признаков, соответствующих нашим измерениям объекта, существует так же выделенный признак, либо группа признаков, которые мы называем классифицирующими признаками, и в выяснении их значений при заданном векторе X и состоит задача, которую выполняют естественные и искусственные распознающие системы.





Выбор исходного описания объекта: геометрический или структурный подход.

Геометрический основывается на различии формы объектов.

Структурный подход основан на анализе элементов объекта.

При этом сначала выделяют фрагменты всех объектов и характеристики их взаимного расположения. Все это образует исходную БД - словарь, позволяющий строить различные гипотезы. Из них отбираются наиболее существенные для данного конкретного случая, и затем выполняется описание единого образа объектов, который используется для идентификации других объектов.

Вывод

Сложная задача распознавания образов требует разработки гибкой распознающей системы, которая могла бы классифицировать любой образ, имеющийся на изображении

В целом проблема распознавания образов состоит из двух частей:

- обучение (система должна приобрести способность реагировать одинаковыми реакциями на все объекты одного образа)
- распознавание (распознавание новых объектов)

Основные принципы распознающих систем

-заложить в компьютер как можно больше известных образов-шаблонов и сравнивать их с поступающими для распознавания неизвестными образами

-на первой стадии обязательно обрабатывают изображение и выделяют характерные признаки

-обучение

Лекция 2 «Общая постановка проблемы распознавания объектов и явлений»

Общие положения задач распознавания

Во-первых, решение задач распознавания требует в общем случае построения специальной системы распознавания.

Во-вторых, решение задачи распознавания необходимо (также в общем случае) для того, чтобы система управления, стоящая над системой распознавания, могла принимать правильные решения.

Системы распознавания должны строиться так, чтобы обеспечивать системе управления возможность наиболее эффективно распоряжаться своими ресурсами, допустимым набором решений, а само построение систем распознавания, как и любых технических систем, не может быть осуществлено без учета соответствующих ограничений.

Содержательная трактовка проблемы распознавания

Процесс распознавания состоит в том, что система распознавания на основании сопоставления апостериорной информации относительно каждого поступившего на вход системы объекта или явления с априорным описанием классов принимает решение о принадлежности этого объекта (явления) к одному из классов.

Правило, которое каждому объекту ставит в соответствие определенное наименование класса, называют **решающим правилом**.

Суть проблемы распознавания заключается в определении решающих правил, нахождении в признаковом пространстве таких границ (решающих границ), придерживаясь которых признаки пространства оптимальным образом, например с точки зрения **минимизации ошибок распознавания**, подразделяются на области, соответствующие классам.

При определении решающих правил (решающих границ в признаковом пространстве) в зависимости от объема исходной априорной информации рассматриваются следующие ситуации:

1. Количество исходной информации достаточно для того, чтобы путем ее анализа и непосредственной обработки определить решающие правила (**системы распознавания без обучения**)

2. Количество исходной информации недостаточно для определения решающих правил на основе ее непосредственной обработки, в связи с чем реализуется процедура обучения (**обучающиеся системы распознавания**).

Задача отыскания решающих правил базируется на том, что **алфавит классов объектов и априорный словарь признаков, предназначенных для их описаний, известны.**

Ситуация, когда **словарь признаков известен, но неизвестен алфавит классов.** При этом, **определен некоторый набор правил**, в соответствии с которыми на основании процедуры самообучения находится искомый алфавит классов. Затем определяются решающие правила (**самообучающиеся системы**).

Назначение систем распознавания — получить информацию, необходимую для принятия определенных решений, о принадлежности неизвестного объекта (явления) к тому или иному классу.

Ограничения на построение систем распознавания.

1. При прочих равных условиях повышение эффективности принимаемых решений следует связывать со **степенью детализации** определения или назначения либо характера распознаваемого объекта или явления. **Степень детализации** определяется количеством классов, на которое подразделено множество объектов или явлений. Если система управления располагает m различными решениями, то в алфавите классов системы распознавания, целесообразно предусмотреть $m+1$ классов. Тогда, если распознанный объект относится к классу Ω_1 принимается решение I_1 , если к классу Ω_2 — решение h и т. д., если же объект относится к классу Ω_{m+1} , решение не принимается.

2. Эффективность принимаемых системой управления решений (в том числе, при заданном алфавите классов) зависит от **точности определения принадлежности** распознаваемого объекта или явления к соответствующему классу. Точность же определения или ошибка распознавания при заданном по точности априорном описании классов определяется **размерностью и информативностью признакового пространства, объемом и качеством апостериорной информации о значениях признаков (параметров), которыми характеризуется распознаваемый объект.**

Пусть заданы три класса Ω_1 , Ω_2 и Ω_3 объектов распределениями $f_1(x)$, $f_2(x)$, $f_3(x)$ априорными вероятностями появления объектов соответствующих классов $P(\Omega_1)=P(\Omega_2)=P(\Omega_3)=P$, а также потерями $c_{11} = c_{22} = c_{33} = 0$ и $c_{12} = c_{21} = c_{13} = c_{31} = c_{23} = c_{32} = c$.

Представлены законы распределений. Средний (байесовский) риск:

$$\begin{aligned} \bar{R}_1 &= Pc \left[\int_{-\infty}^a f_2(x) dx + \int_{-\infty}^a f_3(x) dx + \int_a^b f_1(x) dx + \int_a^b f_3(x) dx + \right. \\ &\quad \left. + \int_b^{\infty} f_1(x) dx + \int_b^{\infty} f_2(x) dx \right] = \\ &= Pc \left[\int_{-\infty}^a f_2(x) dx + \int_{-\infty}^b f_3(x) dx + \int_a^{\infty} f_1(x) dx + \int_b^{\infty} f_2(x) dx \right]. \end{aligned}$$

Положим теперь, что объекты, относящиеся к классам Ω_1 и Ω_2 , решено объединить в один класс Ω_4 , описание которого:

$$f_4(x) = \frac{P(\Omega_1)f_1(x) + P(\Omega_2)f_2(x)}{P(\Omega_1) + P(\Omega_2)} = \frac{f_1(x) + f_2(x)}{2};$$

$$P(\Omega_4) = 2P, \quad C_{43} = C_{34} = C.$$

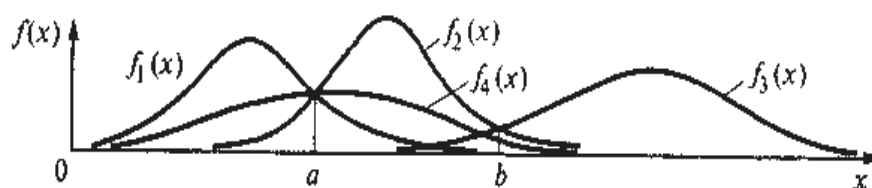
Средний риск в данном случае в предположении неизменности границы b составит:

$$\begin{aligned} \bar{R}_2 &= c \left[P \int_{-\infty}^b f_3(x) dx + 2P \int_b^{\infty} f_4(x) dx = \right. \\ &= Pc \left[\int_{-\infty}^b f_3(x) dx + \int_b^{\infty} f_1(x) dx + \int_b^{\infty} f_2(x) dx \right]. \end{aligned}$$

Из сравнения величин \bar{R}_1 и \bar{R}_2 видно, что $\bar{R}_1 > \bar{R}_2$ на величину:

$$\Delta R = \bar{R}_1 - \bar{R}_2 = Pc \left[\int_{-\infty}^a f_2(x) dx + \int_a^b f_1(x) dx \right].$$

Следовательно, при заданном признаковом пространстве и прочих равных условиях уменьшение числа классов приводит к уменьшению ошибок распознавания и, наоборот, при увеличении числа классов системы распознавания в целях поддержания на заданном уровне или даже уменьшения среднего риска (вероятности ошибочных решений) надо расширять словарь признаков (естественно, при прочих равных условиях).



Таким образом, стремление по возможности наиболее эффективно использовать набор возможных решений системы управления приводит к необходимости увеличения алфавита классов до $m+1$.

По мере увеличения алфавита классов ошибки распознавания растут, а это уменьшает эффективность использования возможных решений.

Необходим компромисс между размерами алфавита классов и объемом рабочего словаря признаков системы, базирующийся на исходных данных относительно набора возможных решений и величины ресурсов, отпущенных на создание измерительной аппаратуры, реализующей словарь признаков, позволяет обеспечить решение задачи построения системы распознавания оптимальным образом.

Составляющая эффективности управлений, зависящая от функционирования системы распознавания, обуславливается двумя факторами.

- Первый фактор связан со степенью детализации распознавания объектов или явлений, наибольшее значение которой будет в том случае, если число классов, содержащихся в алфавите классов системы распознавания, равно количеству возможных решений (плюс единица — последний класс, объекты которого не распознаются).

- Второй фактор — точность решения задачи распознавания. Естественно, чем она выше, тем меньше вероятность принять решение, не соответствующее особенностям данного объекта или явления.

Цель построения системы распознавания - в разработке таких алфавита классов и словаря признаков, которые в условиях ограниченных ресурсов на построение системы распознавания обеспечивают максимальную эффективность системы управления, принимающей соответствующее решение в зависимости от результатов решения задачи распознавания. И находить наилучшие решающие правила, решающие границы между классами.

Важно, что при построении логических систем распознавания, использующих либо алгоритмы распознавания, основанные на методах алгебры логики, либо структурных (лингвистических) систем, решающие правила вообще не определяются.

Постановка задачи распознавания

Пусть задано множество объектов или явлений $\Omega = \{w_1, \dots, w_z\}$, а также множество возможных решений $L = \{l_1, \dots, l_k\}$, которые могут быть приняты системой управления по результатам решения задачи распознавания. Введем в рассмотрение множество возможных вариантов разбиения объектов на классы $A = \{A_1, \dots, A_r\}$. Будем полагать, что если выбран вариант разбиения $A_\alpha, \alpha=1, \dots, r$, то множество Ω подразделяется на m_α классов, т. е.

$$A_\alpha: \Omega_q^{A_\alpha} \cap \Omega_g^{A_\alpha} = \emptyset, \bigcup_{i=1}^{m_\alpha} \Omega_i^{A_\alpha} = \Omega, q, g = 1, \dots, m_\alpha, q \neq g.$$

Пусть первоначальная информация позволяет построить априорное признаковое пространство (составить априорный словарь признаков), описываемое многомерным вектором $x_\alpha = \{x_1, \dots, x_N\}$

Информация относительно множества решений $L = \{1_1, \dots, 1_k\}$ позволяет произвести исходное разбиение множества объектов на классы, т. е. составить априорный алфавит классов.

В первом варианте подразделения объектов на классы ($\alpha = 1$), т. е. когда $A_\alpha = A_1$ их число равно $m_\alpha = m_1 = k + 1$. Исходное множество объектов $\Omega = \{w_1, \dots, w_z\}$ (обучающую выборку) подразделим на подмножества — классы $\Omega_1^{A_1}, \Omega_2^{A_1}, \dots, \Omega_{m_1}^{A_1}$.

Пусть первоначальная информация позволяет построить априорное признаковое пространство (составить априорный словарь признаков), описываемое многомерным вектором $x_\alpha = \{x_1, \dots, x_N\}$. Информация относительно множества решений $L = \{1_1, \dots, 1_k\}$ позволяет произвести исходное разбиение множества объектов на классы, т. е. составить априорный алфавит классов. В первом варианте подразделения объектов на классы ($\alpha = 1$), т. е. когда $A_\alpha = A_1$, их число равно $m_\alpha = m_1 = k + 1$. Исходное множество объектов $\Omega = \{w_1, \dots, w_z\}$ (обучающую выборку) подразделим на подмножества — классы $\Omega_1^{A_1}, \Omega_2^{A_1}, \dots, \Omega_{m_1}^{A_1}$.

Если обучающая выборка достаточно представительна, то непосредственной обработкой исходной информации можно определить описания классов.

При статистическом подходе к задаче распознавания такими описаниями являются априорные вероятности $P(\Omega_1^{A_1})$ появления объектов соответствующих классов, а также условные плотности распределения значений признаков по классам, т. е. функции $f_{\Omega_i^{A_1}}^A(x_1, \dots, x_N), i = 1, \dots, m_1$.

Если объем исходной априорной информации недостаточен для непосредственного описания классов, то они могут быть получены с помощью процедуры обучения.

Наличие описаний классов позволяет определить решающие правила (решающие границы), использование которых обеспечивает минимизацию ошибок при распознавании неизвестных объектов.

$P_{\text{пр}}(\Omega_i^{A_1} | x_\alpha)$ - Оценка апостериорной вероятности правильного решения

Обозначим оценку апостериорной вероятности правильного решения задачи распознавания, усредненную по всем возможным значениям признаков априорного словаря, описываемого вектором x_d . Эта оценка может быть получена проведением статистических испытаний (метод Монте — Карло) математической модели системы распознавания.

Метод Монте-Карло (методы Монте-Карло, ММК) — общее название группы численных методов, основанных на получении большого числа реализаций стохастического (случайного) процесса, который формируется таким образом, чтобы его вероятностные характеристики совпадали с аналогичными величинами решаемой задачи

Если бы не было финансовых ограничений на величину ресурсов, на построение измерительных устройств, предназначенных для определения признаков x_1, x_2, \dots , то можно было бы полагать, что основные характеристики системы распознавания — алфавит классов и словарь признаков — определены, и можно приступить к построению системы распознавания.

В условиях ограничений, когда реализовать априорное признаковое пространство $x_{\alpha\alpha} = \{x_1, \dots, x_N\}$ в полном объеме не представляется возможным, приходится его сокращать по сравнению с априорным, т. е. переходить от априорного словаря признаков к **рабочему**.

Положим, что вектор $\lambda = \{\lambda_1, \dots, \lambda_N\}$, компоненты которого $\lambda_j = \begin{cases} 1 \\ 0 \end{cases}$ (в зависимости от того, используется ли данный признак априорного словаря в рабочем словаре или нет).

Введем обозначение для рабочего словаря $x_p = \{x_{j_1}, \dots, x_{j_n}\}$, где $j_1, \dots, j_n \in 1, \dots, N$, т. е. множество признаков рабочего словаря состоит из элементов множества признаков априорного словаря (**рабочий словарь представляет собой подмножество множества признаков априорного словаря**).

Обозначим C_j стоимость создания измерительного устройства, обеспечивающего определение x_j -го признака, $j=1, \dots, N$, а C_0 — общую величину ресурсов, ассигнованных на создание всех измерителей. Если :

$$\sum_{j=1}^N C_j = C_0,$$

то в качестве рабочего словаря системы распознавания может быть использован априорный словарь.

Если суммарная стоимость создания комплекса технических средств, обеспечивающих измерение всех признаков априорного словаря, превышает величину C_0

$$\sum_{j=1}^N C_j > C_0.$$

то затраты на создание комплекса технических средств системы, обеспечивающих измерение признаков рабочего словаря, определяются величинами

$$C = \sum_{j=1}^N \lambda_j C_j = \sum_{\beta=1}^n C_{j_\beta} \leq C_0.$$

Обозначим $G(\Omega^{A_1}_i)$ выигрыш, связанный с реализацией возможных решений при распознавании объекта w , отнесенного к классу Ω^A_i в варианте классификации A_1 .

Тогда математическое ожидание выигрыша от выбора варианта A_1 при использовании априорного словаря признаков

$$R = M[G(\Omega_i^{A_1})] = \sum_{i=1}^{m_1} P_{\text{пр}}(\Omega_i^{A_1} | \mathbf{x}_a) G(\Omega_i^{A_1}).$$

Величину R рассматривают в качестве критерия эффективности системы распознавания. И с его максимизацией связано увеличение эффективности функционирования системы.

В условиях ограничений, определяемых величиной C_0 , возникает следующая экстремальная задача: необходимо в пределах C_0 найти такой вариант разбиения объектов на классы и такое пространство признаков, при которых обеспечивается максимальное значение критерия эффективности системы R . Другими словами, необходимо определить $A_\alpha = A^0$ из множества $A = \{A^1, \dots, A_\alpha, \dots, A_r\}$ и вектор $\lambda = \lambda^0$, которые при наилучшем решающем правиле доставляют экстремальное (максимальное) значение величины R при соблюдении ограничений на величину $C \leq C_0$, т.е.

$$R = R(A^0, \lambda^0) = \max_{A_\alpha} \max_{\lambda} R = \max_{A_\alpha} \max_{\lambda} \sum_{i=1}^{m_\alpha} P_{\text{пр}}(\Omega_i^{A_\alpha} | \lambda) G(\Omega_i^{A_\alpha}),$$

$$C = \sum_{j=1}^N \lambda_j C_j \leq C_0.$$

При этом A^0 определяет алфавит классов, а λ^0 — оптимальный рабочий словарь признаков.

Вывод:

В условиях первоначального (априорного) описания исходного множества объектов на языке априорного словаря признаков необходимо в пределах выделенных ресурсов на построение измерительной аппаратуры определить оптимальный алфавит классов и оптимальный рабочий словарь признаков, которые при наилучшем решающем правиле обеспечивают наиболее эффективное использование решений, принимаемых по результатам распознавания неизвестных объектов или явлений системой управления.

Метод решения задачи распознавания

Наиболее экономный метод решения – это метод математического или физико-математического моделирования. Основная идея модели разрабатываемой системы распознавания — реализация итеративной процедуры, обеспечивающей путем последовательных приближений синтез системы, эффективность работы которой достаточно близко приближается к потенциально достижимой.

Для построения модели необходимы:

1. Множество возможных решений, которые могут быть приняты системой управления на основании результатов распознавания неизвестных объектов или явлений $L = \{l_1, \dots, l_k\}$.

2. Априорный словарь признаков $x_a = \{x_1, \dots, x_N\}$.

3. Исходное множество объектов $\Omega = \{w_1, \dots, w_z\}$.

4. Величина ресурсов C_0 , ассигнованных на построение измерительной аппаратуры системы.

5. Значения выигрышей, получаемых системой управления от конкретных решений из множества возможных решений $L = \{l_1, \dots, l_k\}$, принимаемых по результатам решения задачи распознавания, т. е. величин $G(\Omega^{A\alpha}_i)$, $i=1, \dots, m$; $\alpha=1, \dots, r$.

Первый этап предназначен для построения модели системы распознавания в первом приближении ($a=1$).

Второй и последующие этапы предназначены для уточнения модели системы. Их цель — определить такой вариант разбиения объектов на классы A^0 и такой словарь признаков, при которых критерий R достигает наибольшего значения

Алгоритм реализации 1 этапа следующий.

1. Определяется первый вариант разбиения множества объектов на классы A_1 в соответствии с которым количество классов $m_1=k+1$.

2. Определяется непосредственно либо подмножество множества объектов каждого класса: либо разрабатывается некоторый набор правил относительно значений признаков, содержащихся в априорном словаре, в соответствии с которыми на основе методов самообучения при известном числе классов определяются объекты исходного множества, относящиеся к каждому классу.

3. Производится описание каждого класса на языке признаков априорного словаря, а затем находятся наилучшие решающие границы между классами.

4. Проверяется, достаточна ли величина C_0 для построения измерителей, обеспечивающих определение всех признаков $x_a = \{x_1, \dots, x_N\}$ априорного словаря.

5. Производится описание классов Ω^{A1}_i , $i=1, \dots, m$, на языке рабочего словаря признаков первого приближения и определяются наилучшие решающие границы между ними.

6. Оценивается вероятность правильного решения задачи распознавания.

7. Вычисляется первое приближение значения критерия эффективности системы $R^{(1)}$.

Алгоритм реализации 2 этапов таков.

1. Определяется в алфавите классов первого приближения такой класс Ω^{A1}_v , $v=1, \dots, m$ (либо, исходя из практических соображений, 2—3 класса), для которого величина

$$G(\Omega^{A1}_v) = \min_i G(\Omega^{A1}_i).$$

Это означает, что к классу Ω_v относятся такие объекты, распознавание которых обеспечивает по сравнению с распознаванием объектов других классов наименьший выигрыш.

2. Исключается из алфавита классов первого приближения класс Ω_v^{A1} , а объекты этого класса надлежит отнести к такому классу Ω_μ^{A1} , $\mu=1, \dots, m$, для которого уменьшение величины $G(\Omega_\mu^{A1})$, по сравнению с уменьшением величины $G(\Omega_v^{A1})$, $i=1, \dots, m$, минимально, т. е.

$$\Delta G(\Omega_\mu^{A1}) = \min \Delta G(\Omega_\mu^{A1}), \mu, i=1, \dots, m; \mu \neq i.$$

Таким образом, определяется второй вариант разбиения объектов на классы A_2 , применительно к которому вновь повторяются операции 1—7 и определяется второе приближение значения критерия R^2 эффективности системы. Практически нескольких итераций достаточно для определения такого варианта построения системы, при котором критерий R эффективности системы достигает наибольшего значения.

Системный подход к проблеме распознавания объектов и явлений позволяет в реальных условиях при наличии неизбежных ограничений добиться наибольшей эффективности комплекса «система распознавания + система управления».

Лекция 3 «Классификация систем распознавания»

Общая схема системы распознавания образов:



Любой алгоритм распознавания можно представить как абстрактную функциональную систему R , состоящую из трех компонент:

$R = \{A, S, P\}$, где

$A = \{A_k\}$, $k=1, \dots, K$ – алфавит классов – множество категорий, по которым должны распределить образы,

$S = \{S_j\}$, $j=1, \dots, n$ – словарь признаков – множество характеристик, из которых составляется описание образа,

$P = \{P_l\}$, $l=1, \dots, L$ – множество правил принятия решения.

Функционирование системы:

На вход подается образ – некоторая конфигурация из элементов множества S , к ней применяется определенная последовательность правил из P , в результате конфигурации присваивается индекс, соответствующий одному из элементов множества A .

Качество функционирования системы определяется тем, насколько часто присвоенный образу индекс совпадает с ожидаемым нами результатом.

Компоненты A, S представляют собой информационную часть системы, а P – методологическую.

Система распознавания включает как процесс **синтеза образов**, то есть формирования описаний объектов распознавания и их классов, так и **анализа образов**, то есть сам процесс принятия решений.

Системы распознавания :

- **простые**
- **сложные** в зависимости от того, физически однородная или физически неоднородная информация используется для описания распознаваемых объектов,

имеют ли признаки, на языке которых произведено описание алфавита классов, единую или различную физическую природу.

Примеры: **системы распознавания текста на кассовых чеках** (<http://www.azoft.ru/blog/razrabotka-sistemy-raspoznaniya-teksta-na-kassovyh-chekah/>)

Простые системы распознавания:

Читающие автоматические распознающие устройства, в которых признаки рабочего словаря представляют собой лишь те или иные линейные размеры распознаваемых объектов; автоматы для размена монет, где в качестве признака, используемого при распознавании монет, берется их масса; автоматические устройства, предназначенные для отбраковки деталей, в которых в качестве признаков, применяемых для описания классов бракованных и небракованных деталей, используются либо некоторые линейные размеры, либо масса и т. д.

Сложные системы распознавания:

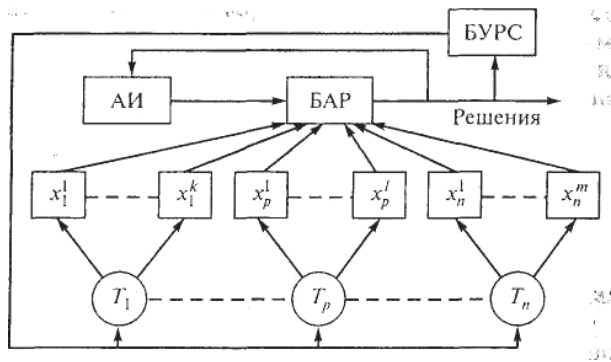
Системы медицинской диагностики, в которых в качестве признаков (симптомов) могут использоваться данные анализа крови и кардиограммы; температура и динамика кровяного давления и т. п.; системы, предназначенные для распознавания образцов геологической разведки, в которых в качестве признаков берутся различные физические и химические свойства, или образцов военной техники вероятного противника и т. д.

Классификации по способу получения апостериорной информации.

Сложные системы делятся:

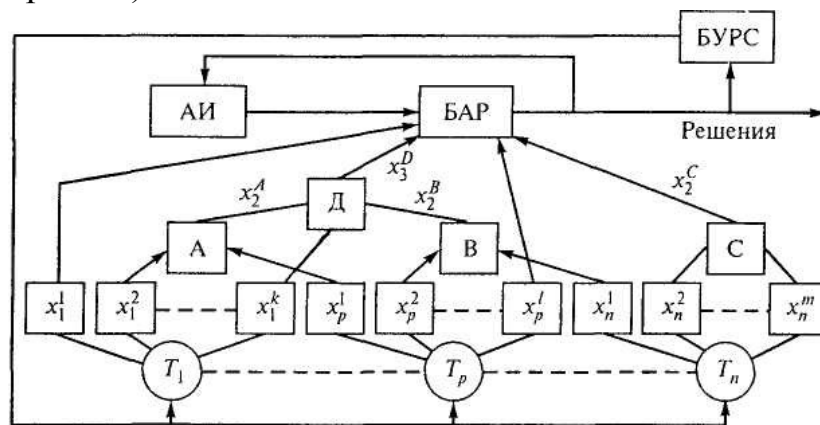
- на одноуровневые и многоуровневые.

Одноуровневые сложные системы. Апостериорная информация о признаках распознаваемых объектов $x_1 \dots, x_N$ определяется прямыми измерениями непосредственно на основе обработки результатов экспериментов (АИ — априорная информация; БАР — блок алгоритмов распознавания; БУРС — блок управления работой средств).



По данным технических средств $T_1, \dots, T_p, \dots, T_n$, на основе обработки полученных реализаций непосредственно находят признаки $x_1^1, \dots, x_1^k, x_1^p, \dots, x_1^l, x_1^n, \dots, x_n^m$ неизвестных объектов или явлений, которые используются для их распознавания.

Многоуровневые сложные системы. В этих системах апостериорная информация о признаках определяется на основе косвенных измерений. Для таких измерений используются специализированные локальные распознающие системы (АИ — априорная информация; БАР — блок алгоритмов распознавания; БУРС — блок управления работой средств).



В многоуровневых системах по данным технических средств $T_1, \dots, T_p, \dots, T_n$, определяются признаки $x_1^1, \dots, x_1^k, x_1^p, \dots, x_1^l, x_1^n, \dots, x_n^m$ (назовем их первичными), которые подразделяются на группы.

Группа 1. Признаки используемые в локальных распознающих устройствах первого (нижнего) уровня для определения признаков второго уровня ($x_2^1, x_2^p, x_2^l, x_2^n, x_2^m, x_n^m$.) На основе этих признаков распознающие устройства первого уровня А, В, С определяют признаки второго уровня x_2^A, x_2^B, x_2^C .

Группа 2. Признаки, непосредственно используемые в распознающих устройствах второго уровня для определения признаков третьего уровня, таким признаком является x_1^k , используемый наряду с признаками второго уровня x_2^A и x_2^B в распознающем устройстве второго уровня D для определения признака третьего уровня x_3^D .

Группа 3. Признаки, используемые в распознающих устройствах третьего уровня для определения признаков четвертого уровня, и т. д.

Группа 4. Признаки, непосредственно используемые в процессе распознавания неизвестных объектов, т. е. признаки, входящие в рабочий словарь признаков системы распознавания, такими признаками являются x^1_1 и x^1_p (назовем эти признаки признаками верхнего уровня).

Классификация по количеству первоначальной априорной информации о распознаваемых объектах или явлениях.

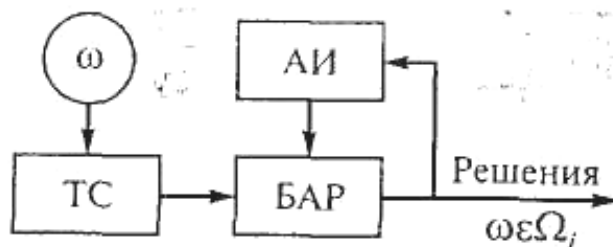
Системы распознавания, как простые, так и сложные, можно разделить на системы:

- без обучения,
- обучающиеся
- самообучающиеся.

Многоуровневые системы распознавания однозначно не подразделяются на указанные классы, так как каждая из локальных систем многоуровневой системы может, в свою очередь, представлять собой систему без обучения, обучающуюся либо самообучающуюся.

Системы без обучения.

В этих системах первоначальной априорной информации достаточно для того, чтобы определить априорный алфавит классов, построить априорный словарь признаков и на основе непосредственной обработки исходных данных произвести описание каждого класса на языке этих признаков, т. е. в первом приближении достаточно определить решающие границы, решающие правила. **Структурная схема системы без обучения.**



ТС — технические средства;

АИ — априорная информация;

БАР — блок алгоритмов распознавания

Обучающиеся системы.

Первоначальной априорной информации достаточно для того, чтобы определить априорный алфавит классов и построить априорный словарь признаков, но недостаточно для описания классов на языке признаков.

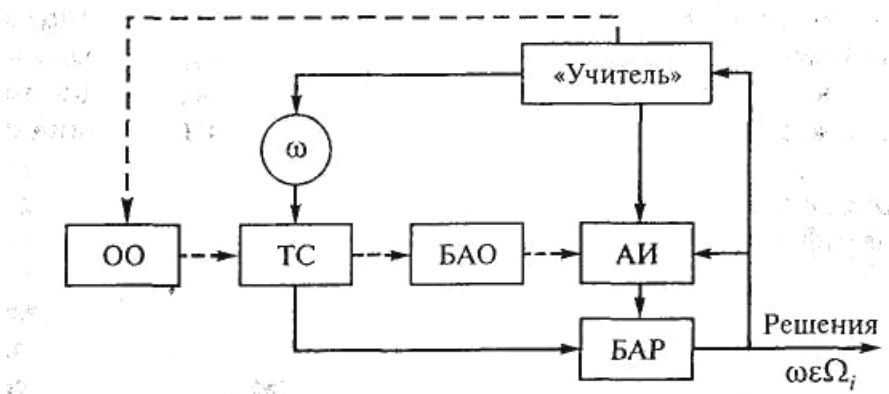
Исходная информация, необходимая для построения обучающихся систем распознавания, позволяет выделить конкретные объекты, принадлежащие различным классам.

$$\left. \begin{aligned} (\omega_1, \omega_2, \dots, \omega_r) &\in \Omega_1; \\ \dots &\dots \dots \dots \dots \dots \dots \\ (\omega_{r+1}, \omega_{r+2}, \dots, \omega_q) &\in \Omega_2; \\ \dots &\dots \dots \dots \dots \dots \dots \\ (\omega_{g+1}, \omega_{g+2}, \dots, \omega_l) &\in \Omega_m. \end{aligned} \right\}$$

Объекты $\omega_1 \dots, \omega_t$ представляют собой обучающие объекты (обучающая последовательность, обучающая выборка).

Цель процедуры обучения — определение разделяющих функций $F_i(x_1, \dots, x_N)$, $i = 1, \dots, m$, путем многократного предъявления системе распознавания различных объектов с указанием классов, к которым эти объекты принадлежат.

Структурная схема обучающейся системы

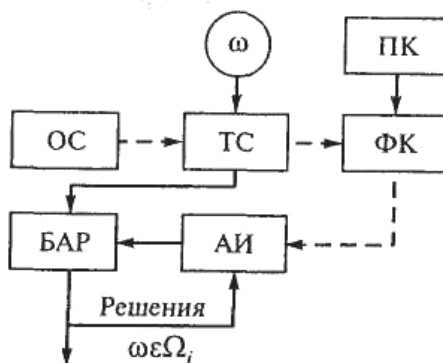


- ОО — обучающие объекты;
- ТС — технические средства;
- БАО — блок алгоритмов обучения;
- АИ — априорная информация;
- БАР — блок алгоритмов распознавания;
- штриховые линии — режим обучения
- сплошные линии — «экзамен»

Самообучающиеся системы.

Первоначальной априорной информации достаточно лишь для определения словаря признаков x_1, \dots, x_N , но недостаточно для проведения классификации объектов.

На стадии формирования системы ей предъявляют исходную совокупность объектов $w_1 \dots, w_l$, заданных значениями своих признаков для $w_1 — (x^1_1 \dots, x^1_N); \dots$; для $w_l — (x^l_1 \dots, x^l_N)$. Но из-за ограниченного объема первоначальной информации система не получает указаний о том, к какому классу объекты исходной совокупности принадлежат. Эти указания заменяются набором правил, в соответствии с которыми на стадии самообучения система распознавания сама вырабатывает классификацию.



ОС — объекты для самообучения; ТС — технические средства; БАР — блок алгоритмов распознавания; АИ — априорная информация; ПК — правила классификации; ФК — режим самообучения; сплошные линии — формирование классов; штриховые линии — распознавание неизвестных объектов

Цель обучения или самообучения — выработать такое количество информации, которое необходимо для функционирования системы распознавания.

Обучающиеся и самообучающиеся системы распознавания применяются, когда отсутствует полная первоначальная априорная информация.

При построении систем распознавания используется **принцип обратной связи: - результаты решения задачи распознавания неизвестных объектов следует использовать для уточнения априорного описания классов.**

Для этого блок априорной информации содержит специальные алгоритмы корректировки априорных описаний классов

Классификация по характеру информации о признаках распознаваемых объектов.

Системы делятся: **на**

- детерминированные
- вероятностные
- логические
- структурные

Детерминированные системы.

Для построения алгоритмов распознавания используются «геометрические» меры близости, основанные на измерении расстояний между распознаваемым объектом и эталонами классов. Применение детерминированных методов распознавания предусматривает наличие координат эталонов классов в признаковом пространстве либо координат объектов, принадлежащих соответствующим классам.

Вероятностные системы.

Для построения алгоритмов распознавания используются вероятностные методы распознавания, основанные на теории статистических решений. Применение вероятностных методов распознавания предусматривает наличие вероятностных зависимостей между признаками распознаваемых объектов и классами, к которым эти объекты относятся.

Логические системы.

Для построения алгоритмов распознавания используются логические методы распознавания, основанные на дискретном анализе и базирующемся на нем исчислении высказываний. Применение логических методов распознавания предусматривает наличие логических связей, выраженных через систему булевых уравнений, в которой переменные — логические признаки распознаваемых объектов, а неизвестные величины — классы, к которым эти объекты относятся

Структурные (лингвистические) системы.

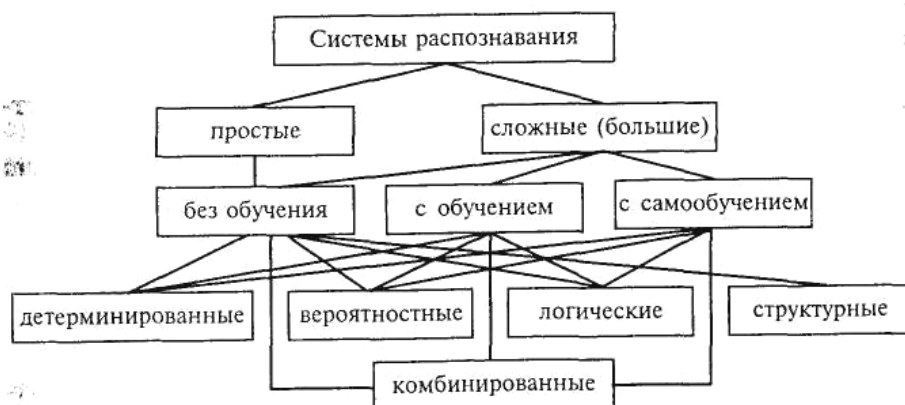
Для построения алгоритмов распознавания используются специальные грамматики, порождающие языки, состоящие из предложений, каждое из которых описывает объекты, принадлежащие конкретному классу. Применение структурных методов распознавания требует наличия совокупностей предложений, описывающих все множество объектов, принадлежащих всем классам алфавита классов системы распознавания.

Множество предложений должно быть подразделено на подмножества по числу классов системы. Элементами подмножеств являются предложения, описывающие объекты, принадлежащие данному подмножеству (классу). Априорными описаниями классов являются совокупности предложений, каждое из которых соответствует конкретному объекту, принадлежащему данному классу.

Комбинированные системы.

Для построения алгоритмов распознавания используется специально разработанный метод вычисления оценок. Такие алгоритмы распознавания называют алгоритмами вычисления оценок (АВО). Их применение требует наличия таблиц, где

содержатся объекты, принадлежащие соответствующим классам, а также значения признаков, которыми характеризуются эти объекты. Признаки могут быть детерминированными, логическими, вероятностными и структурными.



Лекция 4 «Качественное описание задачи распознавания»

Распознавание представляет собой задачу преобразования входной информации, в качестве которой уместно рассматривать некоторые параметры, признаки распознаваемых образов, в выходную, представляющую собой заключение о том, к какому классу относится распознаваемый образ.

Распознавание образов есть один из разделов кибернетики.

Кибернетика есть наука об общих законах преобразования информации в сложных системах.

Создание искусственного интеллекта — это, в том числе и построение распознающих систем, приближающихся по своим возможностям к возможностям человека в решении задач распознавания.

Пример задачи распознавания

Задача: сторона А должна распознать самолеты стороны В.

Для построения системы распознавания необходимо:

- провести детальный анализ всей доступной информации об авиации стороны В
- исходя из анализа тактико-технических характеристик своих средств противодействия самолетам стороны В, оценить, какие решения она может принимать в случае налета самолетов стороны В - это:

- 1) применить средство противодействия S_1 ,
- 2) применить средство противодействия S_2 ;
- 3) совместно использовать средства S_1 и S_2 .

Положим, первый класс — бомбардировщики, второй класс — истребители, третий класс — штурмовики.

Этапы выполнения:

1. Определить классы

Наличие конкретных технических средств обнаружения самолетов и определения их параметров, а также недостаточный объем исходной (априорной) информации о классах самолетов стороны В может привести к тому, что с точки зрения эффективности стороне А целесообразно ввести в рассмотрение только два класса.

2. Определить, с помощью каких параметров или признаков можно описать выделенные классы самолетов.

3. Из полученного перечня исключить те признаки, относительно которых не представляется возможным определить их значения применительно к каждому классу самолетов.

4. Из полученного перечня признаков надо выделить признаки, которые могут быть реально определены (например, крейсерская и максимальная скорости,

предельная высота полета, число и тип двигателей, длина фюзеляжа, размах крыльев и др.).

5. На основе априорных данных следует описать на языке выбранных признаков каждый класс самолетов.

Одни признаки имеют качественный характер (тип двигателей), другие — количественный (скорость, высота полета и т. д.).

В описании классов должны содержаться сведения :

1. Какие признаки качественного характера присущи каждому классу.
2. Диапазоны или законы распределений значений количественных признаков.

Выполнена подготовительная работа: проанализирована априорная информация о самолетах, произведена их классификация, выбрана система признаков и описаны все классы самолетов на языке этих признаков.

Сопоставление полученных апостериорных данных об этом самолете с данными, заключенными в априорном описании всех классов самолетов на языке признаков, позволяет определить, к какому классу относится неизвестный самолет, т. е. позволяет произвести его распознавание.

Задача распознавания.

1. Имеется некоторая совокупность объектов или явлений.
2. В соответствии с выбранным принципом классификации она подразделена на ряд классов, т. е. составлен **алфавит классов**.
3. Разработан **словарь признаков**, на языке которого описывается каждый класс объектов.
4. Созданы **технические средства**, обеспечивающие определение признаков.
5. На вычислительных средствах системы распознавания **реализован алгоритм распознавания**, позволяющий сопоставлять апостериорные данные о неизвестном объекте с априорной информацией и на основе сопоставления определять, к какому классу он может быть отнесен.

В рамках кибернетики во второй половине 50-х годов XX в. начало формироваться новое научное направление, связанное с разработкой теоретических основ и практической реализацией устройств, а затем и систем, предназначенных для распознавания неизвестных объектов, явлений, процессов.

Первые работы в области распознавания образов были посвящены главным образом теории и практике построения читающих автоматов, и само слово «образ» использовалось для обозначения напечатанного или написанного от руки знака, изображающего букву или цифру. Математическим аппаратом постановки и решения задач распознавания с момента их возникновения явилась теория статистических решений.

Приоритетны - многоуровневые систем распознавания, представляющих собой совокупность технических средств и программно-алгоритмических комплексов.

К **техническим средствам** относятся измерительные системы или измерительная аппаратура, предназначенная для обнаружения распознаваемых объектов и определения признаков, на языке которых они описываются, а также локальные вычислительные машины, входящие в состав измерительных систем, и центральные ЭВМ системы.

Программно-алгоритмический комплекс - совокупность алгоритмов, предназначенных для обработки апостериорной измерительной информации и определения признаков распознаваемых объектов, программно реализованных на вычислительных средствах измерительных систем и **алгоритмов распознавания**.

Область применения:

- Системы технической диагностики машин и механизмов
- Система распознавания цели
- Многомерные медицинские данные с помощью ЭВМ могут быть соотнесены в когнитивный графический образ в виде интегральных функциональных профилей или сцен,
- В сельском хозяйстве. Здесь следует иметь в виду распознавание размера урожая конкретных сельскохозяйственных культур на определенных участках поля по данным аэро-и космических наблюдений, значительное уменьшение ручного труда при сортировке плодов по их форме, цвету и размерам, массовую медицинскую диагностику сельскохозяйственных животных, автоматическую дойку роботами, снабженными системами распознавания

Проектирование систем распознавания – построение уточняющейся математической или физико-математической модели проектируемой системы.

Первая итерация может быть названа априорной (исходной). На ее основе составляется априорный алфавит классов и разрабатывается априорный словарь признаков, определяются границы классов в априорном признаковом пространстве. Определение алфавита классов и словаря признаков.

Вторая итерация – нахождение новых границ между классами, возможной корректировки алфавита классов и словаря признаков по уточненной модели системы распознавания служит основой. Этот процесс итеративный.

Математическое обеспечение систем распознавания – это математическая модель системы, методы и алгоритмы обработки измерительной информации, получаемой техническими средствами системы и предназначенной для определения признаков распознаваемых объектов; методы и алгоритмы оптимального управления процессом распознавания; методы и алгоритмы оценки эффективности системы распознавания

как на стадии проектирования, так и в процессе ее функционирования и т. д.

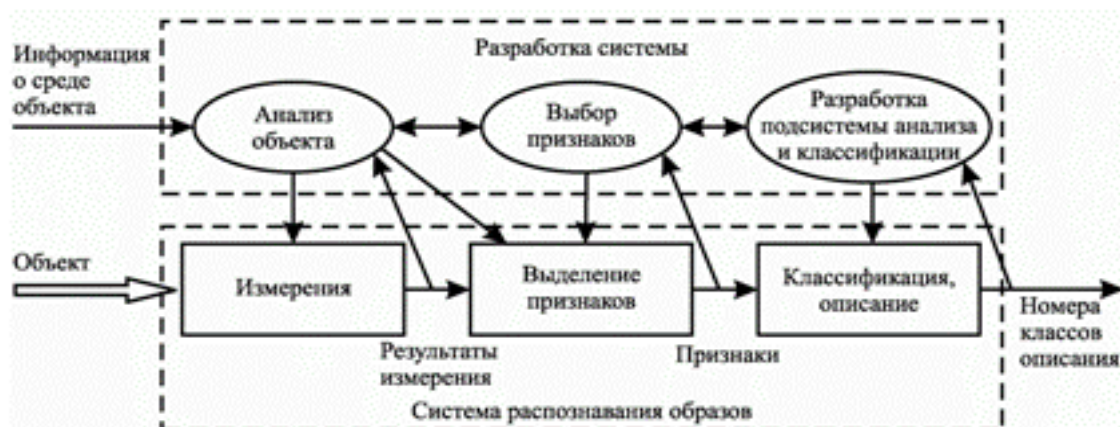
Основные задачи построения систем распознавания

Каждая система распознавания приспособлена для распознавания только данного вида объектов или явлений

Задача 1 заключается в том, чтобы определить полный перечень признаков (параметров), характеризующих объекты или явления, для распознавания которых разрабатывается данная система.

Признаки объектов делятся на :

- детерминированные,
- вероятностные,
- логические
- структурные.



Детерминированные признаки — признаки, принимающие конкретные числовые значения (например, размах крыла $l_{кр} = 25$ м, длина фюзеляжа $l_{ф} = 50$ м, масса самолета $G=70$ тит. д.), которые могут рассматриваться в качестве координат точки в признаковом пространстве, соответствующей данному объекту. При рассмотрении признаков в качестве детерминированных ошибками измерений пренебрегают.

Вероятностные признаки — признаки, случайные значения которых распределены по всем классам объектов, при этом решение о принадлежности распознаваемого объекта к тому или другому классу может приниматься только на основании конкретных значений признаков данного объекта, определенных в результате проведения соответствующих опытов.

Логические признаки распознаваемых объектов можно рассматривать как элементарные высказывания, принимающие два значения истинности («да», «нет» или «истина», «ложь») с полной определенностью. К логическим признакам относятся прежде всего признаки, не имеющие количественного выражения. К логическим

можно отнести также признаки, у которых важна не величина признака у распознаваемого объекта, а лишь факт попадания или непадения ее в заданный интервал. Например: симптомы болезни в медицинской диагностики

Структурные (лингвистические, синтаксические) признаки представляют собой производные элементы (символы) структуры объекта. Иначе эти элементы (константы) называют терминалами. Каждый объект может рассматриваться как цепочка терминалов или как предложение. Эти предложения и описывают объекты. При этом если предложение, описывающее неизвестный распознаваемый объект, относится к языку данного класса, то этот объект и принимается принадлежащим к этому классу. Например, при распознавании букв русского алфавита терминалами являются вертикальная, горизонтальная, наклонная черточки, наличие угла и т. д.

Задача 2 заключается в проведении первоначальной классификации распознаваемых объектов или явлений, в составлении априорного алфавита классов. Основное в данной задаче — выбор надлежащего принципа классификации.

Задача 3 состоит в разработке априорного словаря признаков. Словарь разрабатывается на основе результатов решения первой задачи с учетом того, что в априорный словарь признаков включаются только те признаки, относительно которых может быть получена априорная информация, необходимая для описания классов на языке этих признаков.

Задача 4 состоит в описании всех классов априорного алфавита классов на языке признаков, включенных в априорный словарь признаков. (Для ее решения могут быть использованы методы непосредственной обработки исходных данных, обучения или самообучения)

Если признаки распознаваемых объектов — детерминированные, то описанием каждого класса объектов на языке этих признаков является его эталон, т. е. точка, сумма расстояний которой от точек, описывающих объекты, принадлежащие данному классу, минимальна.

Если признаки распознаваемых объектов логические и имеют количественные выражения, то для описания классов объектов на языке признаков необходимо определить диапазоны значений признаков $\Delta x_j^i, j=1, \dots, N$, соответствующие классам $\Omega_h, i=1, \dots, m$.

Если признаки распознаваемых объектов есть суждения качественного характера, то каждый рассматривается как элементарное логическое высказывание A', B', C', \dots . Для описания классов на языке этих признаков необходимо выяснить, какими из них характеризуется каждый класс, после этого установить зависимости в форме булевых соотношений между признаками $A, B, C, \dots; A', B', C', \dots$ и классами $\Omega_1, \dots, \Omega_m$.

Если распределение объектов по областям D_i N -мерного пространства признаков для всех значений $i=1, \dots, m$ вероятностное, то для описания классов необходимо

определить характеристики этих распределений: функции плотности вероятности $f_i(x_1, \dots, x_N)$ значений параметров x_1, \dots, x_N при условии, что объекты принадлежат классу Ω_i ; априорные вероятности $P(\Omega_i)$ того, что объект, случайным образом выбранный из общей совокупности, окажется принадлежащим классу Ω_1 .

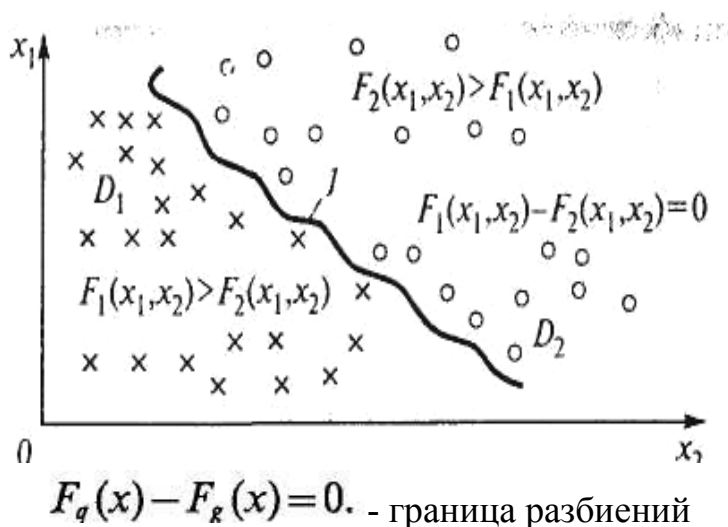
Если признаки распознаваемых объектов — структурные, то описаниями классов объектов являются языки, состоящие из предложений, каждое из которых характеризует структурные особенности объектов, принадлежащих исключительно одному из классов.

Задача 5 заключается в разбиении априорного пространства признаков на области, соответствующие классам априорного алфавита классов

Пусть в априорном словаре признаков содержится упорядоченный набор параметров объектов или явлений — признаки x_1, \dots, x_N , которые можно рассматривать как составляющие вектора $x_a = \{x_1, \dots, x_N\}$, описывающего априорное пространство признаков систем распознавания (априорное признаковое пространство) размерности N ; конкретные точки этого пространства представляют собой распознаваемые объекты.

Положим, произведено разбиение объектов на классы $\Omega_1, \dots, \Omega_m$. Требуется выделить в пространстве признаков области $D_i, i = 1, \dots, m$, эквивалентные классам, т. е. характеризуемые следующей необходимой зависимостью: если объект, имеющий признаки x^0_1, \dots, x^0_N , относится к классу Ω_i , то представляющая его в признаковом пространстве точка принадлежит области D_i .

Алгебраический вид задачи: требуется построить разделяющие функции $F_i(x_1, \dots, x_N), i = 1, \dots, m$, обладающие следующим свойством: если объект, имеющий признаки x^0_1, \dots, x^0_N , относится к классу Ω_i то величина $F_i(x^0_1, \dots, x^0_N)$ должна быть наибольшей. Такой же она должна быть и для всех других значений признаков объектов, относящихся к классу Ω_i . Если x_q обозначает вектор признаков объекта, принадлежащего к Ω_q -му классу, то для всех значений вектора x_q $F_q(x_q) > F_g(x_q), q, g = 1, \dots, m, q \neq g$.



Задача 6 состоит в выборе алгоритмов распознавания, обеспечивающих отнесение распознаваемого объекта или явления к тому или другому классу или их некоторой совокупности.

Алгоритмы распознавания основываются на сравнении той или другой меры близости или меры сходства распознаваемого объекта с каждым классом. При этом если выбранная мера близости L данного объекта w с каким-либо классом Ω_g , $g = 1, \dots, m$, превышает меру его близости с другими классами, то принимается решение о принадлежности этого объекта классу Ω_g , т. е.

В алгоритмах распознавания, базирующихся на использовании детерминированных признаков, в качестве меры близости используется среднее квадратичное расстояние между данным объектом w и совокупностью объектов $\{w_{g1}, \dots, w_{gk_g}\}$, представляющих собой класс Ω_g :

$$L(w, \Omega_g) = \sqrt{k_g^{-1} \sum_{s=1}^{k_g} d^2(w, w_{gs})},$$

В случае, если необходимо учитывать веса W_j признаков x_j , $j=1, \dots, N$, объекта w и признаков x_{gsj} объектов w_{gs} класса Ω_g , может быть применена метрика следующего вида:

$$L(w, \Omega_g) = \sqrt{k_g^{-1} \sum_{s=1}^{k_g} \sum_{j=1}^N W_j (x_j - x_{gsj})^2}.$$

В алгоритмах распознавания, базирующихся на использовании вероятностных признаков, в качестве меры близости используется риск, связанный с решением о принадлежности распознаваемого объекта к классу Ω_i , $i=1, \dots, m$. Пусть даны описания классов $\{f_i(x), P(\Omega_i)\}$, $x = \{x_1, \dots, x_N\}$ и риски правильных и ошибочных решений, представляющие собой элементы платежной матрицы вида:

$$C = \begin{vmatrix} c_{11} & c_{12} & \dots & c_{1m} \\ \cdot & \cdot & \cdot & \cdot \\ c_{m1} & c_{m2} & \dots & c_{mm} \end{vmatrix}$$

По главной диагонали матрицы расположены потери при правильных решениях, а по обеим сторонам от нее — потери при ошибочных решениях. Если $c_{ii} < 0$, $i=1, \dots, m$, то такие отрицательные потери можно рассматривать как выигрыш при правильных решениях.

Пусть в результате экспериментов установлено, что значения признаков у распознаваемого объекта w составляют $x_1 = x_1^0$, $x_2 = x_2^0$, ..., $x_N = x_N^0$. Обозначим это событие a_N . Тогда значение риска, связанного с решением вида $w \in \Omega_g$ при условии, что имеет место событие a_N , будет

$$R(w \in \Omega_g | a_N) = R(\Omega_g | a_N) = \sum_{i=1}^m c_{ig} P(\Omega_i | a_N),$$

где условная апостериорная вероятность того, что $w \in \Omega_i$ в соответствии с теоремой гипотез или формулой Байеса

$$P(\Omega_i/a_N) = P(\Omega_i) f_i(x_1^0, \dots, x_N^0) / \sum_{i=1}^m P(\Omega_i) f_i(x_1^0, \dots, x_N^0). \quad (1.6)$$

В общем случае решение вида $w \in \Omega_g$ принимается в случае, если

$$R(\Omega_g/a_N) = \min_i R(\Omega_i/a_N).$$

В алгоритмах распознавания, базирующихся на использовании логических признаков, не используется понятие «мера близости». Когда построено описание классов на языке логических признаков в виде соответствующих булевых соотношений (эквивалентности или импликаций), при подстановке в эти соотношения значений признаков, характеризующих распознаваемый объект, автоматически возникает ответ: к какому классу или классам этот объект может быть отнесен и к каким он не относится.

В алгоритмах распознавания, базирующихся на использовании структурных (лингвистических) признаков, понятие меры близости также может не использоваться. Когда построены языки, описывающие классы в виде совокупностей предложений, характеризующих структурные особенности объекта, относящиеся к каждому классу, то распознавание неизвестного объекта осуществляется идентификацией предложения, описывающего этот объект, с одним из предложений языка — элемента описания соответствующего класса.

Задача 7 состоит в определении рабочего алфавита классов и рабочего словаря признаков системы распознавания. Она представляет собой общую постановку проблемы распознавания.

Суть ее заключается в разработке такого алфавита классов и такого словаря признаков (назовем их оптимальными), которые в условиях ограничений на построение системы распознавания обеспечивают максимальное значение показателя эффективности системы управления, принимающей в зависимости от результатов распознавания неизвестных объектов соответствующие решения.

Задачу можно решить с помощью математической (физико-математической) модели системы распознавания путем последовательных приближений. Первое приближение системы — априорный словарь признаков и априорный алфавит классов.

При построении рабочих алфавита классов и словаря признаков необходимо учитывать:

1. возможность создания технических средств наблюдений, обеспечивающих на основе проведения экспериментов определение признаков распознаваемых объектов и целесообразность их использования.

2. обеспечение в условиях названных ограничений наибольшей точности решения задачи распознавания, так как она непосредственно влияет на эффективность управленческих решений

Задача 8 состоит в разработке специальных алгоритмов управления работой системы.

Задача алгоритмов управления в том, чтобы процесс функционирования системы распознавания был оптимальным и выбранный критерий качества этого процесса достигал экстремального значения.

В качестве подобного критерия могут использоваться, например, вероятность правильного решения задачи распознавания, среднее время ее решения, расходы, связанные с реализацией процесса распознавания, и т. д.

Достижение экстремальной величины критериев выполняется при некоторых ограничивающих условиях (время решения задачи, расходы)

Задача 9 состоит в выборе показателей эффективности системы распознавания и оценке их значений.

В качестве показателей эффективности системы могут рассматриваться вероятность правильных решений, среднее время решения задач распознавания, величина расходов, связанных с получением апостериорной информации, и т. д.

Оценка значений выбранной совокупности показателей эффективности, как правило, проводится на основе экспериментальных исследований либо реальной системы распознавания, либо с помощью ее физической или математической модели.

Лекция 5 «Классификация с помощью решающих функций»

Определения:

1) множество образов U ;

2) пространство признаков X ,

$$x \in U \Rightarrow x \in X;$$

3) множество классов $\Omega = \{ \omega_1, \dots, \omega_m \}$;

4) множество предпочтений $0 \leq 1$ $\{X_0, X_1, \dots, X_m\}$ классов

$\Omega = \{ \omega_1, \dots, \omega_m \}$ в пространстве X

$$x \in \omega_i, \text{ если } x \in X_i, i = 1, \dots, m;$$

если $x_0 \in X$, то образ x попал в область «неопределенности»;

$\{X_0, X_1, \dots, X_m\}$ - полная группа множеств, т.е. $\bigcup_{i=0}^m X_i = X$, $X_i \cap X_j = \emptyset$ для всех $i \neq j$.

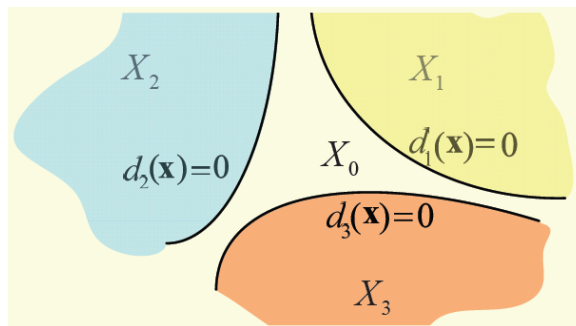
Понятие решающих функций

$d_j(x)$ – решающая (дискриминантная) функция,

$X_j = \{x \in R^n : d_j(x) > 0\}$ – область предпочтения,

$S_j = \{x \in R^n : d_j(x) = 0\}$ – разделяющая поверхность,

образ $x \in \omega_i$, если $d_j(x) < 0 \forall j \neq i$ и $d_i(x) > 0$.



Линейные решающие функции (ЛРФ)

ЛРФ – функция вида

$$d(\mathbf{x}) = d(x_1, \dots, x_n) = w_1 x_1 + \dots + w_n x_n + w_{n+1} = (\mathbf{w}, \mathbf{x}),$$

где $\mathbf{w} = (w_1, \dots, w_n, w_{n+1})^T$ - вектор весовых коэффициентов ЛРФ,

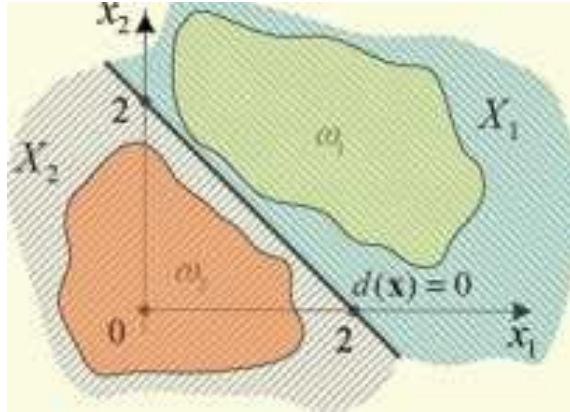
$\mathbf{x} = (x_1, \dots, x_n, 1)^T$ - вектор признаков образа x .

Разделяющая поверхность

$$S = \{x \in R^n : d(\mathbf{x}) = 0\}$$

- гиперплоскость в R^n .

$d(\mathbf{x}) = d(x_1, x_2) = w_1 x_1 + w_2 x_2 + w_3 = 0$,
 где $d(\mathbf{x}) = x_1 + x_2 - 2$, $\mathbf{w} = (1, 1, -2)$. Тогда
 $x \in \varpi_1$, если $d(\mathbf{x}) > 0 \Leftrightarrow x_1 + x_2 - 2 > 0$
 и $x \in \varpi_2$, если $d(\mathbf{x}) < 0 \Leftrightarrow x_1 + x_2 - 2 < 0$.



Условия линейной разделимости классов

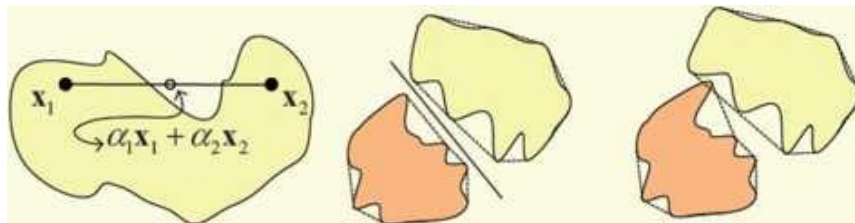
Теорема 1. 1) Два множества векторов линейно разделимы тогда и только тогда, когда их выпуклые оболочки не пересекаются;

2) линейно независимая система векторов в R^n линейно разделима на любые два класса.

Пусть $y_i = \begin{cases} 1, & \mathbf{x}_i \in \varpi_1, \\ -1, & \mathbf{x}_i \in \varpi_2 \end{cases}$ — метки векторов двух классов.

Теорема 1'. Векторы двух классов $\{x_1, \dots, x_k\} \in \varpi_1$ и $\{x_{k+1}, \dots, x_n\} \in \varpi_2$ в R^N не разделимы никакой гиперплоскостью тогда и только тогда, когда выпуклая оболочка векторов $\{y_i \mathbf{x}_i\}_{i=1}^N$ содержит начало координат.

Следствие (из т. Хана-Банаха). Если A и B - непересекающиеся выпуклые множества в R^n , то существует гиперплоскость, разделяющая эти множества.



$con(M) := \left\{ \sum_{i=1}^p \alpha_i \mathbf{x}_i : \mathbf{x}_i \in M, \alpha_i \geq 0, \sum_{i=1}^p \alpha_i = 1 \right\}$ - выпуклая оболочка множества M .

Теорема Хана — Банаха и некоторые её следствия

Приведём сначала теорему Хана — Банаха и некоторые её следствия в той форме, в которой они нам понадобятся.

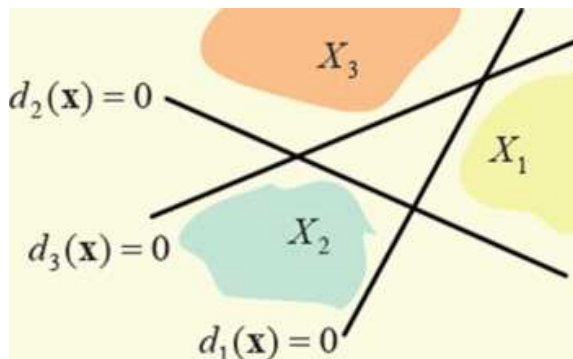
Теорема (Хан, Банах). Пусть X — комплексное нормированное пространство, f_0 — линейный ограниченный функционал, определённый на подпространстве $L \subset X$. Тогда существует линейный ограниченный функционал f , определённый на всём X и удовлетворяющий условиям

$$\begin{aligned}\langle f, x \rangle &= \langle f_0, x \rangle, \quad x \in L, \\ \|f\| &= \|f_0\|_{L}.\end{aligned}$$

Иначе говоря, ограниченный линейный функционал, заданный на подпространстве нормированного пространства, можно продолжить на всё пространство с сохранением нормы.

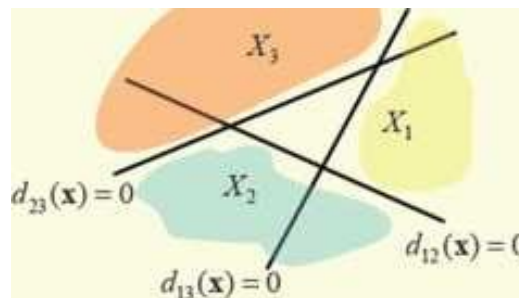
Случаи линейной разделимости классов

1. Каждый класс отделяется от всех остальных классов одной линейной решающей функцией



$x \in X_i$, если $d_i(x) > 0$ и $d_j(x) < 0$ для всех $j \neq i$

2. Каждые два класса можно разделить одной линейной разделяющей поверхностью



Если функция $d_{ij}(x) = (\mathbf{w}_{ij}, \mathbf{x})$ разделяет i -й и j -й классы, то образ $x \in \omega_i$, если $d_{ij}(x) > 0$ для всех $j \neq i$.

3. Любой из m классов отделяется от всех остальных классов с помощью m решающих функций $d_i(\mathbf{x})$, $i = 1, 2, \dots, m$.

Образ $x \in \omega_i$, если $d_i(\mathbf{x}) > d_j(\mathbf{x})$ для всех $j \neq i$.

Если $d_{ij}(\mathbf{x}) = d_i(\mathbf{x}) - d_j(\mathbf{x})$, то набор решающих функций будет удовлетворять случаю 2.

Общий подход к нахождению ЛРФ

Пусть (Ξ, Y) - множество прецедентов,

$\Xi = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ - обучающая выборка,

$Y = \{y_1, \dots, y_N\}$ - множество меток двух классов,

$d(\mathbf{x}) = (\mathbf{w}, \mathbf{x})$ - ЛРФ.

Задача: найти вектор \mathbf{w} такой, что

$$(\mathbf{w}, \mathbf{x}) > 0, \text{ если } \mathbf{x} \in \omega_1 \text{ и } (\mathbf{w}, \mathbf{x}) < 0, \text{ если } \mathbf{x} \in \omega_2.$$

Для унифицированных векторов

$$\mathbf{x}' = \begin{cases} \mathbf{x}, & \mathbf{x} \in \omega_1, \\ -\mathbf{x}, & \mathbf{x} \in \omega_2 \end{cases}$$

вектор \mathbf{w} должен удовлетворять СЛАН

$$(\mathbf{w}, \mathbf{x}_i') > 0 \text{ для всех } i = 1, \dots, N$$

Метод градиентного спуска рассматривает задачу поиска минимума функции $f(\mathbf{x})$, записываемой в виде:

Пусть функция $f(\mathbf{x}): \mathbb{R}^n \rightarrow \mathbb{R}$ такова, что можно вычислить ее градиент.

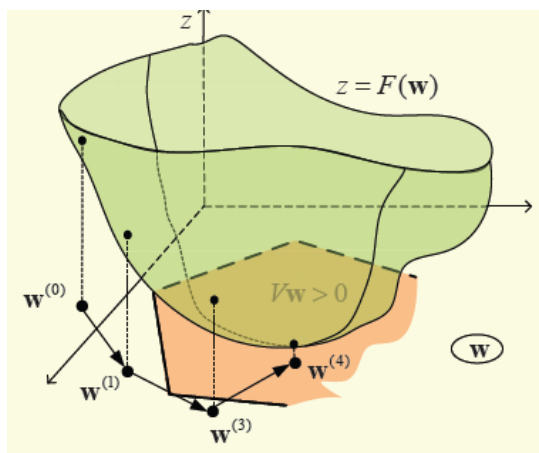
Тогда можно применить метод градиентного спуска, его варианты:

- Метод градиентного спуска с постоянным шагом
- Метод градиентного спуска с дроблением шага
- Метод наискорейшего спуска

Основная идея метода заключается в том, чтобы осуществлять оптимизацию в направлении наискорейшего спуска, а это направление задаётся антиградиентом :

$$f(\mathbf{x}) \rightarrow \min_{\mathbf{x} \in \mathbb{R}^n} \text{Метод_градиентного_спуска}$$

Метод градиентного спуска



$$\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} - h_k \text{grad}(F(\mathbf{w}^{(k)})), k = 1, 2, \dots$$

В этом случае вводится, так называемая функция критерия $F(\mathbf{w})$, и ищется вектор \mathbf{w} , минимизирующий функцию $F(\mathbf{w})$. Минимум функции $F(\mathbf{w})$ можно найти методом градиентного спуска с помощью следующей итерационной процедуры где h_k – итерационный шаг. Функцию критерия можно вводить разными способами. Важно, чтобы минимум этой функции достигался для того вектора \mathbf{w} , для которого выполняется условие

$$(\mathbf{w}, \mathbf{x}'_i) > 0 \text{ для всех } i = 1, \dots, N$$

Кроме того, эта функция должна удовлетворять определенным требованиям, гарантирующим сходимость итерационной процедуры.

Функция критерия

Пример.

$$F(\bar{\mathbf{w}}) = -\sum_{\mathbf{x}' \in E(\bar{\mathbf{w}})} (\bar{\mathbf{w}}, \mathbf{x}'), \text{ где } E(\bar{\mathbf{w}}) = \{\mathbf{x}' \in \Xi : (\bar{\mathbf{w}}, \mathbf{x}') \leq 0\}.$$

Равносильные задачи:

$$\text{найти } \mathbf{w} \in R^n : V\mathbf{w} > 0 \Leftrightarrow \text{найти } \mathbf{w} \in R^n : V\mathbf{w} = \mathbf{y}, \mathbf{y} \in (R^+)^N$$

Переопределенная задача:

$$\text{найти } \mathbf{w} \in R^n, \mathbf{y} \in (R^+)^N : V\mathbf{w} = \mathbf{y}.$$

Псевдорешение:

$$\text{найти } \mathbf{w} \in R^n, \mathbf{y} \in (R^+)^N : F(\mathbf{w}, \mathbf{y}) = \frac{1}{2} \|V\mathbf{w} - \mathbf{y}\|_2^2 \rightarrow \min.$$

Так как $F(\mathbf{w}, \mathbf{y})$ - квадратичный неотрицательно определенный функционал в выпуклой области, то существует единственное псевдорешение.

Алгоритм Хо-Кашьяпа

1. Выбирается произвольный вектор $\mathbf{y}^{(0)}$ с N положительными координатами, вычисляется $\mathbf{w}^{(0)} = V^{\otimes} \mathbf{y}^{(0)}$ и полагается $k = 0$.

2. Проверяется условие останова $V\mathbf{w}^{(k)} > 0$. Если оно выполняется, то алгоритм завершает работу, в противном случае - переход к пункту 3.

3. Вычисляются векторы $\mathbf{y}^{(k+1)}$ и $\mathbf{w}^{(k+1)}$ по формулам (1) и (2), наращивается k .
Переход к пункту 2.

Замечание.

Если $(V\mathbf{w}^{(k)} - \mathbf{y}^{(k)})^+ = \mathbf{0}$, но $V\mathbf{w}^{(k)} - \mathbf{y}^{(k)} \neq \mathbf{0}$, то классы не являются линейно разделимыми.

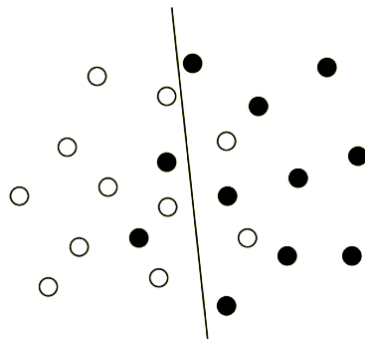
Если классы являются линейно разделимыми, то алгоритм вычисления весового вектора \mathbf{w} по формулам (2.3) и (2.4) сходится. Такой подход к нахождению решающей функции называется алгоритмом наименьшей среднеквадратичной ошибки (НСКО-алгоритмом) или алгоритмом Хо-Кашьяпа.

$$\mathbf{y}^{(k+1)} = \mathbf{y}^{(k)} - h_k \text{grad}_{\mathbf{y}}(F(\mathbf{w}^{(k)}, \mathbf{y}^{(k)})) = \mathbf{y}^{(k)} + h_k (V\mathbf{w}^{(k)} - \mathbf{y}^{(k)})^+, \quad k = 1, 2, \dots, \quad (2.3)$$

$$\mathbf{w}^{(k+1)} = V^{\otimes} \mathbf{y}^{(k+1)}, \quad k = 1, 2, \dots \quad (2.4)$$

Обобщенная постановка задачи линейного разделения классов:

Найти такую ЛРФ, чтобы ошибка неправильной классификации была минимальной.



В случае, когда классы не являются линейно разделимыми, можно поставить другую задачу: найти такую линейную решающую функцию, чтобы ошибка неправильной классификации была минимальной. Такой подход реализует важное свойство хорошей системы распознавания – свойство обобщения, которое предполагает, что система распознавания должна уметь классифицировать элементы «похожие» на элементы обучающей выборки.

Обобщенные решающие функции (ОРФ)

Если классы нельзя разделить с помощью ЛРФ в пространстве R^n , то следует вложить эти классы в пространство R^l большей размерности $l > n$ с помощью

некоторого отображения $\varphi: R^n \rightarrow R^l$. Причем это отображение должно быть таким, чтобы классы в R^l можно было линейно разделить.

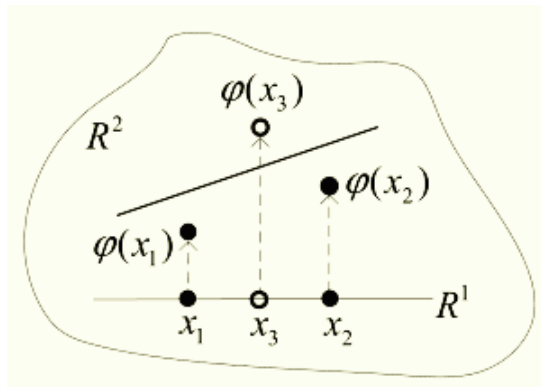
Пусть классы $\Xi_1 = \{x_i : i \in I_1\}$, $\Xi_2 = \{x_i : i \in I_2\}$ линейно неразделимы в R^n .

Задача: найти такое отображение $\varphi: R^n \rightarrow R^l$, $l > n$, чтобы классы $\tilde{\Xi}_1 = \{\varphi(x_i) : i \in I_1\}$ и $\tilde{\Xi}_2 = \{\varphi(x_i) : i \in I_2\}$ были линейно разделимы в R^l

Пример.

$\Xi_1 = \{x_1, x_2\}$, $\Xi_2 = \{x_3\}$ - линейно неразделимы в R^1 , а

$\tilde{\Xi}_1 = \{\varphi(x_1), \varphi(x_2)\}$, $\tilde{\Xi}_2 = \{\varphi(x_3)\}$ - линейно разделимы в R^2 .



R^l - спрямляющее пространство,

φ - спрямляющее отображение.

Для построения φ используют ОРФ вида

$$d(\mathbf{x}) = w_1 f_1(\mathbf{x}) + \dots + w_{l-1} f_{l-1}(\mathbf{x}) + w_l,$$

где $f_i(\mathbf{x})$ - скалярные функции в R^n .

Тогда спрямляющее отображение φ имеет вид

$$\varphi: \mathbf{x} \rightarrow (f_1(\mathbf{x}), \dots, f_l(\mathbf{x}))^T = \mathbf{x}^* \in R^l, \mathbf{x}^* = (x_1^*, \dots, x_l^*)^T,$$

где $f_l(\mathbf{x}) \equiv 1$ и

$$d(\mathbf{x}^*) = w_1 x_1^* + \dots + w_{l-1} x_{l-1}^* + w_l = (\mathbf{w}, \mathbf{x}^*).$$

Важный класс ОРФ - *мономиальные* функции вида

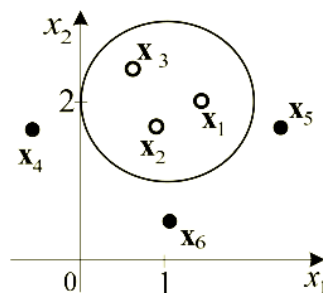
$$f_{s_1, \dots, s_n}(x_1, \dots, x_n) = x_1^{s_1} \cdot \dots \cdot x_n^{s_n}.$$

Пример. Классы $\Xi_1 = \{x_1, x_2, x_3\}$, $\Xi_2 = \{x_4, x_5, x_6\}$ линейно неразделимы, но разделимы с помощью ОРФ

$$d(\mathbf{x}) = (x_1 - 1)^2 + (x_2 - 2)^2 - 1 = x_1^2 + x_2^2 - 2x_1 - 4x_2 + 4$$

или

$$d(\mathbf{x}^*) = 1 \cdot x_1^* + 1 \cdot x_2^* + 0 \cdot x_3^* - 2 \cdot x_4^* - 4 \cdot x_5^* + 4 \cdot 1.$$



Задача понижения размерности

Задачу понижения размерности можно рассматривать как задачу выбора наиболее информативных признаков образов. Точнее, по заданной выборке векторов, принадлежащих разным линейно разделимым классам, требуется найти такое подпространство R^p исходного пространства R^n $p < n$, чтобы после вычисления проекций $\mathbf{x}' = pr\mathbf{x} \in R^p$ векторов $\mathbf{x} \in R^n$ на это подпространство проекции классов оставались линейно разделимыми.

Метод главных компонент. Корреляционный подход (Principle Component Analysis - PCA)

Постановка задачи.

Дана обучающая выборка $\Xi = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \subseteq R^n$, $\sum_{i=1}^N \mathbf{x}_i = \mathbf{0}$.

Требуется найти такое подпространство R^p , $p < n$, чтобы векторы выборки

$\Xi' = \{\mathbf{x}'_1, \dots, \mathbf{x}'_N\} \subseteq R^p$, $\mathbf{x}'_i = pr\mathbf{x}_i$ были слабо коррелированными между собой.

Автокорреляционная матрица

$$R(\mathbf{x}) = \mathbf{x}\mathbf{x}^T$$

характеризует корреляцию между признаками вектора \mathbf{x} (если $R(\mathbf{x})$ - диагональная матрица, то корреляции между признаками нет).

Автокорреляционная матрица векторов-образов в классе ϖ_i

$$R_i = \frac{1}{|\varpi_i|} \sum_{\mathbf{x} \in \varpi_i} R(\mathbf{x}),$$

Автокорреляционная матрица всей выборки

$$R = \frac{1}{m} \sum_{i=1}^m R_i,$$

где m - число классов.

Задача: требуется найти такое подпространство, чтобы автокорреляционная матрица R' проекций векторов-образов на него была диагональной.

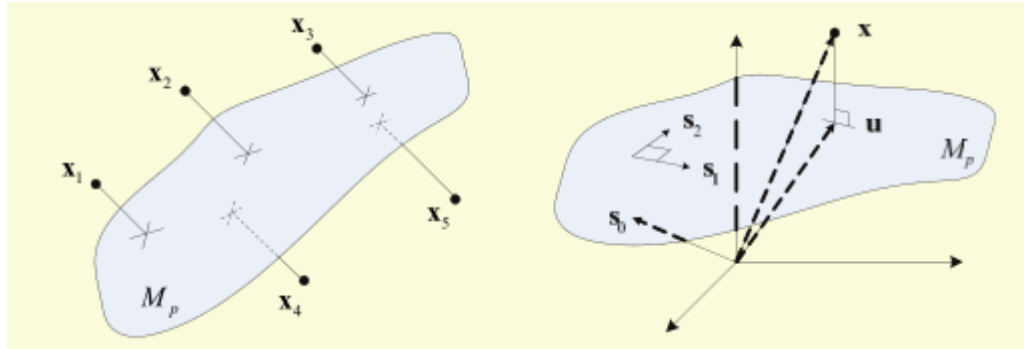
Метод главных компонент. Алгебраический подход

Постановка задачи.

Пусть $\Xi = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ - обучающая выборка векторов в R^n . Необходимо найти p -мерное линейное многообразие M_p ($p < n$), которое минимизирует сумму квадратов расстояний от всех точек множества Ξ до многообразия M_p , т.е.

$$Q(M_p) = \sum_{i=1}^N d^2(\mathbf{x}_i, M_p) \rightarrow \min,$$

где $d(\mathbf{x}, M_p)$ - расстояние от точки $\mathbf{x} \in R^n$ до многообразия M_p .



Векторы линейного p -мерного многообразия M_p в R^n имеют вид

$$\mathbf{u} = \mathbf{s}_0 + \sum_{j=1}^p \alpha_j \mathbf{s}_j, \alpha_j \in R, \quad (*)$$

где $\mathbf{s}_0 \in R^n$ - вектор сдвига,

$\{\mathbf{s}_j\}_{j=1}^p$ - ортонормированная система векторов в R^n

Тогда

$$d(\mathbf{x}, M_p) = d(\mathbf{x}, \mathbf{u}) = \|\mathbf{x} - \mathbf{u}\|,$$

где $\mathbf{u} \in M_p$: $\mathbf{x} - \mathbf{u} \perp \mathbf{s}_j \Leftrightarrow$

$$0 = (\mathbf{s}_j, \mathbf{x} - \mathbf{u}) = (\mathbf{s}_j, \mathbf{x} - \mathbf{s}_0 - \sum_{k=1}^p \alpha_k \mathbf{s}_k) = (\mathbf{s}_j, \mathbf{x} - \mathbf{s}_0) - \alpha_j$$

$$\Rightarrow \alpha_j = (\mathbf{s}_j, \mathbf{x} - \mathbf{s}_0),$$

и

$$d^2(\mathbf{x}, M_p) = \|\mathbf{x} - \mathbf{u}\|^2 = \left\| \mathbf{x} - \mathbf{s}_0 - \sum_{j=1}^p (\mathbf{s}_j, \mathbf{x} - \mathbf{s}_0) \mathbf{s}_j \right\|^2 =$$

$$\|\mathbf{x} - \mathbf{s}_0\|^2 - 2 \sum_{j=1}^p (\mathbf{s}_j, \mathbf{x} - \mathbf{s}_0)^2 + \sum_{j=1}^p \|\mathbf{s}_j\|^2 (\mathbf{s}_j, \mathbf{x} - \mathbf{s}_0)^2 =$$

$$\|\mathbf{x} - \mathbf{s}_0\|^2 - \sum_{j=1}^p (\mathbf{s}_j, \mathbf{x} - \mathbf{s}_0)^2.$$

Линейный дискриминант Фишера

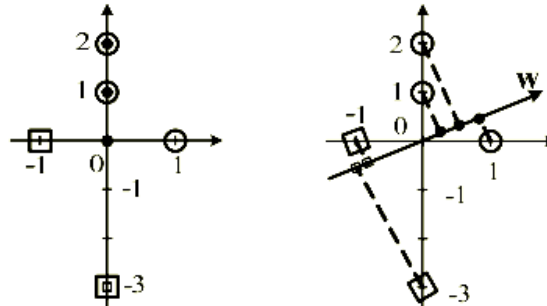
Постановка задачи

Найти такое подпространство, чтобы:

1) расстояние между средними значениями проекций классов было максимальным;

2) полный разброс спроецированных выборочных значений был минимальным.

Прямая проецирования, удовлетворяющая этим требованиям, называется *линейным дискриминантом Фишера*



Пусть \mathbf{w} , $\|\mathbf{w}\| = 1$, - направляющий вектор прямой проецирования. Тогда $\mathbf{x}' = \mathbf{w}^T \mathbf{x}$ - проекция точки на прямую. Для нахождения \mathbf{w} Р. Фишер предложил использовать следующий критерий

$$f(\mathbf{w}) = \frac{|\mathbf{m}'_1 - \mathbf{m}'_2|^2}{s_1'^2 + s_2'^2} \rightarrow \max,$$

где $\mathbf{m}'_i = \frac{1}{|\omega_i|} \sum_{\mathbf{x} \in \omega_i} \mathbf{x}'$ - выборочные математические ожидания проекций векторов i -го класса, $i = 1, 2$,

$s_i'^2 = \sum_{\mathbf{x} \in \omega_i} (\mathbf{x}' - \mathbf{m}'_i)^2$ - разброс спроецированных выборочных значений внутри i -го класса, $i = 1, 2$.

Так как

$$\mathbf{m}'_i = \frac{1}{|\omega_i|} \sum_{\mathbf{x} \in \omega_i} \mathbf{x}' = \frac{1}{|\omega_i|} \sum_{\mathbf{x} \in \omega_i} \mathbf{w}^T \mathbf{x} = \mathbf{w}^T \mathbf{m}_i, \quad i = 1, 2,$$

то

$$|\mathbf{m}'_1 - \mathbf{m}'_2|^2 = |\mathbf{w}^T (\mathbf{m}_1 - \mathbf{m}_2)|^2 = \mathbf{w}^T (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T \mathbf{w} = \mathbf{w}^T S_m \mathbf{w},$$

где $S_m = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T$ - матрица разброса между классами.

Средние выборочные дисперсии проекций векторов в классах будут (с точностью до нормирующего множителя) равны. Средние выборочные дисперсии проекций векторов в классах будут (с точностью до нормирующего множителя) равны

Аналогично,

$$s_i'^2 = \sum_{\mathbf{x} \in \omega_i} (\mathbf{x}' - \mathbf{m}'_i)^2 = \sum_{\mathbf{x} \in \omega_i} (\mathbf{w}^T (\mathbf{x} - \mathbf{m}_i))^2 = \sum_{\mathbf{x} \in \omega_i} \mathbf{w}^T (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T \mathbf{w} = \mathbf{w}^T S_i \mathbf{w},$$

где $S_i = \sum_{\mathbf{x} \in \omega_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T$, $i = 1, 2$, - матрицы разброса векторов в классах.

Пусть $S = S_1 + S_2$ - матрица разброса векторов всей выборки.

Тогда $s_1'^2 + s_2'^2 = \mathbf{w}^T S \mathbf{w}$ и функция критерия примет вид

$$f(\mathbf{w}) = \frac{\mathbf{w}^T S_m \mathbf{w}}{\mathbf{w}^T S \mathbf{w}}.$$

Пример. Пусть заданы двумерные образы - векторы $\mathbf{x}_1 = (0, 2)^T$, $\mathbf{x}_2 = (0, 1)^T$, $\mathbf{x}_3 = (1, 0)^T \in X_1$ и $\mathbf{x}_4 = (-1, 0)^T$, $\mathbf{x}_5 = (0, -3)^T \in X_2$ (рис. 2.12), принадлежащие областям предпочтения X_1 и X_2 двух классов.

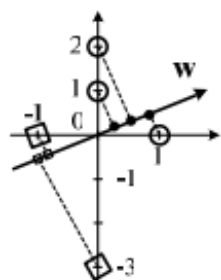


Рис. 2.13

Тогда $\mathbf{m}_1 = (1/3, 1)^T$, $\mathbf{m}_2 = (-1/2, -3/2)^T$ и

$$S_1 = \frac{1}{3} \begin{pmatrix} 2 & -3 \\ -3 & 6 \end{pmatrix}, \quad S_2 = \frac{1}{2} \begin{pmatrix} 1 & -3 \\ -3 & 9 \end{pmatrix}, \quad S = S_1 + S_2 = \frac{1}{6} \begin{pmatrix} 7 & -15 \\ -15 & 39 \end{pmatrix},$$

$$\mathbf{w} = S^{-1}(\mathbf{m}_1 - \mathbf{m}_2) = \frac{1}{8} \begin{pmatrix} 39 & 15 \\ 15 & 7 \end{pmatrix} \cdot \frac{5}{6} \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \frac{5}{4} \begin{pmatrix} 7 \\ 3 \end{pmatrix}.$$

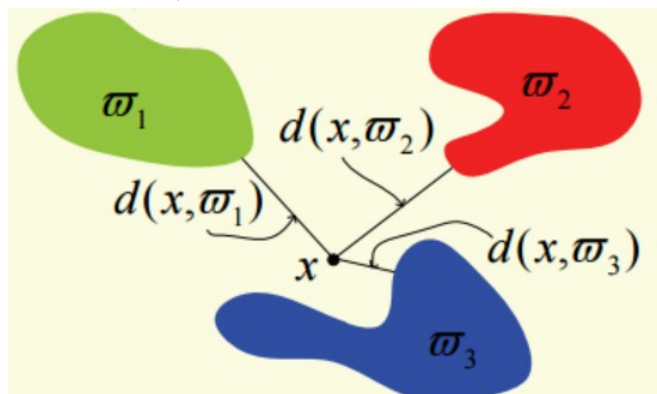
Следовательно, $\mathbf{x}'_1 = 15/2$, $\mathbf{x}'_2 = 15/4$, $\mathbf{x}'_3 = 35/4$, $\mathbf{x}'_4 = -35/4$, $\mathbf{x}'_5 = -45/4$. Видно, что в этом случае проекции векторов классов будут лучше разделены и локализованы (рис. 2.13).

Классификация с помощью функций расстояния

- постановка задачи классификации с помощью функции расстояния;
- стандартизация признаков;
- способы измерения расстояний между векторами признаков;
- способы определения расстояния между вектором-образом и классом;
- способы определения расстояний между классами.

Постановка задачи

Найти такую функцию $d(x, \varpi_i)$, что $x \in \varpi_i$, если $d(x, \varpi_i) \leq d(x, \varpi_j)$ для всех $j \neq i$



$d(x, \varpi_i)$ - функция расстояния.

Меры близости и аксиомы метрики

Мера близости между:

- двумя образами $d(x, y)$,
- образом и классом $d(x, \varpi)$
- двумя классами $d(\varpi_i, \varpi_j)$

Способы стандартизации признаков

Дана выборка $\Xi = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ векторов $\mathbf{x}_i = (x_{i1}, \dots, x_{im})^T$
 x_{ik} - признаки (гены).

Задача: привести все признаки к единому масштабу.

Способы стандартизации признаков:

1) $x_{ik} \rightarrow \frac{x_{ik} - \tilde{m}_i}{\tilde{\sigma}_i}$, где $\tilde{m}_i = \frac{1}{N} \sum_{k=1}^N x_{ik}$ - среднее выборочное значение i -й координаты,

$\tilde{\sigma}_i = \sqrt{\frac{1}{N} \sum_{k=1}^N (x_{ik} - \tilde{m}_i)^2}$ - выборочное среднеквадратичное отклонение;

2) $x_{ik} \rightarrow \frac{x_{ik} - \min_k x_{ik}}{\max_k x_{ik} - \min_k x_{ik}}$.

Способы измерения расстояний между векторами признаков

а) метрика Евклида

$$d_2(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2 = \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2};$$

б) манхаттановская метрика

$$d_1(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_1 = |x_1 - y_1| + \dots + |x_n - y_n|$$

(если $x_i \in \{-1, 1\}$, то d_1 - метрика Хэмминга);

в) равномерная метрика

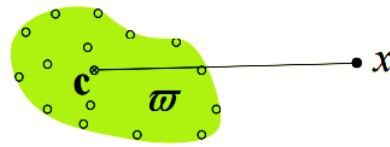
$$d_\infty(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_\infty = \max_{1 \leq i \leq n} |x_i - y_i|;$$

г) метрика Минковского

$$d_p(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_p = \sqrt[p]{|x_1 - y_1|^p + \dots + |x_n - y_n|^p}, \quad (p \geq 1)$$

Способы определения расстояния между вектором-образом и классом

1. Определение расстояния до центра класса



$$d(x, \varpi) = d(\mathbf{x}, \mathbf{c})$$

Правило классификации:

$$x \in \varpi_i, \text{ если } d(\mathbf{x}, \mathbf{c}_i) < d(\mathbf{x}, \mathbf{c}_j) \quad \forall j \neq i, \quad \mathbf{c}_i = \frac{1}{|\varpi_i|} \sum_{\mathbf{x} \in \varpi_i} \mathbf{x}$$

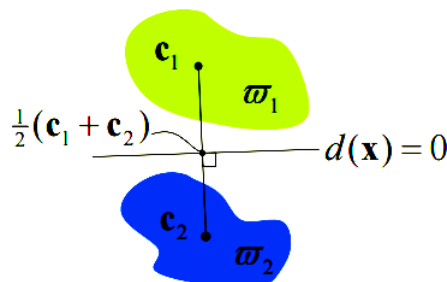
Применимость:

- 1) элементы класса «компактно» расположены в пространстве;
- 2) цена ошибки неправильной классификации невелика.

Этот способ применяется в том случае, когда класс «хорошо описывается» одним эталонным образом – центром класса (качество такого описания может быть измерено величиной дисперсии элементов класса), а цена ошибки неправильной классификации не очень велика.

Клетки Вороного

а) классификация по двум классам (евклидова метрика):



Решающая функция

$$d(\mathbf{x}) = \|\mathbf{x} - \mathbf{c}_1\|_2^2 - \|\mathbf{x} - \mathbf{c}_2\|_2^2 = (\mathbf{x} - \mathbf{c}_1, \mathbf{x} - \mathbf{c}_1) - (\mathbf{x} - \mathbf{c}_2, \mathbf{x} - \mathbf{c}_2) = \mathbf{x}^2 - 2\mathbf{c}_1 \cdot \mathbf{x} + \mathbf{c}_1^2 - \mathbf{x}^2 + 2\mathbf{c}_2 \cdot \mathbf{x} - \mathbf{c}_2^2 = 2(\mathbf{c}_2 - \mathbf{c}_1) \cdot \left(\mathbf{x} - \frac{1}{2}(\mathbf{c}_1 + \mathbf{c}_2)\right).$$

В общем случае, с помощью функции расстояния все пространство признаков с заданными в нем центрами классов $\mathbf{c}_1, \dots, \mathbf{c}_m$ разбивается на отдельные области X_1, \dots, X_m такие, что $d(\mathbf{x}, \mathbf{c}_j) < d(\mathbf{x}, \mathbf{c}_i)$ для любой точки $\mathbf{x} \in X_j$ и всех $i \neq j$. Такие области называются *клетками Вороного* а множество всех клеток Вороного - *диаграммой Вороного*. В евклидовой метрике границами клеток Вороного всегда являются некоторые выпуклые многогранники.

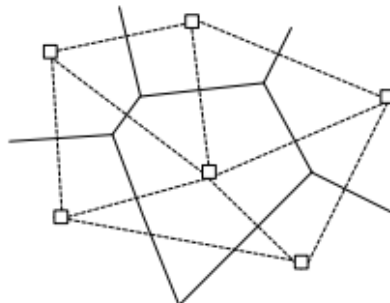
Диаграмму Вороного в евклидовой метрике на плоскости можно построить с помощью процедуры, известной в вычислительной геометрии, как триангуляция Делоне. Триангуляция Делоне – это планарный граф, **удовлетворяющий условиям:**

- 1) все его внутренние области являются треугольниками;

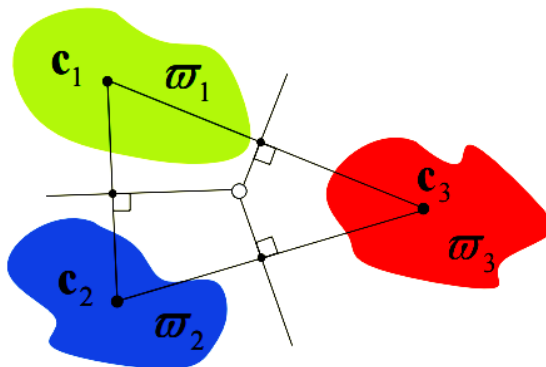
2) минимальный многоугольник, охватывающий все треугольники, является выпуклым;

3) внутрь окружности, описанной вокруг любого треугольника, не попадает ни одна точка триангуляции.

По диаграмме Вороного легко строится триангуляция Делоне и наоборот:

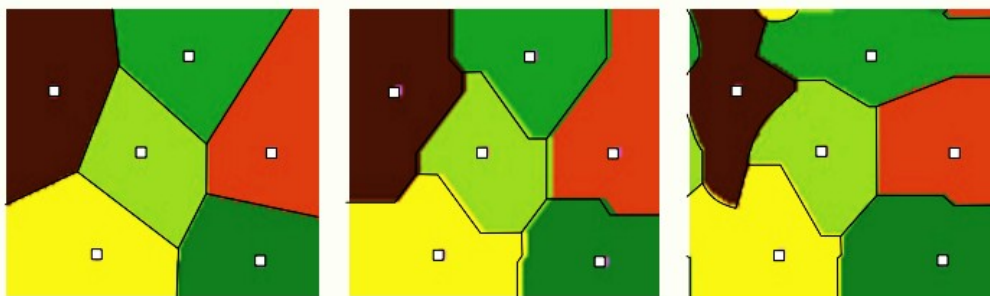


б) классификация по трем классам (евклидова метрика):



Пусть c_1, \dots, c_m - центры классов в R^n ,

$X_i = \{x \in R^n : d(x, c_i) < d(x, c_j) \forall j \neq i\}, i = 1, \dots, m$ - клетки Вороного.

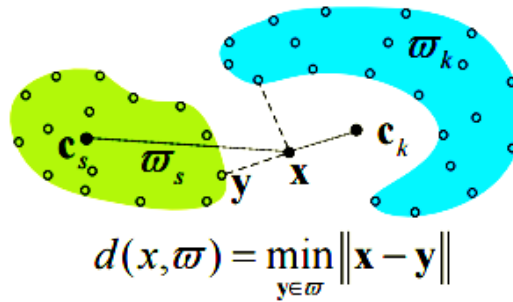


метрика Евклида

метрика Хэмминга

метрика Канберра

2. Метод ближайшего соседа



Правило классификации:

$x \in \varpi_s$, если $\|x - x_s\| = \min \{\|x - x_i\| : i = 1, \dots, N\}$ и $x_s \in \varpi_s$.

Применимость:

- 1) ошибки в данных невелики;
- 2) цена ошибки неправильной классификации - велика.

Алгоритмы кластеризации

- дескриптивная постановка задачи кластеризации;
- основные цели кластеризации;
- математическая постановка задачи кластеризации;
- основные критерии качества кластеризации;
- алгоритм k -внутригрупповых средних;
- теорема о сходимости алгоритма k -means;
- алгоритмы расстановки центров кластеров;
- алгоритм FOREL;
- алгоритм ИСОМАД

Постановка задачи

Задача. Найти такое разбиение обучающей выборки $\Xi = \{x_1, \dots, x_N\}$ на непересекающиеся подмножества (кластеры) X_1, \dots, X_m : $X_1 \cup \dots \cup X_m = \Xi$, $X_i \cap X_j = \emptyset \forall i \neq j$, чтобы все точки одного кластера состояли из «похожих» элементов, а точки разных кластеров существенно отличались.

Основные параметры кластеризации:

- 1) критерий «похожести» элементов Q ;
- 2) используемая метрика d ;
- 3) число кластеров.

Основные цели кластеризации

1) нахождение групп схожих элементов с целью дальнейшей независимой их обработки. Параметры кластеризации должны обеспечивать минимальность числа кластеров;

2) получение выборки эталонных элементов - типичных представителей кластеров. Параметры кластеризации должны обеспечивать формирование однородных кластеров;

3) нахождение элементов, не попадающих ни в один из кластеров, при этом сами кластеры должны быть небольшими;

4) формирование иерархической структуры выборки (задача таксономии).

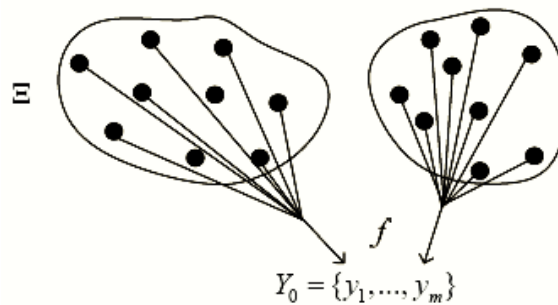
Математическая постановка задачи кластеризации

Пусть $\Xi = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ обучающая выборка,

Ψ - множества допустимых меток (определяется целями кластеризации). Найти множества меток $Y_0 \in \Psi$ и функцию $f: \Xi \rightarrow Y_0$, чтобы

$$Y_0 = \arg \min_{Y \in \Psi, f} Q(Y, f),$$

где $Q(Y, f)$ - выбранный критерий качества кластеризации.



Алгоритм k -внутригрупповых средних (k -means)

Пусть m - фиксированное число кластеров. Найти такую функцию кластеризации $f: \Xi \rightarrow Y$, $|Y| = m$, чтобы $Q^{(3)}(f) \rightarrow \min$.

Алгоритм k -means

1. Назначаются начальные центры кластеров $\mathbf{c}_1^{(0)}, \dots, \mathbf{c}_m^{(0)}$ и полагается $k = 0$.
2. Выборка Ξ разбивается на m кластеров по методу ближайшего соседа - получим кластеры $X_1^{(k)}, \dots, X_m^{(k)}$

3. Рассчитываются новые центры кластеров по формуле:

$$\mathbf{c}_i^{(k+1)} = \frac{1}{|X_i^{(k)}|} \sum_{\mathbf{x} \in X_i^{(k)}} \mathbf{x}.$$

4. Если $\mathbf{c}_i^{(k+1)} = \mathbf{c}_i^{(k)} \forall i$, то - останов, иначе - к п.2.

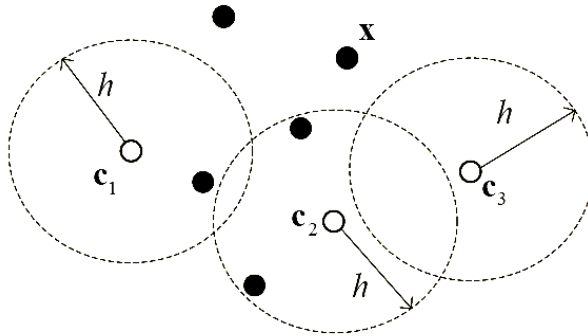
Алгоритмы расстановки центров кластеров

1. Алгоритм простейшей расстановки центров кластеров

Вводится некоторый порог $h > 0$.

1) в качестве первого центра кластера назначается первый элемент выборки $\mathbf{c}_1 = \mathbf{x}_1$

2) если уже выбраны k центров кластеров, то в качестве $k+1$ -го центра выбирается такой элемент выборки \mathbf{x}_j , что минимальное расстояние от \mathbf{x}_j до центров $\mathbf{c}_i, i = 1, \dots, k$, будет больше h .



2) Алгоритм, основанный на методе просеивания

Пусть $f(\mathbf{x})$ - плотность распределения элементов выборки.

Свойства:

1) $f(\mathbf{x}) \geq 0$;

2) $f(\mathbf{x})$ принимает тем большее значение, чем \mathbf{x} ближе к точке сгущения элементов выборки. Например,

$$f(\mathbf{x}) = f_h(\mathbf{x}) = \frac{1}{h^2} \sum_{i: \|\mathbf{x} - \mathbf{x}_i\| < h} (h^2 - \|\mathbf{x} - \mathbf{x}_i\|^2), \quad h > 0.$$

Упорядочим элементы выборки по убыванию $f(\mathbf{x}_1) \geq f(\mathbf{x}_2) \geq f(\mathbf{x}_3) \geq \dots$

Осуществим простейшую расстановку центров кластеров, выбирая в качестве новых центров те элементы выборки, в которых значение плотности будет наибольшим.

Максиминный алгоритм

1. В качестве первого центра кластера выбирается элемент $c_1 = x_1$.
2. В качестве второго центра кластера выбирается тот элемент $c_2 = x_{j_2}$, который находится на наибольшем расстоянии от c_1 , т.е. $\|x_{j_2} - c_1\| = \max_{x \in \Xi} \|x - c_1\|$.

3. Предположим, что выбраны k центров $C^{(k)} = \{c_1, \dots, c_k\}$ кластеров. В качестве очередного $(k+1)$ -го центра кластера выбирается тот элемент $x_{j_{k+1}}$, который находится на наибольшем расстоянии от ближайшего из центров c_1, \dots, c_k (рис. 4.2), т.е.

$$\min_{c \in C^{(k)}} \|x_{j_{k+1}} - c\| = \max_{x \in \Xi \setminus C^{(k)}} \min_{c \in C^{(k)}} \|x - c\|.$$

4. Проверяется условие останова.

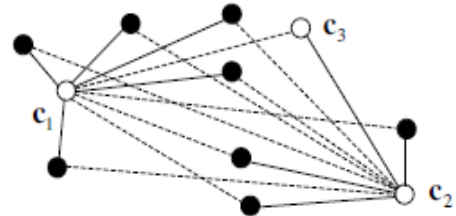


Рис. 4.2

Алгоритм FOREL (FORmal ELeMent)

Предложен Н.Г. Загоруйко и В.Н. Ёлкиной в 1967г. задается параметр $r > 0$ - радиус гипершара $B_r(e)$.

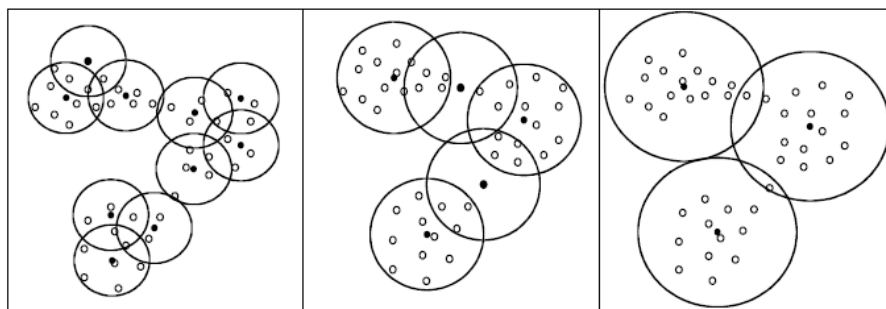
Алгоритм

1. Выбирается некоторый элемент $x \in \Xi$ обучающей выборки, $e^{(1)} := x$.
2. Вычисляется центр тяжести

$$e^{(2)} = \frac{1}{|B_r(e^{(1)}) \cap \Xi|} \sum_{x \in B_r(e^{(1)}) \cap \Xi} x$$

Выполнение пункта 2 повторяется до тех пор, пока последовательность $e^{(1)}, e^{(2)}, \dots$ не стабилизируется в точке e .

3. Полагаем $\Xi := \Xi \setminus B_r(e)$ и переходим к пункту 1. Условием останова алгоритма является $\Xi = \emptyset$.



Алгоритм ИСОМАД (ISODATA)

Итеративный Самоорганизующийся Метод Анализа Данных (ISODATA - Iterative Self-Organizing Data Analysis Techniques), разработан в 1965 году Бэллом (Ball G.) и Хэллом (Hall D.)

Основные процедуры изменения числа кластеров.

1. Удаление кластеров. Если кластер содержит мало элементов $|X_i| < q_1$ то он удаляется, его элементы распределяются по другим кластерам, а центр кластера c_i удаляется из списка центров кластеров.

2. Разделение кластеров. Если дисперсия i -го мастера $D_i > q_3$, то i -й кластер разделяют на два. Для этого вычисляется «направление» в R^n , вдоль которого дисперсия кластера максимальна. Кластер разделяется на два гиперплоскостью, проходящей через центр кластера и перпендикулярной вычисленному направлению.

3. Слияние кластеров. Пусть $l_{ij} = \|c_i - c_j\|$ - расстояние между центрами кластеров. Если $l_{ij} < q_3$, то кластеры X_i и X_j объединяются. Новый центр кластера вычисляется по формуле:

$$c = \frac{1}{|X_i| + |X_j|} (c_i |X_i| + c_j |X_j|).$$

Метод опорных векторов

Метод SVM (Support Vector Machine) появился в ряде работ Владимира Вапника и др. в 60-80-е годы.

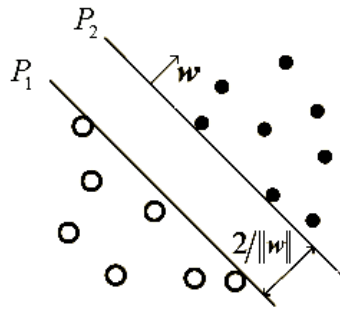
Вапник Владимир Наумович (р. 1935) - российский математик. Работал в ИГТУ РАН, с 1990 г. живет и работает в Англии. Работы по прикладной статистике, автор статистической теории обучения.

- Линейно разделимый случай;
- Линейно неразделимый случай;
- SVM, минимизирующий ошибку неправильной классификации

Линейно разделимый случай

Постановка задачи.

Имеется множество прецедентов (Ξ, Y) , где $\Xi = \{x_1, \dots, x_N\}$ - обучающая выборка, $x \in R^n$, а $Y = (y_1, \dots, y_N)$ - множество меток двух классов ω_1 и ω_2 . Известно, что ω_1 и ω_2 линейно разделимы. Требуется найти две такие параллельные разделяющие гиперплоскости P_1 и P_2 , что $d(P_1, P_2) \rightarrow \max$.



Линейно неразделимый случай

В 1992 году Бернард Бозер (Boser B.), Изабелл Гийон (Guyon I.) и Владимир Вапник адаптировали SVM для нелинейного разделения классов. Для этого рассмотрим вложение пространства признаков R^n в пространство H большей размерности с помощью отображения $\varphi : R^n \rightarrow H$. Разделяющую функцию будем искать в виде

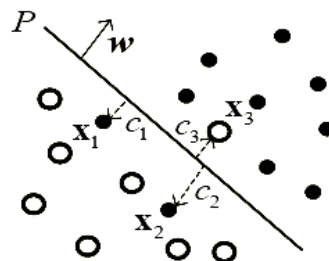
$$f(\mathbf{x}) = (\mathbf{w}, \varphi(\mathbf{x})) + b, \quad \mathbf{w} = \sum_{i=1}^N \lambda_i y_i \varphi(\mathbf{x}_i),$$

где коэффициенты λ_i зависят от y_i и от значения $(\varphi(\mathbf{x}_i), \varphi(\mathbf{x}_j))$

SVM, минимизирующий ошибку неправильной классификации

Постановка задачи.

Имеется множество прецедентов (Ξ, Y) , где $\Xi = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ - обучающая выборка, $\mathbf{x}_i \in R^n$, а $Y = (y_1, \dots, y_N)$ - множество меток двух классов ω_1 и ω_2 . Известно, что ω_1 и ω_2 линейно неразделимы. Требуется найти такую гиперплоскость P , что сумма расстояний от P до неправильно классифицируемых точек не превосходила заданного значения K , а расстояние до ближайшей правильно классифицируемой точки было максимальным.



Математическая постановка.

Требуется найти такой вектор \mathbf{w} и число b , чтобы $1/2\|\mathbf{w}\|^2 \rightarrow \min$ и выполнялись неравенства

$$y_i((\mathbf{w}, \mathbf{x}_i) + b) \geq 1 - c_i, c_i \geq 0, i = 1, \dots, N, \sum_{i=1}^N c_i \leq K.$$

При $K = 0$ - задача линейного разделения;

K пропорциональна сумме расстояний от P до неправильно классифицируемых точек;

i -я точка классифицируется неправильно $\Leftrightarrow c_i > 1$.

Равносильная постановка. Требуется найти такой вектор \mathbf{w} и число b , чтобы

$$\frac{1}{2} \|\mathbf{w}\|^2 + \frac{1}{K} \sum_{i=1}^N c_i \rightarrow \min$$

и выполнялись неравенства

$$y_i((\mathbf{w}, \mathbf{x}_i) + b) \geq 1 - c_i, c_i \geq 0, i = 1, \dots, N$$

Преимущества и недостатки SVM

Это наиболее быстрый метод нахождения решающих функций;

- метод находит разделяющую полосу максимальной ширины, что позволяет в дальнейшем осуществлять более уверенную классификацию;
- метод чувствителен к шумам и стандартизации данных;
- не существует общего подхода к автоматическому выбору ядра (и построению спрямляющего подпространства в целом) в случае линейной неразделимости классов.

Лекция 6 «Нейронные сети и проблемы распознавания».

Термин «нейронная сеть» появился в середине XX века. Первые работы, в которых были получены основные результаты в данном направлении, были проделаны Уорреном Мак-Каллоком (1898-1969) и Уолтером Питтсом (1923-1969).

В 1943 году ими была разработана компьютерная модель нейронной сети на основе математических алгоритмов и теории деятельности головного мозга. Они выдвинули предположение, что нейроны можно упрощённо рассматривать как устройства, оперирующие двоичными числами, и назвали эту модель «пороговой логикой».

Данная модель заложила основы двух различных подходов исследований нейронных сетей.

Один подход был ориентирован - на изучение биологических процессов в головном мозге, другой – на применение нейронных сетей как метода искусственного интеллекта для решения различных прикладных задач.

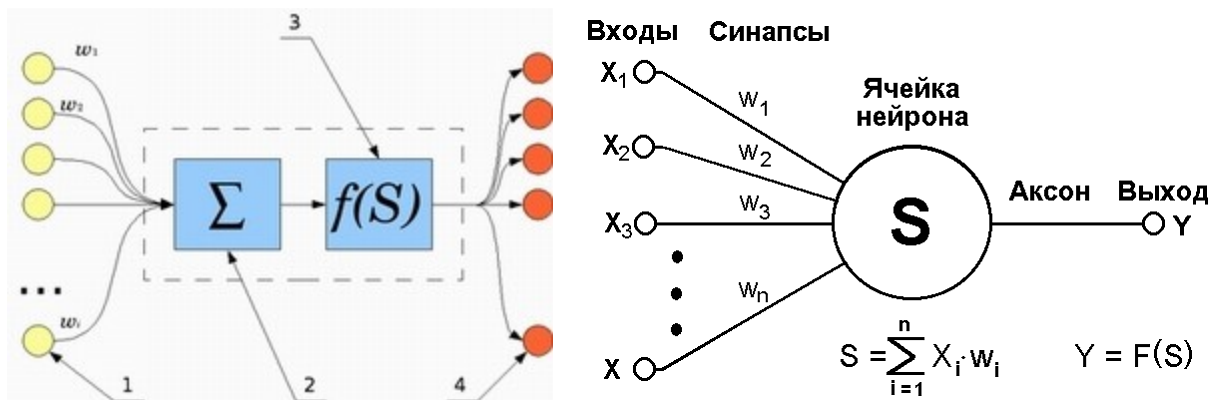


Рис. 6.1 – Схема искусственного нейрона:

- 1 - нейроны, выходные сигналы которых поступают на вход данному;
- 2 - сумматор входных сигналов;
- 3 - вычислитель передаточной функции;
- 4 - нейроны, на входы которых подаётся выходной сигнал данного нейрона;
- 5 - w_i — веса входных сигналов или силы синаптических связей с пресинаптическими нейронами

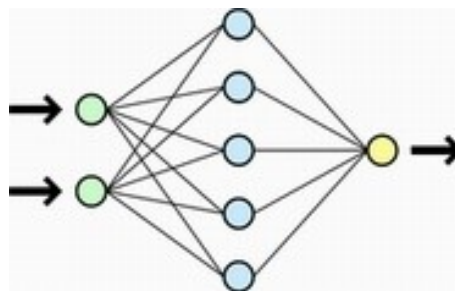


Рис. 6.2 – Искусственная нейронная сеть, состоящая из нейронов Мак-Каллока - Питтса

В 1949 году канадский физиолог и психолог Дональд Хебб (1904-1985) высказал идеи о характере соединения нейронов мозга и их взаимодействии. Он первым предположил, что обучение заключается в первую очередь в изменениях силы синаптических связей. Теория Хебба считается типичным случаем самообучения, при котором испытуемая система спонтанно обучается выполнять поставленную задачу без вмешательства со стороны экспериментатора.

В немного вольной трактовке правило обучения Хебба. имеет очень простой смысл: связи нейронов, активирующихся совместно, должны усиливаться, связи нейронов, срабатывающих независимо, должны ослабевать.

В 1954 году в Массачусетском технологическом институте с использованием компьютеров Фарли и Кларк разработали имитацию сети Хебба.

Состояние выхода линейного сумматора можно записать:

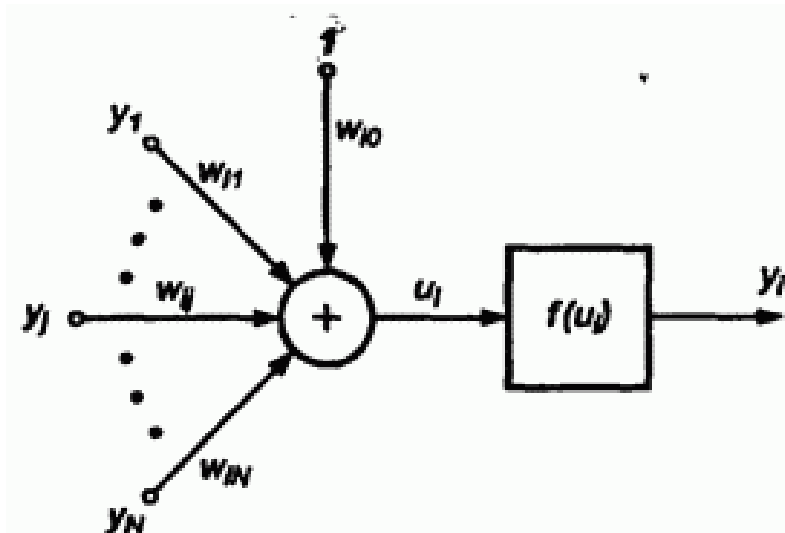


Рис. 6.3 – Модель линейного нейрона Хебба

В основе многих нейросетевых моделей лежит правило обучения Хебба: связи нейронов, активирующихся совместно, должны усиливаться, связи нейронов, срабатывающих независимо, должны ослабевать.

$$y = \sum_{i=1}^m \omega_i x_i$$

В 1957 году Фрэнком Розенблаттом (1928—1971) были разработаны математическая и компьютерная модели восприятия информации мозгом на основе двухслойной обучающейся нейронной сети. При обучении данная сеть использовала арифметические действия сложения и вычитания. Розенблатт описал также схему не только основного перцептрона, но и схему логического сложения.

В 1958 году им была предложена модель электронного устройства, которое должно было имитировать процессы человеческого мышления.

1960 г - первая действующая машина - электронная машина «Марк-1», которая могла научиться распознавать некоторые из букв, написанных на карточках, которые подносили к его «глазам», напоминающим кинокамеры.

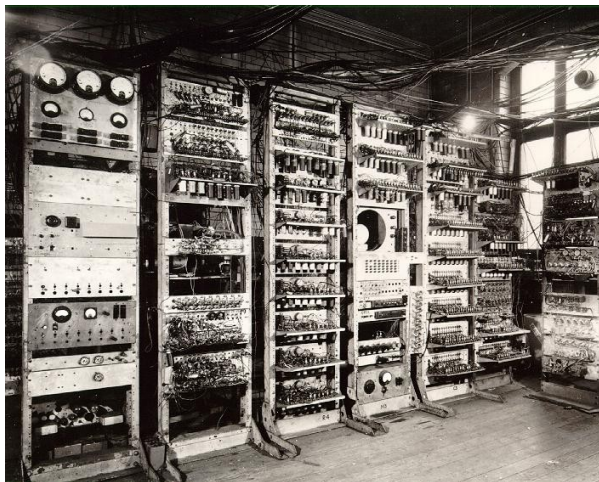


Рис. 6.4 –Манчестерский Марк I

Марк-1 – один из первых полностью электронных компьютеров с хранимой в оперативной памяти программой, тем самым практически реализующий архитектуру фон Неймана. Проект был санкционирован Министерством обороны Великобритании и финансировался Министерством снабжения Великобритании из Государственного бюджета Великобритании для использования в рамках ядерной программы Великобритании. Стал первым ЭВМ, который использовал в качестве носителя информации магнитный барабан

В 1969 г интерес к исследованию нейронных сетей угас (после публикации работы по машинному обучению Минского и Пейперта в 1969 году).

Основные вычислительные проблемы, возникающие при компьютерной реализации искусственных нейронных сетей:

- однослойные нейронные сети не могли совершать «сложение по модулю 2», то есть реализовать функцию «Исключающее ИЛИ»;
- компьютеры не обладали достаточной вычислительной мощностью, чтобы эффективно обрабатывать огромный объем вычислений, необходимый для больших нейронных сетей.

В 1974 г. А.И. Галушкиным и 1975 году Полом Дж. Вербосом была **разработан метод обратного распространения ошибки**, который позволил эффективно решать задачу обучения многослойных сетей и решить проблему со «сложением по модулю 2». Основная идея этого метода состоит в распространении сигналов ошибки от

выходов сети к её входам, в направлении, обратном прямому распространению сигналов в обычном режиме работы

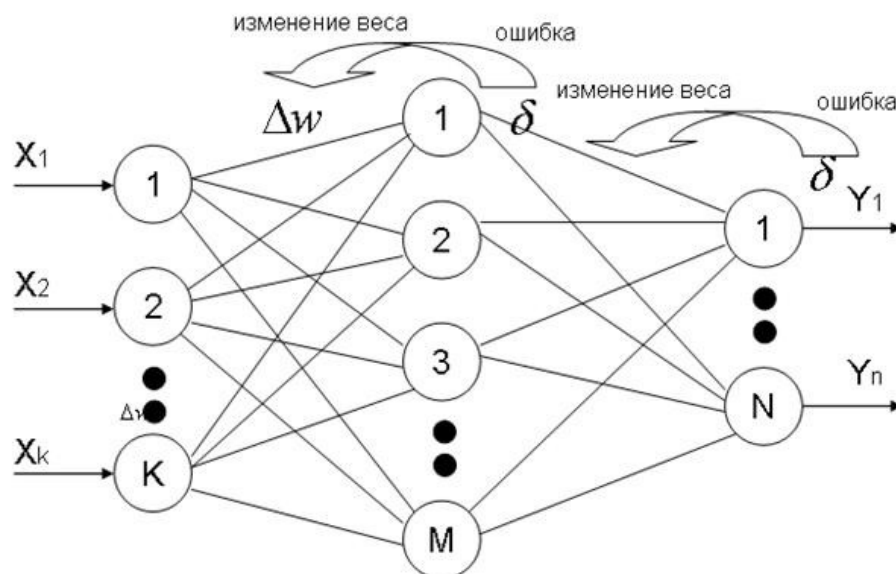


Рис. 6.5 – Метод обратного распространения ошибки

В 1975 году Фукусимой был разработан когнитрон, который стал одной из первых многослойных нейронных сетей. Фактическая структура сети и методы, используемые в когнитроне для настройки относительных весов связей, варьировались от одной стратегии к другой.

Каждая из стратегий имела свои преимущества и недостатки.

Сети могли распространять информацию только в одном направлении или перебрасывать информацию из одного конца в другой, пока не активировались все узлы и сеть не приходила в конечное состояние.

Достичь двусторонней передачи информации между нейронами удалось лишь в сети Хопфилда (1982), и специализация этих узлов для конкретных целей была введена в первых гибридных сетях.

Перцептоны.

Перцептрон – это некоторый класс моделей мозга или отдельной его системы (например, зрительной).

Понятие «перцептрон» ввел в 1957 году американский нейрофизиолог Френк Розенблатт (1928-1969), который создал в 1961 г. первую нейросетевую машину «Марк-1».

Схематично устройство перцептрона показано на рис. 6.6.

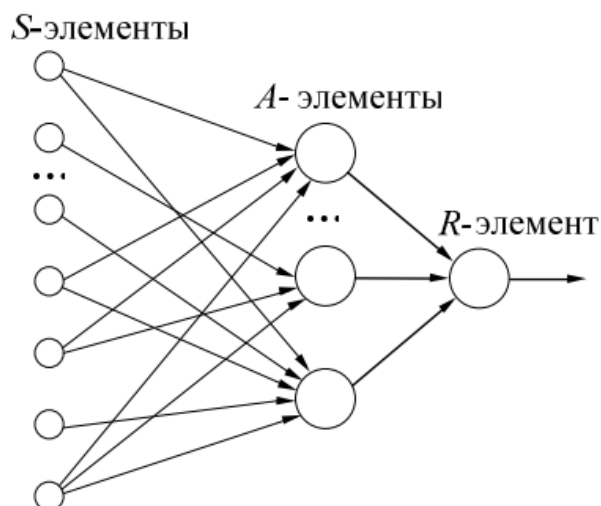


Рис. 6.6 – Схема перцептрона

S – набор чувствительных сенсорных элементов (сетчатка),

A – набор ассоциирующих элементов (нейронов),

R – реагирующий элемент (т.е. нейрон, передающий сигнал управления мышцам или железам).

Принципы функционирования перцептрона:

- 1) Входы в нейрон подразделяются на тормозящие и возбуждающие.
- 2) Сенсорные элементы возбуждаются, если в результате воздействия раздражителя (например, света) величина входного сигнала окажется больше некоторого порогового значения.
- 3) S-элементы случайным образом связаны с A-нейронами.
- 4) A-нейрон возбуждается и посылает сигнал на реагирующий элемент, если число возбуждающих сигналов больше числа тормозящих.
- 5) Сигналы, поступающие на реагирующий элемент, суммируются с некоторыми весами. Реагирующий элемент выбирает некоторое действие, если

$$R(\mathbf{x}) = \sum_{i=0}^n w_i x_i = (\mathbf{w}, \mathbf{x}) > 0,$$

где $w = (w_0, w_1, \dots, w_n)$, $x = (1, x_1, \dots, x_n)$, x_i – сигнал, поступающий на реагирующий элемент от i -го нейрона ($x_0 \equiv 1$ – сигнал смещения).

Постановка задачи обучения перцептрона:

Имеется некоторая обучающая выборка $\Xi = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ векторов двух классов \mathcal{O}_1 и \mathcal{O}_2 . Требуется построить линейную решающую функцию (ЛРФ) $R(\mathbf{x}) = (\mathbf{w}, \mathbf{x})$, которая бы правильно разделяла элементы обучающей выборки, т.е.

$$\begin{aligned} (\mathbf{w}, \mathbf{x}_i) &> 0, & \text{если } \mathbf{x}_i \in \mathcal{O}_1, \\ (\mathbf{w}, \mathbf{x}_i) &< 0, & \text{если } \mathbf{x}_i \in \mathcal{O}_2. \end{aligned}$$

Пусть $\Xi_n = \{\mathbf{x}_1, \dots, \mathbf{x}_N, \mathbf{x}_1, \dots, \mathbf{x}_N, \dots\}$.

Алгоритм обучения персептрона:

- 1) Иницируется начальное значение $w^{(0)}$ весового вектора персептрона.
- 2) Персептрону предъявляется очередной вектор x_k из обучающей выборки E_n и осуществляется коррекция весового вектора по правилу:

$$w^{(k+1)} = \begin{cases} w^{(k)}, & \text{если } (w^{(k)}, x_k) > 0 \text{ и } x_k \in \overline{\omega}_1, \\ w^{(k)}, & \text{если } (w^{(k)}, x_k) < 0 \text{ и } x_k \in \overline{\omega}_2, \\ w^{(k)} + x_k, & \text{если } (w^{(k)}, x_k) \leq 0 \text{ и } x_k \in \overline{\omega}_1, \\ w^{(k)} - x_k, & \text{если } (w^{(k)}, x_k) \geq 0 \text{ и } x_k \in \overline{\omega}_2. \end{cases}$$

- 3) Проверяется условие останова для найденного весового вектора: алгоритм завершает свою работу, если он N раз подряд правильно классифицирует элементы обучающей выборки. В противном случае переходим к пункту 2.

Идеология нейроинформатики.

В основе **нейроинформатики** (раздела искусственного интеллекта, который занимается исследованием нейронных сетей (НС)) лежит два представления:

- о строении мозга;
- о процессах обучения.

При рассмотрении строения мозга ключевым элементом является понятие простейшего элемента – «нейрона». Второе представление базируется на возможности, по аналогии с живыми организмами, формировать путем обучения такие связи между нейронами, чтобы множество нейронов (нейронная сеть) могло решать определенную задачу.

Принцип обучения НС.

Обучение НС обычно строится по принципу «поощрение-наказание»: системе предъявляется набор примеров с заданными ответами. Нейроны преобразуют входные сигналы, выдают ответ – также набор сигналов. Отклонение от правильного ответа штрафуются путем изменения внутренней настройки сети. Обучение состоит в минимизации отклонения желаемого результата от действительного.

Отличия между нейрокомпьютерами и «фон-неймановскими» компьютерами:

- 1) нейрокомпьютер способен решать не одну, а целый **класс задач**;
- 2) эффективность решения **«интеллектуальных»** задач – распознавания, узнавания, чтения и т.д.
- 3) нейрокомпьютер имеет **однородную** аппаратную реализацию, т.е. конструируется из однотипных элементов – нейронов;
- 4) нейросетевая архитектура обладает **высокой** степенью **распараллеливания** вычислений;

- 5) нейрокompьютер достаточно просто позволяет осуществлять **изменение конфигурации** вычислительной системы;
- 6) НС, как правило, не программируется, а **обучается** на решение какой-либо задачи, в отличие от обычного компьютера;
- 7) НС обладают высокой **надежностью** и **устойчивостью** к небольшим изменениям или повреждениям сети.

Элементы нейронных сетей.

Первая модель кибернетического нейрона была предложена в 1943 году в статье известного нейрофизиолога Уоренна МакКаллока (McCulloch W.S.) МакКаллок У. и его ученика, в то время студента Уолтера Питтса (Pitts W.) «Исчисление идей, имманентных нервной активности».

Формальный нейрон моделирует естественный нейрон. Известно, что в коре головного мозга человека порядка $\approx 10^{11}$ нейронов, каждый из которых связан с $\approx 10^3 - 10^4$ – другими нейронами. Естественный нейрон имеет множество отростков – *дендритов*, по которым нервные импульсы поступают в нейрон, и одно длинное волокно – *аксон*, по которому нервный импульс от данного нейрона передается на другие нейроны. Аксон данного нейрона соединяется с дендритами других нейронов с помощью так называемых *синапсов*. Если величина суммарного заряда, поступившего в клетку, превышает некоторое пороговое значение, то нейрон возбуждается и передает импульс через аксон и синапсы на другие нейроны. В настоящее время нейрофизиологам известны **около 50 разных типов нейронов**.

Формальный нейрон:

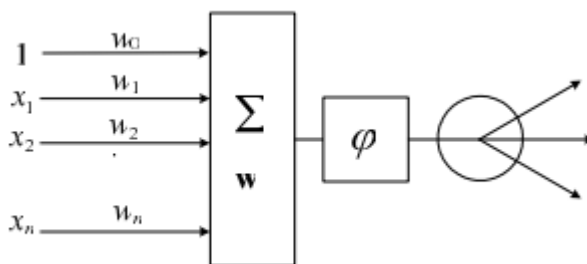


Рис. 6.7 – Схема формального нейрона

$X = (1, x_1, \dots, x_n)^T$ – вектор входного сигнала (вход нейрона – *дендрит*);

$w = (w_0, w_1, \dots, w_n)^T$ – вектор настраиваемых параметров (*синапсов*);

Σ – *адаптивный сумматор*, вычисляет скалярное произведение $t = (w, x)$;

ϕ – нелинейный преобразователь (*функция активации*), преобразует t в выходной сигнал $\phi(t)$ (выход нейрона - *аксон*).

Примеры функций активации:

- 1) функция Хэвисайда $\eta(t) = \begin{cases} 1, & t > 0, \\ 0, & t \leq 0, \end{cases}$
- 2) функция знака (сигнум) $\text{sgn}(t) = \begin{cases} 1, & t > 0, \\ -1, & t \leq 0, \end{cases}$
- 3) тождественная функция $\varphi(t) = t$;
- 4) гиперболический тангенс $th(t)$;
- 5) гауссовская функция $e^{-t^2/2}$;
- 6) логарифмическая функция $\ln(t + \sqrt{t^2 + 1})$;
- 7) сигмоидная функция $(1 + e^{-t})^{-1}$.

Архитектуры нейронных сетей:

Среди множества НС можно выделить две базовые архитектуры – слоистые и полносвязные сети.

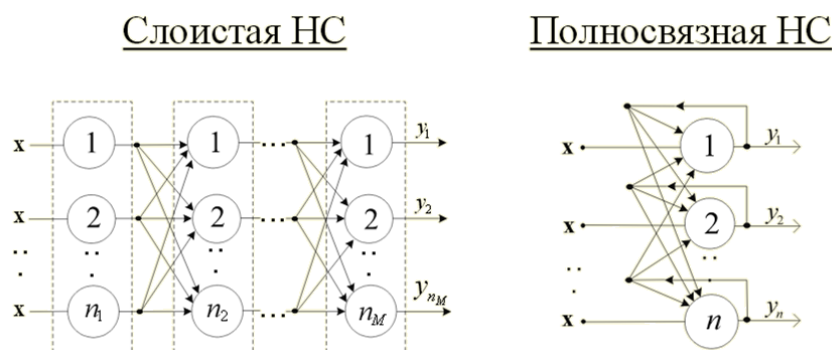


Рис. 6.8 – Базовые архитектуры НС

Базовые математические задачи, решаемые нейронными сетями.

1) **Базовые задачи, решаемые одним нейроном:** вычисление скалярного произведения и нелинейной функции от него.

Приложение: нахождение ЛРФ, разделяющей два класса.

2) **Базовые задачи, решаемые слоем из m нейронов:** НС из m нейронов с может вычислять произведение матрицы $W = [w_1, \dots, w_m]^T$ на вектор x : $s = Wx$, w_i – вектор синаптических связей i -го нейрона ($i = 1, \dots, m$), x – входной вектор, поступающий на все нейроны, $s_i = (w_i, x)$ – значение на выходе i -го нейрона ($i = 1, \dots, m$).

Приложение: вычисление градиента квадратичной формы

$$K(\mathbf{x}) = \frac{1}{2}(\mathbf{x}, W\mathbf{x}) : \text{grad}K(\mathbf{x}) = W\mathbf{x}.$$

3) **Базовые задачи, решаемые полносвязной сетью:** вычисление минимума квадратички $P(\mathbf{x}) = \frac{1}{2}(\mathbf{x}, W\mathbf{x}) + (\mathbf{b}, \mathbf{x})$

Так как $\text{grad}P(\mathbf{x}) = W\mathbf{x} + \mathbf{b}$, то минимум $P(\mathbf{x})$ найдем методом градиентного спуска:

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - h(W\mathbf{x}^{(k)} + \mathbf{b}), \quad k = 1, 2, \dots,$$

где I – единичная матрица. Процедура реализуется на полносвязной сети: каждый j -й нейрон связан с i -м нейроном ($i \neq j$) с помощью веса $-hw_{ij}$, с сами собой – с помощью веса $1 - hw_{ij}$. Единичный сигнал на вход j -й нейрона подается с весом $-hb_j$.

Основные алгоритмы обучения нейронных сетей:

Общий принцип обучения НС – принцип «поощрения и наказания» реализуется с помощью функционально-оптимизационного подхода, основанного на введении некоторого функционала $F_{\Xi, Y}(\mathbf{w})$, который характеризует величину отклонения откликов НС при предъявлении ей векторов обучающей выборки $\Xi = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ от соответствующих эталонных откликов $Y = \{y_1, \dots, y_N\}$. Обучение НС состоит в нахождении такого вектора параметров \mathbf{w} . НС, при котором $F_{\Xi, Y}(\mathbf{w}) \rightarrow \min$. Минимизация, как правило, осуществляется методом градиентного спуска в итерационной процедуре, реализуемой самой НС.

Алгоритмы обучения одного нейрона. Алгоритм Хебба:

Хебб Дональд Олдинг (Hebb D.O.) (1904 – 1985) – канадский физиолог и нейропсихолог.

Правило Хебба (1949): обучение нервных клеток мозга происходит путем усиления связей между теми нейронами, которые сильно возбуждаются.

Постановка задачи: требуется настроить один нейрон на распознавание образов двух классов. Пусть $\Xi = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, $\mathbf{x}_i = (x_{ij})$, $x_{ij} \in \{-1, 1\}$ – обучающая выборка биполярных векторов, $Y = \{y_1, \dots, y_N\}$ – множество эталонных откликов нейрона: $y_i = 1$, если $x_i \in \omega_1$ и $y_i = -1$, если $x_i \in \omega_2$.

Правило Хебба: если нейрон правильно классифицирует вектор, то порог синаптических связей снижается пропорционально этому вектору.

Алгоритм Хебба обучения одного нейрона:

1) Иницируются начальные значения весового вектора – вектора синапсов: $\mathbf{w}^{(0)}$.

2) Для всех пар (x_i, y_i) , $x_i \in \Xi$, $y_i \in Y$ выполняется коррекция весового вектора по формуле $w^{(k+1)} = w^{(k)} + x_i y_i$.

3) Проверяется условие останова для найденного весового вектора w : для каждой пары (x_i, y_i) , $x_i \in \Xi$, $y_i \in Y$ вычисляются значения $s_i = \text{sgn}((w, x_i))$. Если $s_i = y_i$ для всех $i = 1, \dots, N$, то алгоритм прекращает свою работу, в противном случае – переход к пункту 2.

Обучение многослойной нейронной сети методом обратного распространения ошибки

Метод обратного распространения ошибки (error back propagation) был предложен Румельхартом, Хинтоном и Уильямсом (Rumelhart D.E., Hinton G.E., Williams R.J.) в 1986 году.

Имеется сеть, состоящая из M слоев, $\Xi = \{x_1, \dots, x_N\}$ – обучающая выборка, $Y = \{y_1, \dots, y_N\}$ – множество правильных откликов сети, $w_{ij}^{(n)}$ – синаптическая связь между i -м нейроном n -1-го слоя и j -м нейроном n -го слоя, $x^{(n)}$ – вектор на выходе n -го слоя, $n = 1, \dots, M$.

Постановка задачи:

Найти такое множество весовых матриц $W^{(n)} = (w_{ij}^{(n)})$, $n = 2, \dots, M$, чтобы среднеквадратичная ошибка неправильной классификации была минимальной, т.е.

$$F(W^{(2)}, \dots, W^{(M)}) = \sum_{x \in \Xi} \|x^{(M)} - y\|^2 = \sum_{x \in \Xi} \sum_j (x_j^{(M)} - y_j)^2 \rightarrow \min$$

Алгоритм обучения НС методом обратного распространения ошибки:

1) Иницируются начальные значения весовых матриц $W(2), \dots, W(M)$.

2) На вход первого слоя сети подается очередной обучающий вектор x и вычисляются все значения $s_j^{(n)} = \sum_i x_i^{(n-1)} w_{ij}^{(n)}$, $n = 1, \dots, M$, $j = 1, \dots, N$.

3) По формуле

$$\delta_j^{(n)} = \left(\sum_k \delta_k^{(n+1)} w_{jk}^{(n+1)} \right) \frac{dx_j^{(n)}}{ds_j^{(n)}}, \quad n = M-1, M-2, \dots, 1$$

(или $\delta_j^{(M)} = (x_j^{(M)} - y_j) \frac{dx_j^{(M)}}{ds_j^{(M)}}$ для последнего слоя) вычисляются значения ошибок $\delta_j^{(n)}$

$$\Delta w_{ij}^{(n)} = -h \frac{\partial F}{\partial w_{ij}^{(n)}} = -h \frac{\partial F}{\partial x_j^{(n)}} \frac{dx_j^{(n)}}{ds_j^{(n)}} \frac{\partial s_j^{(n)}}{\partial w_{ij}^{(n)}} = -h \delta_j^{(n)} x_i^{(n)}$$

4) По формуле осуществляется коррекция веса на данном слое.

5) Выполняя обратное распространение ошибки по формуле

$$\delta_j^{(n)} = \left(\sum_k \delta_k^{(n+1)} w_{jk}^{(n+1)} \right) \frac{dx_j^{(n)}}{ds_j^{(n)}}, \quad n = M-1, M-2, \dots, 1$$

, корректируем веса на других слоях.

6) Проверяем условие останова – если $F(k+1) = F(k)$, то алгоритм завершает работу, в противном случае – переход к пункту 2.

Алгоритм и сеть Кохонена

Кохонен Теуво (Kohonen T.) (р.1943), финский ученый

Постановка задачи: Пусть $\Xi = \{x_1, \dots, x_N\}$ – выборка векторов в R^n . Найти разбиение Ξ на m классов w_1, \dots, w_m (ядер выборки Ξ) так, чтобы

$$F(w_1, \dots, w_m) = \frac{1}{2} \sum_{i=1}^m \sum_{x \in X_i} \|x - w_i\|^2 \rightarrow \min$$

где X_i – множество тех векторов выборки Ξ , которые «ближе» к ядру w_i .

Сеть Кохонена состоит из m нейронов: i -й нейрон вычисляет меру близости входного вектора-образа x к ядру w_i .

Алгоритм Кохонена:

Устанавливаются начальные значения ядер $w_i^{(k)}$, $k = 0$.

На вход сети подается вектор x обучающей выборки, вычисляются расстояния $d_i = \|w_i^{(k)} - x\|$ между вектором x и всеми ядрами $w_i^{(k)}$. Определяется тот j -й нейрон, для которого расстояние d_j – минимально.

Осуществляется коррекция весового вектора $w_j^{(k)}$ j -го нейрона по формуле $w_j^{(k+1)} = w_j^{(k)} + h(x - w_j^{(k)})$.

Проверяется условие стабилизации ядер: $w_i^{(k+1)} = w_i^{(k)}$ для всех $i = 1, \dots, m$. В противном случае – переход к пункту 2.

Сети ассоциативной памяти

Сеть ассоциативной памяти – это нейронная сеть, которая решает следующую задачу:

В синаптических связях НС «зашивается» информация о некоторых эталонных e_1, \dots, e_m . При поступлении на вход сети вектора признаков x сеть должна «вспомнить» тот эталонный образ, который «ближе» всего в заданной метрике к входному вектору x .

Наиболее популярные сети ассоциативной памяти:

- сеть Хопфилда;
- сет Хэмминга.

Алгоритм и сеть Хопфилда

Нейронная сеть Хопфилда решает задачу ассоциативного «узнавания» в евклидовой метрике – для каждого предъявленного биполярного вектора x НС находит наиболее близкий к нему образ-эталон e_p и выдает его на выходе.

Эта сеть была разработана в 1982 году известным американским физиком, специалистом в области физики твердого тела Джоном Хопфилдом (Hopfield J.J.) (р. 1933).

Информация об образах-эталонах «зашифрована» в синаптических связях самой НС. Для биполярных векторов

$$\|x - e_p\|^2 = \|x\|^2 + \|e_p\|^2 - 2(x, e_p) = 2n - 2(x, e_p)$$

Если $x, e_p \in R^n$. Тогда

$$\|x - e_p\| \rightarrow \min \Leftrightarrow (x, e_p) \rightarrow \max$$

Функционал энергии (функционал Ляпунова)

$$F(x) = -\frac{1}{2} \sum_p (x, e_p)^2 + \alpha \sum_i (x_i^2 - 1)^2, \quad \alpha > 0.$$

Задача: Найти $x: F(x) \rightarrow \min$.

Алгоритм Хопфилда:

- 1) На вход сети подается вектор x , полагается $y^{(0)} = x, k = 0$.
- 2) Рассчитываются новые состояния нейронов по формуле $s_i^{(k+1)} = \sum_j q_{ij} y_j^{(k)}$ и новые значения аксонов по формуле $y_i^{(k+1)} = \text{sgn}(s_i^{(k+1)})$.
- 3) Проверяется условие стабилизации аксонов: $y^{(k+1)} = y^{(k)}$, то алгоритм завершает работу, в противном случае – переход к пункту 2.

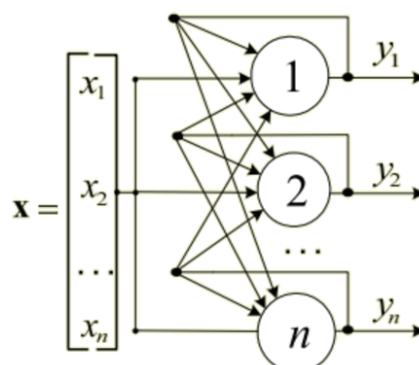


Рис. 6.9 – Нейронная сеть Хопфилда

Алгоритм и сеть Хэмминга

НС Хэмминга решает задачу ассоциативного «узнавания» в метрике Хэмминга – для каждого предъявленного биполярного вектора x НС находит наиболее близкий к нему образ-эталон e_j и выдает на выходе номер этого эталона. Разработана Ричардом Хэммингом (1915-1988) американским математиком, специалистом по теории кодирования. Свойство расстояния Хэмминга для биполярных векторов: если $x = (x_i)$ и $y = (y_i)$, $x_i, y_i \in \{-1, 1\}$, то

$$\|x - y\|_1 = \sum_i |x_i - y_i| = 2d(x, y), \text{ где } d(x, y) - \text{число различных компонент векторов.}$$

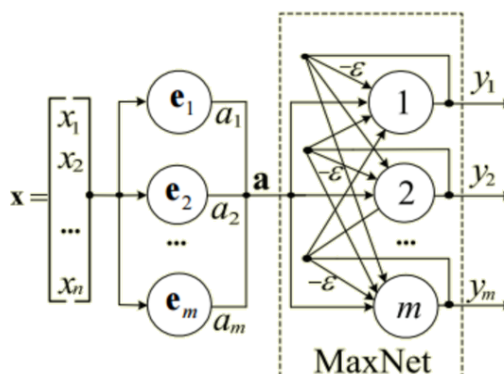


Рис. 6.9 – Нейронная сеть Хэмминга

Алгоритм Хэмминга:

1) На вход первого слоя подается вектор $\bar{x} = (1, x_1, \dots, x_n)^T$, вычисляется вектор мер близостей $\mathbf{a} = (a_1, \dots, a_m)^T$, $y^{(0)} = \mathbf{a}$, $k = 0$.

2) На вход каждого p -го нейрона второго слоя поступает вектор $y^{(k)}$. Рассчитываются новые состояния нейронов по формуле $s_i^{(k+1)} = \sum_j q_{ij} y_j^{(k)}$, где

$$q_{ij} = \begin{cases} 1, & i = j, \\ -\epsilon, & i \neq j, \end{cases} \quad 0 < \epsilon \leq 1/n.$$

Вычисляются новые значения аксонов

$$y_i^{(k+1)} = \varphi(s_i^{(k+1)}), \text{ где } \varphi(t) = \begin{cases} t, & t \geq 0, \\ 0, & t < 0. \end{cases}$$

3) Проверяется условие стабилизации аксонов: если $y^{(k+1)} = y^{(k)}$, то алгоритм завершает работу, в противном случае – переход к пункту 2.

Метод потенциальных функций

Этот подход был развит в 1962-1965 годах в работах советских математиков М.А. Айзермана, Э.М. Браверманна, Л.И. Розоноэра и др. Э.М. В этом методе для построения по множеству прецедентов решающей функции $d(\cdot) x$, разделяющей точки двух классов, используется следующая физическая аналогия. Каждая точка обучающей выборки отождествляется с единичным гравитационным зарядом. Множество таких зарядов создает гравитационное поле. Если имеются несколько множеств точечных зарядов, соответствующих разным классам, то пробный заряд

«притянется» к тому классу, который в данной точке пространства создает больший потенциал.

Метод потенциальных функций - метрический классификатор, частный случай метода ближайших соседей. Позволяет с помощью простого алгоритма оценивать вес («важность») объектов обучающей выборки при решении задачи классификации.

Общая идея метода иллюстрируется на примере электростатического взаимодействия элементарных частиц. Известно, что потенциал («мера воздействия») электрического поля элементарной заряженной частицы в некоторой точке пространства пропорционален отношению заряда частицы (Q) к расстоянию до частицы (r):

$$\varphi(r) \sim \frac{Q}{r}$$

Метод потенциальных функций реализует полную аналогию указанного выше примера. При классификации объект проверяется на близость к объектам из обучающей выборки. Считается, что объекты из обучающей выборки «заряжены» своим классом, а мера «важности» каждого из них при классификации зависит от его «заряда» и расстояния до классифицируемого объекта.

http://www.machinelearning.ru/wiki/index.php?title=%D0%9C%D0%B5%D1%82%D0%BE%D0%B4_%D0%BF%D0%BE%D1%82%D0%B5%D0%BD%D1%86%D0%B8%D0%B0%D0%BB%D1%8C%D0%BD%D1%8B%D1%85_%D1%84%D1%83%D0%BD%D0%BA%D1%86%D0%B8%D0%B9

Лекция 7 «Нейронная сеть в распознавании образов».

Искусственные нейронные сети представляют собой математическую модель функционирования биологических нейронных сетей – сетей нервных клеток живого организма. Как и в биологической нейронной сети, основным элементом искусственной нейросети является нейрон. Соединенные между собой нейроны образуют слои, количество которых может варьироваться в зависимости от сложности нейросети и решаемых ею задач. Кроме того, нейронные сети способны к обучению. Схема простой нейронной сети, имеющей два входных и один выходной элемент:

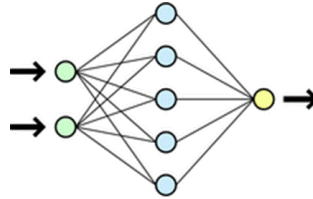


Рис. 7.1 – Общая схема нейронной сети

На сегодняшний день популярны благодаря своей простоте сети с прямым распространением сигнала и сети с обратным распространением ошибки. Они позволяют решать широкий спектр задач, например, классификацию, прогнозирование, распознавание образов.

Каждый нейрон имеет свои вход/входы, функцию активации, выход/выходы. В качестве функции активации используют обычно сигмоидальные функции, поскольку они позволяют свести бесконечное число значений входного сигнала к конечному отрезку значений выходных. Например, логистическая функция

$$a = \frac{1}{1 + e^{-n}}$$

Её график выглядит так:

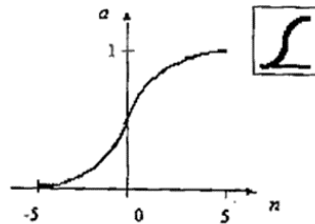


Рис. 7.2 – График функции $a = \frac{1}{1 + e^{-n}}$

Как видно из графика, область определения данной функции бесконечна, а область значений изменяется от 0 до 1.

Вид двухслойной сети с сигмоидальным тангенсом в качестве функции активации для входного слоя и линейной для выходного представлен на рисунке 7.1

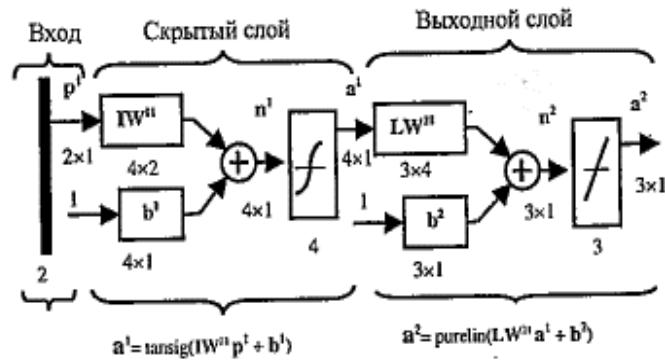


Рис. 7.3 – Структура двухслойной нейронной сети

При обучении сети рассчитывается некоторый функционал, характеризующий качество обучения:

$$J = \frac{1}{2} \sum_{q=1}^Q \sum_{i=1}^{S^M} (t_i^q - a_i^{qS^M})^2,$$

Здесь J – функционал, Q – объём выборки обучения, M – число слоёв сети, S^M – число нейронов выходного слоя, a^q – вектор сигнала на выходе сети, t^q – вектор целевых значений сигнала для выборки с номером q .

$$J = \frac{1}{2} \sum_{q=1}^Q \sum_{i=1}^{S^1} (t_i^q - a_i^{qS^1})^2 = \frac{1}{2} \sum_{q=1}^Q \sum_{i=1}^{S^1} (t_i^q - f(n_i^q))^2, \quad i = \overline{1, S^1},$$

Это вид функционала ошибки для однослойной сети. Здесь $f(n^q)$ – функция активации,

$$n_i^q = \sum_{j=0}^R w_{ij} p_j^q$$

– сигнал на входе функции активации i -того нейрона, p^q – вектор входного сигнала, R – число элементов вектора входа, w_{ij} – весовые коэффициенты сети.

Если предположить, что функция активации дифференцируется, то мы можем рассчитать градиент функционала ошибки, на анализе которого и построено большинство функций обучения нейросетей:

$$\frac{\partial J}{\partial w_{ij}} = - \sum_{q=1}^Q (t_i^q - f(n_i^q)) \frac{\partial (f(n_i^q))}{\partial w_{ij}} = - \sum_{q=1}^Q (t_i^q - f(n_i^q)) f'(n_i^q) p_j^q.$$

Если положить $\Delta_i^q = (t_i^q - f(n_i^q)) f'(n_i^q) = (t_i^q - a_i^q) f'(n_i^q)$, $i = \overline{1, S^1}$, то вид предыдущего выражения упрощается следующим образом:

$$\frac{\partial J}{\partial w_{ij}} = - \sum_{q=1}^Q \Delta_i^q f'(n_i^q) p_j^q, \quad i = \overline{1, S^1}.$$

Методы обучения можно условно разделить на несколько порядков.

1) Методы нулевого порядка используют только информацию о значении функции в заданных точках

2) Методы первого порядка используют градиент функционала ошибки по настраиваемым параметрам: $x_{k+1} = x_k - \alpha_k g_k$

Здесь x_k – вектор параметров, α_k – параметр скорости обучения, g_k – градиент функционала.

3) Методы второго порядка используют значения второй производной функционала. Количество используемых итераций они могут сократить, но возрастают необходимые затраты на вычисления.

Создание нейросети для распознавания букв латинского алфавита

Выполнение в среде MATLAB подразумевает последовательно несколько этапов.

- 1) Создание обучающей и целевой последовательностей.
- 2) Создание сети.
- 3) Обучение сети.
- 4) Проверка качества обучения.

1) Создание обучающей последовательности

Создать обучающую последовательность можно двумя способами. Необходимо либо вручную набрать массив 35x26, каждый столбец которого будет представлять очередную букву латинского алфавита, либо воспользоваться функцией `prprob()`:

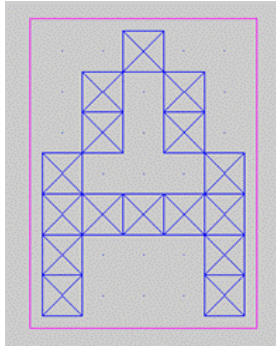
```
[ALPHABET,TARGETS] = PRPROB();
```

В этом случае сразу образуется и массив букв, который будет использоваться как обучающая последовательность, и массив векторов целей. ALPHABET и TARGETS являются именами переменных.

Каждый вектор в массиве ALPHABET будет представлять собой развёрнутое изображение буквы. Но чтобы получить более наглядное изображение, надо воспользоваться функцией `plotchar()`:

```
>> alphabet=prprob();  
>> plotchar(alphabet(:,1))
```

Это выведет следующий график:



2) Создание сети

Чтобы создать сеть с прямым распространением сигнала, надо воспользоваться функцией `newff()`. Ей подаются следующие аргументы:

- Вектор минимальных и максимальных значений каждого входа. Целесообразно использовать функцию `minmax()`;
- Количество нейронов в каждом слое в виде вектор-строки;
- Множество строк, каждая из которых называет используемую функцию активации нейронов в слое;
- Строка с именем функции обучения сети.

Далее, необходимо настроить такие параметры, как количество этапов обучения, функцию качества обучения, коэффициенты и постоянные слоёв.

Пример

```
net=newff(minmax(alphabet),[S1 S2],{'logsig','logsig'},'traingdx');
net.performFcn='sse';
net.LW{2,1}=net.LW{2,1}*.01;
net.b{2}=net.b{2}*.01;
net.trainParam.goal=.1;
net.trainParam.epochs=5000;
net.trainParam.mc=.95;
net.trainParam.show=100;
```

3) Обучение сети

Обучение сети осуществляется посредством функции `train()`, аргументами которой являются нейросеть, обучающая последовательность, целевая последовательность.

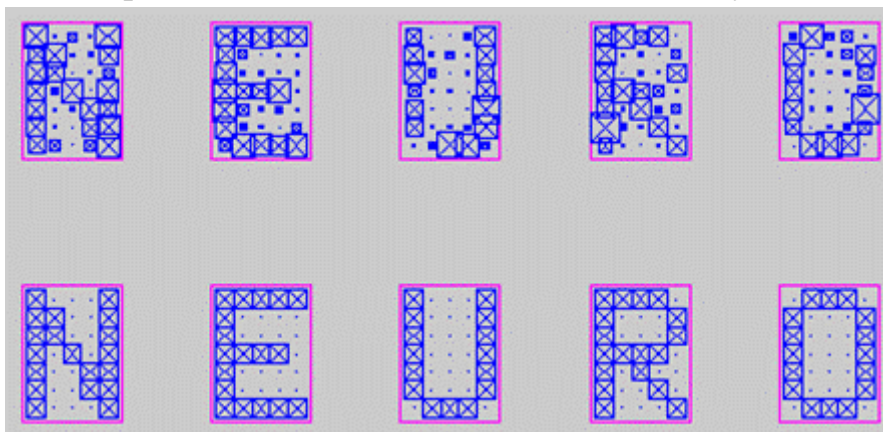
Пример

```
net=train(net,P,T);
```

4) Проверка качества обучения

Для проверки качества обучения необходимо добавить шум в обучающую последовательность, так как она будет использоваться для проверки обучения. Для этого нужно воспользоваться функцией `randn()`, которая возвращает массив заданной размерности с нормальным распределением псевдослучайных чисел. Потом в цикле производится увеличение шума на входной последовательности, сравнение выхода с целевым вектором, подсчёт ошибок.

Пример вывода сети: распознавание слова «N E U R O» в шуме



Пример

```

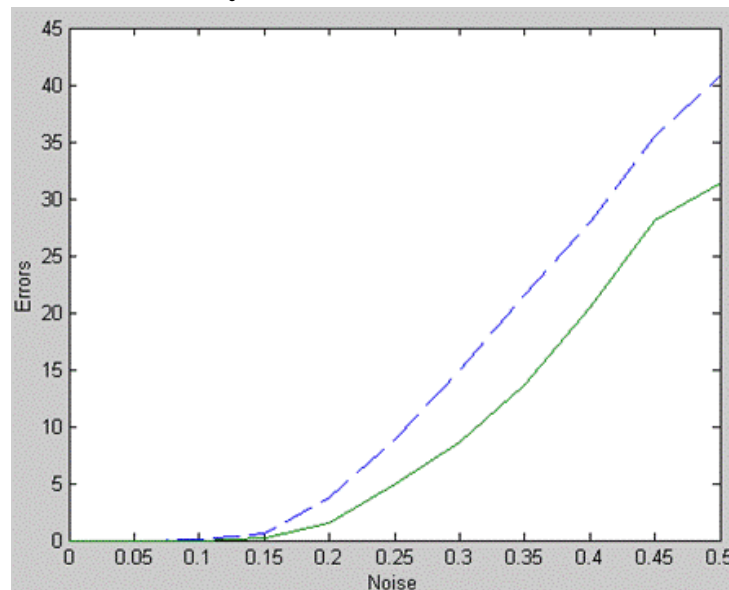
net1=train(net1,P,T);
noise_range=0:.05:.5;
max_test=100;
network1=[];
network2=[];
for noise_level=noise_range
    errors1=0;
    errors2=0;
    for i=1:max_test
        P2=P+randn(35,26)*noise_level;

        A=sim(net,P2);
        AA=compet(A);
        errors1=errors1+sum(sum(abs(AA-T)))/2;

        An=sim(net1,P2);
        AAn=compet(An);
        errors2=errors2+sum(sum(abs(AAn-T)))/2;
        echo off;
    end
    network1=[network1 errors1/26/100];
    network2=[network2 errors2/26/100];
end
plot(noise_range,network1*100,'--',noise_range,network2*100);|

```

Пример графика ошибок на зашумлённых последовательностях:



Создание нейросети для распознавания изображений.

Рассмотрим более сложный пример. Цель примера сводится к распознаванию изображённых на фотографиях предметов от иных изображений.

Для распознавания изображений необходима более сложная структура сети. Простейший пример – отличие различных геометрических фигур друг от друга.

1) Обучающее множество представлено в виде трёх фигур (треугольник, квадрат, круг):



Каждый рисунок загружается командой `imread()`;

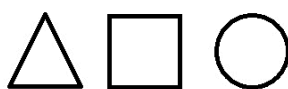
После загрузки необходимо преобразовать полученные данные в вектор-столбец с типом данных `double`.

2) Нейронная сеть содержит три слоя. На каждом используются логистические функции активации.

3) Обучение сети производится на множестве из трёх фигур. Целевой вектор целесообразно представить в виде единичной матрицы размером 3×3 .

4) Обученная нейронная сеть должна отличать одну фигуру от другой. За результат принимается аналитический вывод данных нейросети (большее значение соответствует распознанной фигуре).

Можно усложнить задачу, добавив пустые фигуры:



Нейронная сеть также должна позволить отличить одну фигуру от другой.

Если количество нейронов слишком велико, то будет проявляться эффект переобучения. Недостаточное количество нейронов не позволит эффективно распознавать изображения.

Эффективная конфигурация сети, найденная опытным путём, представляет собой три слоя:

- входной слой из шести нейронов с тангенциальной функцией активации
- промежуточный слой из пяти нейронов с тангенциальной функцией активации
- выходной слой из трёх нейронов и логистической функцией активации

Вывод

Созданные искусственные нейронные сети позволяют распознавать зашумлённые буквы латинского алфавита и изображения фигур.

Лекция 8 «Статистический подход в теории распознавания образов».

Очень часто результат одиночного опыта, произведенного с достаточно сложной системой, не может быть предсказан с полной определенностью, т.к. зависит от ряда случайных факторов и является случайной величиной. Несмотря на это, мы можем сделать важные выводы о поведении системы производя большое число идентичных опытов. В таком случае мы говорим о статистическом описании системы и используем для такого описания методы теории вероятностей. Статистическая теория позволяет предсказать не результат конкретного опыта, а вероятность появления каждого из возможных результатов опыта. Предсказанные вероятности можно сравнить с вероятностями, измеренными на опыте с большим числом одинаковых систем (ансамблем). Если состояние системы не зависит от времени, то с равным успехом можно многократно повторить один и тот же опыт над одной системой.

Теорема Байеса (или **формула Байеса**) — одна из основных теорем элементарной теории вероятностей, которая позволяет определить вероятность какого-либо события при условии, что произошло другое статистически взаимозависимое с ним событие. Другими словами, по формуле Байеса можно более точно пересчитать вероятность, взяв в расчет как ранее известную информацию, так и данные новых наблюдений.

Байес Томас (Bayes T.) (1702 – 1761) – английский математик и пресвитерианский священник.

Если событие A может произойти только при выполнении одного из событий B_1, B_2, \dots, B_n , которые образуют **полную группу несовместных событий**, то вероятность события A вычисляется по формуле:

$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + \dots + P(B_n)P(A|B_n).$$

Эта формула называется **формулой полной вероятности**.

Вновь рассмотрим полную группу несовместных событий B_1, B_2, \dots, B_n , вероятности появления которых $P(B_1), P(B_2), \dots, P(B_n)$. Событие A может произойти только вместе с каким-либо из событий B_1, B_2, \dots, B_n , которые будем **гипотезами**. Тогда по формуле полной вероятности

$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + \dots + P(B_n)P(A|B_n).$$

Если событие A произошло, то это может изменить вероятности гипотез $P(B_1), P(B_2), \dots, P(B_n)$.

По теореме умножения вероятностей

$$P(AB_1) = P(B_1)P(A|B_1) = P(A)P(B|A),$$

откуда

$$P(B_1 | A) = \frac{P(B_1)P(A | B_1)}{P(A)}$$

Аналогично для остальных гипотез

$$P(B_i | A) = \frac{P(B_i)P(A | B_i)}{P(A)}, i = 1, \dots, n.$$

Полученная формула называется **формулой Байеса (формулой Бейеса)**. Вероятности гипотез $P(B_i | A)$ называются **апостериорными вероятностями**, тогда как $P(B_i)$ – **априорными вероятностями**.

В случае, когда один и тот же образ может принадлежать сразу нескольким классам, но частота появления образа с конкретным значением для различных классов не одинакова, то определить оптимальную границу между классами можно только с учетом этого факта. При обработке изображений относительную частоту появления образа в классе можно оценить, построив гистограмму по объектам, принадлежащим заданному классу. При большом количестве образов в k -м классе эта частота стремится к вероятности $p_k(x)$ появления образа x в данном классе, и гистограмму можно рассматривать как дискретную аппроксимацию функции плотности распределения $f_k(x)$

Алгоритмы (реализуют принцип максимума правдоподобия) основаны на статистическом подходе к задаче распознавания и представляют собой классификаторы с обучением.

То есть предполагается, что:

- 1) заданы определенные классы и по ним имеются данные предварительных исследований (тестовые участки при обработке данных ДЗ);
- 2) эти данные позволяют получить представительные выборки образов по всем классам и оценить вероятность появления образа в каждом классе для всего множества X .

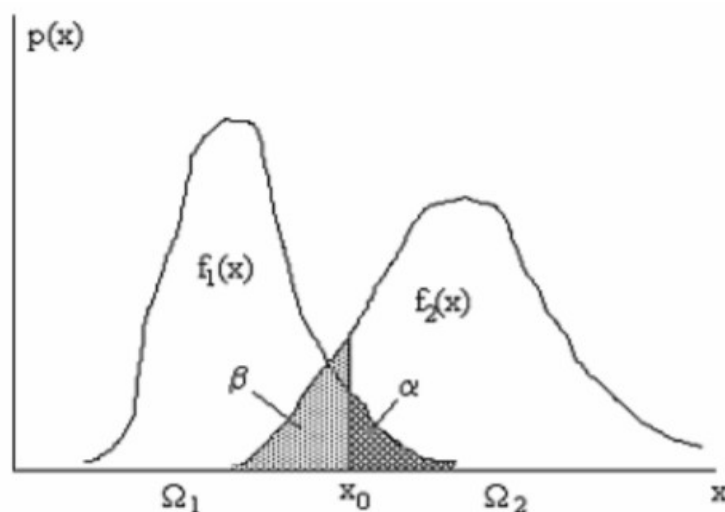
Рассмотрим правила принятия решений в задаче статистической классификации на примере двух классов и одного признака, заданного действительным значением измерения X . Задача заключается в определении на шкале X интервалов Ω_1 и Ω_2 , на которых будут приниматься решения в пользу первого и второго класса соответственно. Для простоты в дальнейшем классы и соответствующие им области принятия решения будем обозначать одним и тем же символом Ω

В теории статистических решений непротиворечивое множество предположений о свойствах случайной величины ξ называется статистической гипотезой. В данном случае мы рассматриваем две Статистическая гипотеза для случая двух классов. статистические гипотезы: ξ , принимающая значения x , имеет распределение с плотностью $f_1(x)$, то есть принадлежит классу Ω_1 , - это гипотеза H_1 против

альтернативы H_2 : ξ имеет распределение с плотностью $f_2(x)$, то есть принадлежит классу Ω_2 . Вероятность появления любого значения x отлична от нуля и для первого, и для второго класса на всем множестве X , поэтому при принятии решения относительно принадлежности некоторого значения x к одному из классов могут возникнуть 4 ситуации.

1. Принимаем гипотезу H_1 , и она верна.
2. Принимаем H_2 , но верна H_1 .
3. Принимаем H_2 , и она верна.
4. Принимаем H_1 , но верна H_2 .

Предположим, что получена вся необходимая статистическая информация о классах: функции плотности статистического распределения $f_1(x)$ и $f_2(x)$, а так же априорные вероятности $P(\Omega_1)$ и $P(\Omega_2)$ появления данных классов.



Точка x_0 , разделяет всю шкалу значений признака x на два интервала: $(-\infty, x_0]$ соответствует области решений в пользу класса Ω_1 , (x_0, ∞) - области решений в пользу класса Ω_2 .

Вероятность возникновения

ситуации 1 соответствует площади под $f_1(x)$ на полуинтервале $(-\infty, x_0]$,

ситуации 2 - площади под $f_1(x)$ на интервале (x_0, ∞) ,

ситуации 3 - площади под $f_2(x)$ на (x_0, ∞) ,

ситуации 4 - площади под $f_2(x)$ на $(-\infty, x_0]$.

Суммарная площадь под $f_1(x)$ и $f_2(x)$ для ситуаций 2 и 4 - это полная вероятность ошибок в нашей схеме принятия решений.

В случае двух альтернативных гипотез ошибку, соответствующую ситуации 2, обычно называют *ошибкой первого рода* (α),

ошибку, соответствующую ситуации 4, - *ошибкой второго рода* (β).

Понятие ошибок первого и второго рода симметрично и зависит от того, какая гипотеза является основной, а какая – альтернативной.

Если бы H_2 была основной гипотезой, ошибка первого рода соответствовала бы ситуации 4.

При классификации пикселей многозональных изображений ошибка первого рода проявляется в появлении на объектах класса Ω_1 точек посторонних классов. В свою очередь, ошибки второго рода проявляются в появлении точек этого класса на других объектах. Когда количество классов невелико, обычно преобладают ошибки второго рода. Это связано с тем, что аналитик не учитывает все присутствующие на изображении типы объектов, в том числе и с характеристиками, близкими к выделяемым классам.

Байесовская стратегия минимального среднего риска

Чтобы построить классификатор на основе описанной схемы, должны выработать правило оптимального разбиения всего множества X на области принятия решения в пользу каждого класса.

В конкретном одномерном случае - правило выбора разделяющей точки x_0 .

Поскольку различие между классами связано с частотой появления тех или иных значений x в этих классах, естественно выбрать точку x_0 таким образом, чтобы при многократном предъявлении классификатору образов x усредненная по совокупности решений ошибка была минимальной.

С этой целью введем “стоимость” каждого принятого решения. Для описанных выше четырех ситуаций с классами Ω_1 и Ω_2 запишем соответствующие платежные коэффициенты в виде матрицы:

$$\begin{vmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{vmatrix} \quad (1)$$

Диагональные элементы такой платежной матрицы соответствуют случаям правильной классификации (ситуации 1 и 3), два других - ошибкам (ситуации 2 и 4). Платежные коэффициенты можно задать так, чтобы они отражали наши предпочтения в отношении той или иной ситуации. Чаще всего ограничиваются условием $c_{11}=c_{22}=0$ и $c_{12}=c_{21}=1$, то есть рассматривают только потери от ошибок.

Рассматривая платежные коэффициенты как наш “риск” в каждой из возможных ситуаций (которые можно рассматривать как случайные события), введем понятие среднего риска для описанных четырех случаев:

$$R = \sum_i r_i p(r_i), \quad i=1, \dots, 4. \quad (2)$$

Здесь r_i - это платежные коэффициенты из матрицы, а $p(r_i)$ - вероятность каждого из четырех событий (вероятность соответствующей выплаты). В случае, когда $c_{11}=c_{22}=0$ и $c_{12}=c_{21}=1$ (то есть «плата» - это штраф за ошибки), функцию R называют также функцией потерь.

Рассмотрим, что представляют собой вероятности $p(r_i)$. Например, вероятность выплаты c_{11} есть вероятность одновременного осуществления двух событий: ξ принадлежит к $f_1(x)$ (вероятность этого события в Ω_1 есть

$$\int_{-\infty}^{x_0} f_1(x) dx$$

и появления самого класса Ω_1 (априорная вероятность этого события - $P(\Omega_1)$), то есть

$$p(r_1) = p(c_{11}) = P(\Omega_1) \int_{-\infty}^{x_0} f_1(x) dx.$$

Следовательно, формулу (2) можно записать так:

$$R = c_{11} P(\Omega_1) \int_{-\infty}^{x_0} f_1(x) dx + c_{12} P(\Omega_1) \int_{x_0}^{\infty} f_1(x) dx + c_{22} P(\Omega_2) \int_{-\infty}^{x_0} f_2(x) dx + c_{21} P(\Omega_2) \int_{x_0}^{\infty} f_2(x) dx. \quad (3)$$

R - есть полный средний риск так называемой стратегии Байеса.

Минимум R в точке x_0 достигается при условии

$$\left. \frac{dR}{dx} \right|_{x=x_0} = 0.$$

Возьмем производную в точке x_0 , учитывая, что $\int f(x) dx = F(x)$, $F(-\infty) = 0$, $F(\infty) = 1$:

$$\left. \frac{dR}{dx} \right|_{x=x_0} = c_{11} P(\Omega_1) f_1(x) - c_{12} P(\Omega_1) f_1(x) - c_{22} P(\Omega_2) f_2(x) + c_{21} P(\Omega_2) f_2(x) = 0. \quad (4)$$

имеем следующее соотношение для $x=x_0$:

$$\frac{P(\Omega_1) f_1(x)}{P(\Omega_2) f_2(x)} = \frac{c_{21} - c_{22}}{c_{12} - c_{11}} = \lambda. \quad (5)$$

Полученная формула (5) называется отношением правдоподобия, а величина λ - коэффициентом правдоподобия.

При значениях $\frac{P(\Omega_1) f_1(x)}{P(\Omega_2) f_2(x)} \geq \lambda$ решение принимается в пользу Ω_1 ,

при $\frac{P(\Omega_1) f_1(x)}{P(\Omega_2) f_2(x)} < \lambda$ - в пользу Ω_2 .

Если положить, что $c_{11}=c_{22}=0$ и $c_{12}=c_{21}=1$, получим:

$$\frac{P(\Omega_1)f_1(x)}{P(\Omega_2)f_2(x)} = 1 \text{ или, в логарифмической форме, } \ln \frac{P(\Omega_1)f_1(x)}{P(\Omega_2)f_2(x)} = 0 \quad (6)$$

Если значения признака для обоих классов распределены по нормальному закону

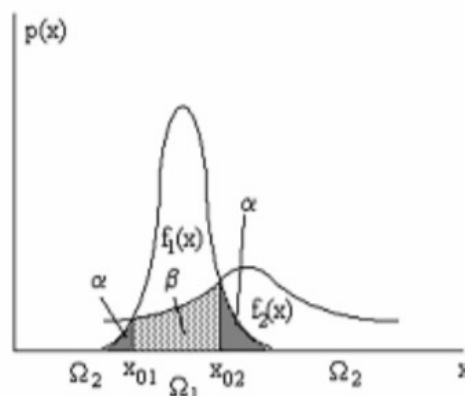
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} \quad (7)$$

со средними m_1 , m_2 и среднеквадратическими отклонениями σ_1 , σ_2 соответственно, отношение правдоподобия в логарифмической форме имеет вид:

$$\ln \frac{P(\Omega_1)}{P(\Omega_2)} + \ln \frac{\sigma_2}{\sigma_1} - \frac{(x-m_1)^2}{2\sigma_1^2} + \frac{(x-m_2)^2}{2\sigma_2^2} = 0. \quad (8)$$

То есть x_0 является решением квадратного уравнения (8).

Случай, когда уравнение имеет два действительных корня, представлен на рисунке:



Пример

Предположим, что необходимо построить классификатор, распознающий печатные буквы русского алфавита. В этом случае система должна отнести символ одному из 33 классов букв русского языка или указать на то, что символ не является буквой русского языка. Для построения системы распознавания определим некоторую систему признаков. Так как символ может быть набран тем или иным шрифтом, то существует большая неопределенность описания образа вектором признаков. При наличии большой статистики $\Xi = \{x_1, \dots, x_N\}$ (большого числа разных шрифтов) эту неопределенность можно описать с помощью статистической оценки плотности $f(x_i)$ распределения признаков в классе той или иной буквы. Такая задача называется задачей непараметрического оценивания. Если ограничиться построением классификатора, распознающего отсканированные буквы одного шрифта, то в этом случае неопределенность описания символа, прежде всего, будет обусловлена наличием шума сканирования. Если шум сканирования имеет вероятностную природу, то общий вид его закона распределения вероятностей заранее известен и требуется

только по выборке определить параметры этого закона. Такая задача называется задачей параметрического оценивания и некоторые способы ее решения, известны в статистике.

Например, плотность распределения вероятностей признаков i -й буквы может иметь вид

$$f_i(\mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}\|\mathbf{x}-\mathbf{m}_i\|^2}, \quad i=1,\dots,33,$$

(так называемое сферическое нормальное многомерное распределение, для соответствующего случайного вектора ξ будем использовать

$$\xi \sim N(\mathbf{m}_i, \sigma^2)$$

Здесь \mathbf{m}_i – математическое ожидание – центр рассеивания вектора признаков i -й буквы (т.е. это вектор признаков незашумленной буквы),

σ^2 – дисперсия шума.

По обучающей выборке

$\mathcal{E} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ можно получить оценки \mathbf{m}_i и σ^2 параметров \mathbf{m}_i и σ^2 функции $f(\mathbf{x}_i)$ методом максимального правдоподобия и они будут равны:

$$\tilde{\mathbf{m}}_i = \frac{1}{|\varpi_i|} \sum_{\mathbf{x}_k \in \varpi_i} \mathbf{x}_k, \quad \tilde{\sigma}^2 = \frac{1}{33} \sum_i \frac{1}{|\varpi_i|} \sum_{\mathbf{x}_k \in \varpi_i} \|\mathbf{x}_k - \mathbf{m}_i\|^2,$$

где ϖ_i – число элементов обучающей выборки, принадлежащих классу ϖ_i . Вероятность появления той или иной буквы-класса можно оценить по формуле

$$\tilde{p}(\varpi_i) = |\varpi_i| / N,$$

где N – общее число элементов выборки.

Далее можно найти и оценки других характеристик среды по формулам

$$f_\xi(\mathbf{x}) = \sum_{k=1}^m p_k f_k(\mathbf{x}) \text{ – формула полной вероятности,}$$

$$p(\varpi_i | \mathbf{x}) = \frac{p_i f_i(\mathbf{x})}{f_\xi(\mathbf{x})} \text{ – формула Байеса.}$$

Литература.

1. Лепский А.Е., Броневич А.Г. Математические методы распознавания образов: Курс лекций. – Таганрог: Изд-во ТТИ ЮФУ, 2009. – 155 с.
2. Волошин Г.Я. МЕТОДЫ РАСПОЗНАВАНИЯ ОБРАЗОВ (конспект лекций) Редактор: Ильин А.А. Размещено на <http://www.allbest.ru/>
3. А. Л. Горелик В. А. Скрипкин «Методы распознавания» Учеб. пособие для вузов. М., «Высш. школа», 1977.

- понятие искусственного интеллекта;
- особенности интеллектуальных задач;
- критерии «интеллектуальности» систем;
- основные подходы к разработке систем искусственного интеллекта;
- основные направления исследований.