

1. Изображение объекта. Виды изображений.

Изображение (описание) объекта – отображение объекта на воспринимающие органы (рецепторы) распознающей системы. Если воспринимающие органы фиксированы, то объект и его изображение отождествляются. *Виды изображений* : 1) объект в виде n -мерного вектора (вектор признаков) $X = (x_1, \dots, x_n)$. Н-р, набор технич-х хар-к машины, м/б бинарным. 2) объект в виде лингвистической структуры. Н-р, текст на естественном языке.

2. Образ (класс).

Пусть задано разбиение множества всех изображений на непересекающиеся подмножества. Каждое такое подмножество назовём классом или образом. Объекты, относящиеся к одному и тому же образу, образуют класс эквивалентности. Иногда под образом понимают множество, элементы которого обладают свойствами рефлексивности и симметричности, но не транзитивности (т.е. отношением толерантности) (для элементов на границе классов).

3. Задача распознавания образов.

Задача распознавания образов заключается в отнесении некоторого объекта из исходного множества к одному из m классов :
 X (описание объекта) \Rightarrow распознающая процедура \Rightarrow класс (образ).

4. Обучающая последовательность.

Пусть из выделенных классов выбрано множество объектов, на основе которого будет осуществляться построение распознающей процедуры. Назовём его обучающей последовательностью x^1, x^2, \dots, x^n .

5. Задача обучения распознаванию образов.

Задача обучения распознаванию образов – по данной обучающей последовательности с использованием некоторого устройства (конструктора) получить распознающую процедуру, которая осуществляла бы правильную классификацию для любого объекта из исходного множества, включая и не вошедшие в эту ОП.

6. Задача обучения с учителем.

Задача обучения с учителем – задача распознавания образов, в которой для каждого объекта из обучающей последовательности известен класс, к которому он относится (т.е. дана маркированная ОП).

7. Задача обучения без учителя.

Задача обучения без учителя (задача самообучения) – задача распознавания образов, в которой для образов из обучающей последовательности не известны классы, к которым они относятся (т.е. дана немаркированная ОП).

8. Решающая функция.

Пусть объект представлен вектором признаков $X=(x_1, \dots, x_n)$ и имеется m образов w_1, \dots, w_m . Рассмотрим $m(m-1)$ решающих функций $d_{ij}(x)$, где $i \neq j$, $i, j = 1 \dots m$, обладающих свойствами: 1) $d_{ij}(x) > 0$, $\forall j \neq i$, если $x \in w_i$. 2) антисимметричность: $d_{ij}(x) = -d_{ji}(x)$. Если есть такой набор РФ, то работа распознающей процедуры сводится к проверке для данного объекта x условия (1) для всех $i = 1 \dots m$. В результате м/б, что: а) объект отнесён к классу i^* . б) объект не отнесён ни к одному из m заданных классов. Множество объектов, не классифицируемых процедурой, составляет область неопределённости. в) по условию (2) невозможно отнесение процедурой объекта x более чем к одному классу: если при данном x усл (1) выполнено для нек-го i , то для всех $j \neq i$ $d_{ij}(x) < 0$, т.е. объект не отнесён ни к какому из классов с номером $j \neq i$. Частный случай: $d_{ij}(x) = d_i^o(x) - d_j^o(x)$, $i \neq j$, где m функций d_i^o обладают свойством: 3) $d_i^o(x) > d_j^o(x)$ для $\forall j \neq i$, если $x \in w_i \Rightarrow$ усл (1,2) выполняются автомат-и. В этом случае РП описывается так: x относится к классу w_i , где $i = \arg \max_j d_j^o(x)$ (знач-е арг-ента j , при кот-м достигается макс функции при фиксир-м x).

9. Примеры задач распознавания.

1. *Тестирование* комбинаторных схем (КС-схема, состоящая из элементов, реализующих основные логические функции).

.....
Её работу можно описать так:

Пусть в схеме есть дефект (элемент x не работает). Тогда опишем её набором:

Задача диагностики КС заключается в том, чтобы фиксируя сигналы на входе и выходе схемы, определить, является ли схема исправной, или она имеет данный дефект. Вектор $x=(x_1, x_2, x_3, x_4, y)$ размерности $n=5$, кол-во классов $m=2$.

2. Контроль состояния ядерного реактора.

.....
Изображение: $X=x_0I, \dots, x_{31}I$ – вектор в евклидовом пространстве. Имеются два класса: w_1 -множество значений вектора X , соответств-х нормальному состоянию реактора, w_2 - аномальному состоянию. *Задача* состоит в определении к какому классу относится объект X .

3. Распознавание символов. Изображение символа проецируется на светочувствительную сетчатку.

Степень затемненности элемента сетчатки можно измерять. Каждому начертанию символа ставится в соответствие вектор $X=(x_1, \dots, x_n)$, компоненты которого отражают степень затемненности. *Класс* – множество векторов, соответствующих различным начертаниям одного и того же символа. Классов столько же сколько символов. $3P$ – дан вектор $X^*=(x_1^*, \dots, x_n^*)$, РП должна решить какой символ спроецирован.

4. Выбор адекватного алгоритма для решения задачи. Имеется некоторое множество задач определенного типа. Предложено m различных эвристических алгоритмов A_1, \dots, A_m для их решения.

Множество задач можно разбить на m классов w_1, \dots, w_m . *Класс* w_i – подмножество задач, которое решается алгоритмом A_i наилучшим в некотором смысле образом. *Изображение* – описание конкретной задачи.

$3P$ – дана конкретная задача, распознающая процедура должна решить какой алгоритм будет решать ее наилучшим образом.

10. Общая характеристика простейших методов распознавания (сравнение с эталоном). Группа наиболее ранних методов, когда класс задаётся характерным представителем (эталоном).

Классифицируемый объект относится *распознающей процедурой* к тому классу, с эталоном которого он согласуется наилучшим образом относительно какой – либо характеристики.

11. Метод совмещения с эталоном.

Для каждого класса выбирается эталон (характерный представитель класса). *Распознающая процедура:* анализируемый объект сравнивается с эталонами классов и относится к тому классу, с эталоном которого он согласуется наилучшим образом. Иногда требуется просто совпадение с эталоном.

12. Метод зондов.

Метод зондов предназначен для распознавания печатных символов. Имеется начертательное поле, в котором расположена система тонкопроводящих зондов. На него накладывается символ-объединение зондов-вырабатывается определенный код символа. Система зондов расположена так, что различным очертаниям одного и того же символа соответствует один и тот же бинарный код. + в том, что решается просто. - в сложности построения таких зондов.

13. Квазитопологический метод.

Квазитопологический метод предназначен для распознавания символов. Каждому символу поставлен в соответствие граф в качестве *эталона* этого символа. РП: к одному классу относятся все очертания символов, графы которых гомеоморфны, то есть могут быть получены друг из друга с помощью взаимнооднозначного непрерывного отображения. -: один и тот же граф может соответствовать нескольким символам.

14. Достоинства и недостатки простейших методов распознавания.

- + простота распознающей процедуры, быстрое действие;
- проблема определения эталонов, узкая специализация эталонов.

15. Классификация по единственному эталону каждого класса.

В обуч. послед-ть входит по одному представителю (эталону) каждого класса. Определим расстояние м-у вектором признаков и эталоном Z_i

$$D_i = \|x - z_i\| = \sqrt{(x - z_i)^T (x - z_i)}$$

РП: $x \rightarrow w_i \quad i = [\arg \min D(z_j, x)] \quad x \in w_i$

относит О-т x к тому из классов, расстояние до которого мин.

Если расстояния одинаковы, то решение выносится произвольно

16. Классификация по множеству эталонов каждого класса.

Классификация по мн-ву эталонов. В обуч. пос-ть входит $N_i \geq 1$ представителей $Z_i^1, Z_i^2, \dots, Z_i^{N_i}$ каждого класса. Найдем расстояние м-у вектором X и классами w

$$D_i = \min_{l=1, N_i} \|x - z_i^l\| = \min_{l=1, N_i} \sqrt{(x - z_i^l)^T (x - z_i^l)}$$

РП: $x \rightarrow w_i \quad i = [\arg \min D(z_j, x)] \quad x \in w_i$

относит О-т x к тому из классов, расстояние до которого мин.

17. Правило ближайшего соседа.

Пусть есть ОП Z_1, \dots, Z_N , причём для каждого вектора Z_j известна его принадлежность к одному из классов W_1, \dots, W_m . Пусть $D(Z_j, x)$ – нек-я мера близости векторов Z_j и x

q-БС-правило ($q \geq 1$ – фиксированное целое число): вычисляется q ближайших соседей x и x относится к тому классу w_i , к которому относится большинство из q ближайших соседей.

18. Достоинства и недостатки методов, основанных на близости описаний.

+ простота реализации. – необходимость хранения всей ОП, отсутствие гарантии правильного распознавания, невозможность оценки качества РП.

19. Меры сходства изображений, используемые для выявления классов (кластеров).

В качестве меры между образами x и z можно использовать: 1) евклидово расстояние: $D = \|x - z\|$. 2)

неметрическая функция сходства $S(x, z) = \frac{x \cdot z}{\|x\| \|z\|}$ – косинус угла между векторами x и z , достигающий

максимума, когда их направления совпадают. 3) мера Танмото – отношение количества совпадающих к количеству различных признаков двух образов.

20. Подходы к построению процедуры классификации (кластеризации).

задание набора правил, основывающихся на использовании выбранной меры сходства для отнесения образов к одному из кластеров.

Эвристический подход Образ относится к тому кластеру, с центром которого он близок в большей степени. При таком подходе хорошо использовать евклидово расстояние, что связано с естественностью его интерпретации как меры близости.

Подход, основанный на использовании показателя качества, связан с разработкой процедур, которые обеспечат минимизацию или максимизацию выбранного показателя качества. Одним из наиболее популярных показателей качества является сумма квадратов ошибки

$$J = \sum_{j=1}^{N_c} \sum_{x \in S_j} \|x - m_j\|^2 \rightarrow \min \quad \text{где } N_c \text{ — число кластеров, } S_j \text{ — множество образов, относящихся к } j\text{-му}$$

кластеру, а $m_j = \frac{1}{N_j} \sum_{x \in S_j} x$ — вектор выборочных средних значений для множества S_j ; N_j количество

образов, входящих в множество S_j . Показатель качества опр-т общую сумму квадратов отклонений хар-к всех образов, входящих в некоторый кластер, от соответствующих средних значений по кластеру.

21. Простой эвристический алгоритм определения кластеров.

Ш0. Пусть задано множество N образов $\{X_1, \dots, X_N\}$. Пусть центр первого кластера Z_1 совпадает с любым из заданных образов и определена произвольная неотрицательная пороговая величина T . Будем считать, что $Z_1 = X_1$. Ш1. Вычислим расстояние D_{21} между образом X_2 и центром кластера Z_1 по формуле евклидова расстояния. Если это расстояние больше значения пороговой величины (T), то новым центром кластера становится $Z_2 = X_2$. В противном случае образ X_2 включается в кластер, центром которого является Z_1 . Пусть Z_2 — центр нового кластера. Ш3. Вычислим расстояния D_{31} и D_{32} от образа X_3 до центров кластеров Z_1 и Z_2 . Если оба расстояния оказываются больше порога T , то учреждается новый центр кластера $Z_3 = X_3$. В противном случае образ X_3 зачисляется в тот кластер, чей центр к нему ближе. Шm. Подобным же образом расстояния от каждого нового образа до каждого известного центра кластера вычисляются и сравниваются с пороговой величиной T , если все эти расстояния превосходят значение порога T , учреждается новый центр кластера. В противном случае образ зачисляется в кластер с самым близким к нему центром. Результаты описанной процедуры определяются выбором первого центра кластера, порядком осмотра образов, значением пороговой величины T , геометрическими характеристиками данных.

22. Эвристический алгоритм максиминного расстояния.

Ш0. Пусть задано множество N образов $\{X_1, \dots, X_N\}$. Ш1. Один из заданных образов произвольным образом назначается центром первого кластера Z_1 . Возьмем $Z_1 = X_1$. Ш2. Затем отыскивается образ, отстоящий от образа X_1 на наибольшее расстояние. Он назначается центром кластера Z_2 . Ш3. Вычисляются расстояния между всеми остальными образами выборки и центрами кластеров Z_1 и Z_2 . В каждой паре этих расстояний выделяется минимальное. После этого выделяется максимальное из этих минимальных расстояний. Если последнее составляет значительную часть расстояния между центрами кластеров z_1 и z_2 (по меньшей мере, половину этого расстояния), то соответствующий образ назначается центром кластера Z_3 . В противном случае выполнение алгоритма прекращается. Шm. В общем случае подобная процедура повторяется до тех пор, пока на каком-либо шаге не будет получено максимальное расстояние, для которого условие, определяющее выделение нового кластера, не выполняется. Оставшиеся образы относятся к тому кластеру, центр которого для них ближайший.

23. Алгоритм K внутригрупповых средних.

Ш1. Выбираются K исходных центров кластеров $Z_1(1), \dots, Z_K(1)$. Этот выбор производится произвольно.

Ш2. Заданное множество образов $\{x\}$ распределяется по K кластерам по следующему правилу:

$x \in S_j(k)$, если $\|x - z_j(k)\| < \|x - z_i(k)\|$ для всех $i = 1, 2, \dots, K, i \neq j$, где $S_j(k)$ — множество образов, входящих в кластер с центром $z_j(k)$. В случае равенства решение принимается произвольным образом. Ш3. На основе пред-го шага определяются новые центры кластеров $z_j(k+1)$, $j = 1, 2, \dots, K$, исходя из условия, что сумма квадратов расстояний между всеми образами, принадлежащими множеству $S_j(K)$, и новым центром кластера должна быть минимальной. Т. е. новые центры кластеров $z_j(k+1)$ выбираются таким

образом, чтобы минимизировать показатель качества $J_j = \sum_{x \in S_j(k)} \|x - z_j(k+1)\|^2, j = 1, 2, \dots, K$

Новые центры кластеров определяются как $z_j(k+1) = \frac{1}{N_j} \sum_{x \in S_j(k)} x, j = 1, 2, \dots, K$

где N_j — число выборочных образов, входящих в множество $S_j(k)$.

Ш4. Проверяется равенство $z_j(k+1) = z_j(k)$ при $j = 1, 2, \dots, K$, которое является условием сходимости алгоритма, и при его достижении выполнение алгоритма заканчивается. В противном случае алгоритм повторяется от шага 2. Качество работы алгоритмов, основанных на вычислении K внутригрупповых средних, зависит от числа выбираемых центров кластеров, от выбора исходных центров кластеров, от последовательности осмотра образов и от геом-х особенностей данных.

24. Алгоритм ISODATA.

$\{X_1, X_2, \dots, X_N\}$ — исходный набор образов. Ш1: Задаются параметры, определяющие процесс кластеризации:

K — необходимое число кластеров; Q_N — параметр, с которым сравнивается количество выборочных образов, вошедших в кластер; Q_s — параметр, характеризующий среднеквадратичное отклонение; Q_c — параметр, характеризующий компактность; L — максимальное количество пар центров кластеров, которые можно объединить; J — допустимое число циклов итерации. Ш2: заданные N образов распределяются по кластерам, соответствующим выбранным исходным центрам, по правилу

$x \in S_j$, если $\|x - z_j\| < \|x - z_i\|$, $i=1,2, \dots, N_c$; $i \neq j$, где

S_j -подмножество образов выборки, включённых в кластер с центром Z_j . Ш3: ликвидируются подмножества образов, в состав которых входит менее Q_N элементов, т. е. если для некоторого j выполняется условие $N_j < Q_N$, то подмножество S_j исключается из рассмотрения и значение N_c уменьшается на 1. Ш4: каждый центр кластера z_j , $j=1, 2, \dots, N_c$, локализуется и корректируется посредством приравнивания его выборочному среднему, найденному по соответствующему

подмножеству S_j , т. е. $z_j = \frac{1}{N_j} \sum_{x \in S_j} x$,

$j=1..N_c$, где N_j —число объектов, вошедших в подмножество S_j . Ш5: вычисляется среднее расстояние \bar{D}_j между объектами, входящими в подмножество S_j , и соответствующим центром кластера по формуле

$\bar{D}_j = \frac{1}{N_j} \sum_{x \in S_j} \|x - z_j\|$, $j=1..N_c$ Ш6: вычисляется обобщённое среднее расстояние между объектами,

находящимися в отдельных кластерах, и соответствующими центрами кластеров по ф-ле

$\bar{D} = \frac{1}{N} \sum_{j=1}^{N_c} N_j \bar{D}_j$ Ш7: а) если текущий цикл итерации – последний, то задается $Q_c=0$, переход к шагу

11. б) Если условие $N_c \leq K/2$ выполняется, то переход к шагу 8. в) Если текущий цикл итерации имеет четный порядковый номер или выполняется условие $N_c \geq K/2$, то переход к шагу 11; в противном случае процесс итерации продолжается. Ш8: для каждого подмножества выборочных образов с помощью

соотношения $\sigma_{ij} = \sqrt{\frac{1}{N_j} (x_{ik} - z_{ij})^2}$, $i=1..n$, $j=1..N_c$ вычисляется вектор среднеквадратичного отклонения

$\sigma_j = (\sigma_{1j}, \sigma_{2j}, \dots, \sigma_{nj})'$, где n есть размерность образа, x_{ik} , есть i -я компонента k -го объекта в подмножестве S_j , z_{ij} есть i -я компонента вектора, представляющего центр кластера z_j , и N_j —количество выборочных образов, включенных в подмножество S_c . Каждая компонента вектора среднеквадратичного отклонения σ_j характеризует среднеквадратичное отклонение образа, входящего в подмножество S_j , по одной из главных осей координат.

Ш9: в каждом векторе среднеквадратичного отклонения σ_j , $j=1, 2, \dots, N_c$, отыскивается максимальная компонента σ_{jmax} . Ш10: Если для любого σ_{jmax} , $j=1, 2, \dots, N_c$, выполняются условия $\sigma_{jmax} > Q_s$, и а) $\bar{D}_j > \bar{D}$ и $N_j > 2(QN+1)$, или б) $N_j < K/2$, то кластер с центром z_j *расщепляется* на два новых кластера с центрами z_j^+ и z_j^- соответственно, кластер с центром z_j ликвидируется, а значение N_c увеличивается на 1. Для определения центра кластера z_j^+ к компоненте вектора z_j , соот-й макс-й компоненте вектора σ_j , прибавляется заданная величина γ_j ; центр кластера z_j^- опред-я вычитанием этой же величины γ_j из той же самой компоненты вектора z_j . В качестве величины γ_j можно выбрать некоторую долю значения макс-й среднеквадратичной компоненты σ_{jmax} , т. е. положить $\gamma_j = k \sigma_{jmax}$, где $0 < k \leq 1$. При выборе γ_j следует руковод-я в основном тем, чтобы ее величина была достаточно большой для различения разницы в расстояниях от произвольного образа до новых двух центров кластеров, но достаточно малой, чтобы общая структура кластеризации существенно не изменилась. Если расщепление происходит на этом шаге, надо перейти к шагу 2, в противном случае продолжать выполнение алгоритма. Ш11: вычисляются расстояния D_{ij} между всеми парами центров кластеров:

$D_{ij} = \|z_i - z_j\|$, $i=1, 2, \dots, N_c-1$; $j=i+1, 2, \dots, N_c$. Ш12: Расстояния D_{ij} сравниваются с параметром Q_c . Те L расстояний, которые оказались меньше Q_c , ранжируются в порядке возрастания:

$[D_{i1j1}, D_{i2j2}, \dots, D_{iLjL}]$ причем $D_{i1j1} < D_{i2j2} < \dots < D_{iLjL}$. а L —максимальное число пар центров кластеров, которые можно объединить. Следующий шаг осуществляет процесс слияния кластеров. Ш13: кластеры с центрами z_{il} и z_{jl} , $i=1, 2, \dots, L$, объединяются (при усл, что в текущем цикле итерации процедура слияния не прим-ь ни к тому, ни к др-у кластеру), причем новый центр кластера опр-я по формуле

$$z^*_l = \frac{1}{N_{i_l} + N_{j_l}} [N_{i_l} (Z_{i_l}) + N_{j_l} (Z_{j_l})]$$

Центры кластеров z_{il} и z_{jl} ликвидируются и значение N_c уменьшается на 1.

Ш14: если текущий цикл итерации—послед-й, то выполнение алг-ма прекращ-ся. В противном случае следует возврат-я либо к ш1, если по предписанию польз-ля меняется к-л из параметров, определяющих процесс кластеризации, либо к ш2, если в очередном цикле итерации параметры процесса должны остаться неизменными. Завершением цикла итерации считается каждый переход к ш1 или 2.

25. Достоинства и недостатки алгоритмов обучения без учителя.

- необходимость проведения обширных экспериментов, трудность применения к данным сложной структуры. +простота реализации, применение эвристических подходов.

26. Изображающие числа и базис.

Базис - таблица, которая представляет все возможные комбинации значений истинности некоторого набора элементов A, B, C, \dots . Для n элементов A_1, \dots, A_n базис содержит n строк и 2^n колонок. Различных базисов существует $(2^n)!$. Тогда базис для одного элемента: $\#A=0\ 1$; для двух элементов A и B : $0\ 1\ 2\ 3$
 $\#A = 0\ 1\ 0\ 1\ \#B = 0\ 0\ 1\ 1$. Строки базиса наз-ся *изображающими числами соответствующих элементов* и обозн-ся приписыванием слева от элемента знака $\#$. *Операции над изображ-и фун-и:* 1) *изображающее число дизъюнкции двух элементов равно сумме изображающих чисел слагаемых: $\#(A + B) = \#A + \#B$, причем сложение $\#A, \#B$ выпол-я поразрядно без переносов в высшие разряды по правилу $0+0=0, 0+1=1+0=1, 1+1=1$.* 2) *изображающее число конъюнкции двух элементов опред-я как произведение их сомножителей: $\#(A \cdot B) = (\#A) \cdot (\#B)$, причем перемножение $\#A, \#B$ выпол-я поразрядно по правилу $0 \cdot 0=0, 0 \cdot 1=1 \cdot 0=0, 1 \cdot 1=1$.* 3) *ич отрицания $\neg A$ получается из ич A заменой в каждом разряде 0 на 1 и 1 на 0.*

27. Восстановление булевой функции по изображаемому числу.

ДНФ: пусть имеется мн-во, состоящее из n элементов A_1, \dots, A_n . Произведение вида $A_1 \cdot A_2 \dots A_{n-1} \cdot A_n$, составленное из элементов A_i , или их отрицаний $\neg A_i$ и содержащее n сомножителей, наз-ся элементарным произведением. Из n элементов можно составить 2^n различных элементарных произведений. Изображающее число каждого элементарного произведения имеет только одну единицу в одном из 2^n разрядов. СДНФ булевой функции — сумма элементарных произведений. Чтобы по данному изображаемому числу восстановить булеву функцию в СДНФ, нужно суммировать элементарные произведения, изображающие числа которых имеют единицы в тех же разрядах, что и изображающее число булевой функции. КНФ Элементарными суммами для n элементов A_1, \dots, A_n наз-ся суммы вида $A_1 + A_2 + \dots + A_{n-1} + A_n$, составленные из элементов A_i или их отрицаний $\neg A_i$ и содержащее n слагаемых. Из n элементов можно составить 2^n элементарных сумм. Ичисла элементарных сумм содержат только один 0 в одном из 2^n разрядов. Кнф булевой функции представляет собой произведение элементарных сумм. Для того, чтобы написать булеву функцию, соответствующую данному изображаемому числу в КНФ, необходимо перемножить элементарные суммы, изображающие числа которых имеют те же 0, что и изображающее число булевой функции.

28. Зависимость и независимость булевых функций..

Каждая булева функция может иметь два значения истинности, n булевых функций могут образовывать 2^n комбинаций значений истинности. n булевых функций $f_1(A, B, C, \dots), \dots, f_n(A, B, C, \dots)$ независимы, если в совокупности при всех возможных значениях аргументов A, B, C, \dots они могут принимать 2^n комбинаций значений истинности. Для проверки независимости функций $f_1(A, B, C, \dots), \dots, f_n(A, B, C, \dots)$ необходимо по отношению к базису $b[A, B, C, \dots]$ вычислить их изображающие числа $\#f_1(A, B, C, \dots) \dots \#f_n(A, B, C, \dots)$ и проверить, образуют ли столбцы этого набора 2^n чисел $0, 1, \dots, 2^n-1$; если 2^n чисел имеется, то функции независимы, в противном случае — зависимы.

29. Нахождение явного вида логической зависимости.

Чтобы найти явную форму логической связи зависимых функций $f_1(A, B, C, \dots) \dots f_n(A, B, C, \dots)$ в виде $F(f_1 \dots f_n) = I$ нужно:

1. В базисе $b[A, B, C, \dots]$ выписываются $\#f_1 \dots \#f_n$ и определяют какие числа отсутствуют в наборе столбцов. Столбцы набора $\#f_1(A, B, C, \dots) \dots \#f_n(A, B, C, \dots)$ представляют собой комбинации значений истинности функций $f_1 \dots f_n$, при которых соответствующие элементарные произведения истинны. Т.к. $\#(F=I) = \#F \Rightarrow$ столбцы указывают номера колонок в $b[f_1 \dots f_n]$, совпадающие с номерами разрядов $\#F(f_1 \dots f_n)$ на которых F истинна. \Rightarrow в соответствующих разрядах $\#F(f_1 \dots f_n)$ должны быть единицы, а в остальных — нули

30. Отыскание решений логического уравнения.

Логическое уравнение - алгебраическое уравнение, элементами которого являются логические числа и операции. Примером булевого уравнения с одним неизвестным может служить соотношение $X \cdot (A+B) = A \cdot B \cdot C$, где X — некоторая булева функция, зависящая от A, B, C , которую требуется найти, так чтобы в результате подстановки $X(A, B, C)$ в данное уравнение оно обращалось в тавтологию.

31. Техника решения логических уравнений с помощью булевых матриц.

Прямоугольная матрица называется булевой, если элементы ее — числа 0 и 1. Операции над булевыми матрицами: 1) Произведение булевых матриц $\|C_{ij}\| = \|a_{ik}\| * \|b_{kj}\|$ опр-ся по правилам обыкновенного матричного умножения, т. е. $C_{ij} = \sum_k a_{ik} \cdot b_{kj}$, с той только разницей, что операция суммирования произведений элементов строк и столбцов заменяется логическим сложением. 2) Соотношение импликаций: $\|a_{ij}\| \rightarrow \|b_{ij}\|$, справедливо для двух булевых матриц: $\|a_{ij}\|$ и $\|b_{ij}\|$, если нет i и j таких, что $\|a_{ij}\|=1$ и $\|b_{ij}\|=0$. 3) Транспонированной матрицей по отношению к $\|a_{ij}\|$ наз-т матрицу $\|b_{ij}\| = \|a_{ji}\|$, получаемую из матрицы $\|a_{ij}\|$ при перемене местами строк и столбцов. Для решения логического уравнения нужно: 1) транспонированную матрицу, состоящую из ич коэф-в одной части уравнения, умножить на матрицу, состоящую из изображающих чисел неизвестных этой же части уравнения. 2) тоже самое проделать с другой частью уравнения. 3) полученные матрицы построчно сравниваем между собой, и если элементы в столбцах совпадают, то вносим в таблицу результатов соответствующие значения неизвестных, по отношению к базису коэффициентов.

32. Две задачи о замене переменных в булевых функциях. (везде $\| \|$).

A, B, C, \dots — элементарные высказывания и совершается замена переменных $A = A(A', B', C', \dots)$, $B = B(A', B', C', \dots)$, $C = C(A', B', C', \dots)$, ...

1 задача: предположим, что задана нек-я функция $F_1(A, B, C, \dots)$ и совершается преобразование вида: A, B, C, \dots - элементарные высказывания заменяются на $A = A(A', B', C', \dots)$, $B = B(A', B', C', \dots)$, $C = C(A', B', C', \dots)$, ..., (*), где A', B', C', \dots - новые переменные. Требуется найти функцию $G_1(A', B', C', \dots) = F_1[A(A', B', C', \dots), B(A', B', C', \dots), C(A', B', C', \dots), \dots]$.

2 задача (обратная): найти такое преобразование переменных вида (*), которое переводило бы функции F_k в функции G_k , т.е. при всех $k = 1, 2, \dots$ $F_k[A(A', B', C', \dots), B(A', B', C', \dots), C(A', B', C', \dots), \dots] = G_k(A', B', C', \dots)$. В отличие от первой задачи, решение данной задачи существует не всегда и, кроме того, может быть неоднозначным. Решается исходя из: $|F_{ki}| * |R_{ij}| = |G_{kj}|$; $|F_{kj}| = |G_{ki}| * |R_{ji}|$; $|R_{ji}| = |R_{ij}|^T$

33. Прямая и обратная логические задачи распознавания.

Пусть вектор признаков подлежащего классификации объекта является бинарным, каждую компоненту которого можно рассматривать как пропозициональную переменную A_1, \dots, A_n . Пусть имеется m классов. Каждому классу поставим в соответствие величину — индикатор класса $\Omega_1, \dots, \Omega_m$ — тоже — пропозициональная переменная, принимающая значение 1, если соответствующий класс имеет место, и 0, если нет. Замечания: в нашей постановке не предполагается, что классы не могут иметь место одновременно. Пусть имеются нек-е заранее известные ограничения, наложенные на значения признаков A_1, \dots, A_n ; на индикаторы $\Omega_1, \dots, \Omega_m$; и на признаки, и на индикаторы классов (совместные ограничения). Причём ограничения определённого типа заданы в виде булевых уравнений. Для ограничений типа :

1. $\Phi_i(A_1, \dots, A_n) = 1 (\Phi_i^{-1}(A_1, \dots, A_n) = \Phi_i^{-2} A_1, \dots, A_n), i = 1..l$ (*)
2. $\Psi_j(\Omega_1, \dots, \Omega_m) = 1 (\Psi_j^{-1}(\Omega_1, \dots, \Omega_m) = \Psi_j^{-2}(\Omega_1, \dots, \Omega_m)), j = 1..p$
3. $\Theta_k(A_1, \dots, A_n; \Omega_1, \dots, \Omega_m) = 1 (\Theta_k^{-1}(A_1, \dots, A_n; \Omega_1, \dots, \Omega_m) = \Theta_k^{-2}(A_1, \dots, A_n; \Omega_1, \dots, \Omega_m)), k = 1..q$

Где все функции - правильно построенные высказывания от пропозиц-х переменных. Пусть получена некоторая информация об объекте—о этих переменных- в виде логического уравнения, которую можно записать в виде:

$$G(A_1, \dots, A_n) = 1. (**)$$

В частности м/б просто известны значения признаков A_1, \dots, A_n . Прямая задача распоз-я: Какой вывод можно сделать о классах w_1, \dots, w_m (точнее об их индикаторах) на основании инфо в (*) и (**)? Или: какое истинное высказывание о классах, обозначенных как $F(\Omega_1, \dots, \Omega_m) = 1$, можно сделать в

предположении истинности формулы (***) при условии выполнения ограничений (*)? Задача распознавания: найти такую функцию (правильно построенную формулу), которая является логическим следствием $G(A_1, \dots, A_n)$ при условии (*), т.е. $G(A_1, \dots, A_n) \models F(\Omega_1, \dots, \Omega_m)$.
Обратная зр: определить множество неизвестных посылок $G(A_1, \dots, A_n)$, из к-х следуют нек-е данные выводы $F(\Omega_1, \dots, \Omega_m)$ при ограничениях (*).

34. Пример логической задачи распознавания.

Пусть объект характ-ся единственным бинарным признаком A_1 ($n=1$).

$m=2$: $w_1, w_2 \Rightarrow \Omega_1, \Omega_2$. Априорная информация (система (**)):

- классы не могут иметь место одновременно: $\Omega_1 \neq \Omega_2$.

- значение признака однозначно определяет класс: $A_1 = \Omega_1$. Пусть об объекте известно: $A_1 = 0 \Leftrightarrow \neg A_1 = 1$ (**), $G \models \neg A_1$. Тогда система (***) имеет вид:

($\neg \neg A_1 \cup F(\Omega_1, \Omega_2) = 1 \& \Omega_1 \neq \Omega_2 \& A_1 = \Omega_1 \Rightarrow F(\Omega_1, \Omega_2) = \neg A_1$ (из 1),

$F(\Omega_1, \Omega_2) = \neg \Omega_1$ (из 3) $\Rightarrow F(\Omega_1, \Omega_2) = \Omega_2$.

35. Алгоритм вычисления оценок (АВО).

Пусть множество объектов $\{\omega\}$ подразделено на классы $\Omega_i, i=1, \dots, m$, и для описания объектов используются признаки $x_j, j=1, \dots, N$. Все объекты описываются одним и тем же набором признаков. Алгоритм распознавания сравнивает описание распознаваемого объекта с описаниями всех объектов и принимает решение о том, к какому классу отнести объект. Классификация основана на вычислении степени похожести (оценки) распознаваемого объекта на объекты, принадлежность которых к классам известна. Эта процедура включает в себя два этапа: сначала подсчитывается оценка для каждого объекта, а затем полученные оценки используются для получения суммарных оценок по каждому из классов Ω_i . *Решающим правилом* является отнесение распознаваемого объекта к классу, которому соответствует максимальная оценка, либо эта оценка превышает оценки всех остальных не менее чем на определенную пороговую величину, либо отношение соответствующей оценки к сумме оценок всех остальных классов не менее пороговой величины.

36. Оsn. идеи, лежащие в осн. АВО. Сущность алгебраического подхода. Рассмотрим полный набор признаков $x = \{x_1, \dots, x_N\}$ и выделим систему подмн-в мн-ва признаков (систему опорных мн-в алгоритма) S_1, \dots, S_l . В АВО обычно рассмат-ся либо все подмн-ва мн-ва признаков фиксированной длины $k, k=2, \dots, N-1$, либо вообще все подмн-ва мн-ва признаков. Удалим произвольный поднабор признаков из строк $\omega_1, \omega_2, \dots, \omega_m$, ω' и обозначим полученные строки $S_{\omega_1}, S_{\omega_2}, \dots, S_{\omega_m}, S_{\omega'}$. Правило близости, позвол-е оценить похожест- строки $S_{\omega'}$, соответс-й распознаваемому объекту ω' , и строки S_{ω_k} , соответс-й произв-му объекту i -й таблицы, состоит в следующем. Пусть «усеченные» строки содержат q первых признаков, т.е. $S_{\omega_k} = (\alpha_1, \dots, \alpha_q)$ и $S_{\omega'} = (\beta_1, \dots, \beta_q)$, и заданы пороги $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_q, \delta$. Строки S_{ω_i} и S_{ω_k} считаются похожими, если выпол-ся не менее чем δ неравенств вида $|\alpha_j - \beta_j| < \varepsilon_j, j=1, \dots, q$. Величины $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_q, \delta$ входят в качестве параметров в АВО. Рассм-м процедуру вычисления оценок по подмножеству S_i . Проверяется близость строки $S_{\omega'}$ со строками $S_{\omega_1}, S_{\omega_2}, \dots, S_{\omega_{l-1}}$, принадлежащими объектам класса Ω_1 . Число строк этого класса, близких по выбранному критерию классиф-й строке $S_{\omega'}$, обозн-ся $\Gamma_{S_i}(\omega', \Omega_1)$; последняя величина представляет собой оценку строки ω' для класса Ω_1 по опорному множеству S_i . Аналогичным образом вычисляются оценки для остальных классов: $\Gamma_{S_1}(\omega', \Omega_1), \dots, \Gamma_{S_l}(\omega', \Omega_m)$

Применение подобной процедуры ко всем остальным опорным множествам алгоритма позволяет получить систему оценок $\Gamma_{S_2}(\omega', \Omega_1), \dots, \Gamma_{S_2}(\omega', \Omega_m), \dots, \Gamma_{S_k}(\omega', \Omega_1), \dots, \Gamma_{S_k}(\omega', \Omega_m)$.

$$\Gamma(\omega', \Omega_1) = \Gamma_{S_1}(\omega', \Omega_1) + \Gamma_{S_2}(\omega', \Omega_1) + \dots + \Gamma_{S_k}(\omega', \Omega_1) = \sum_{si \in S_A} \Gamma_{si}(\omega', \Omega_1)$$

$$\Gamma(\omega', \Omega_m) = \Gamma_{S_1}(\omega', \Omega_m) + \Gamma_{S_2}(\omega', \Omega_m) + \dots + \Gamma_{S_k}(\omega', \Omega_m) = \sum_{si \in S_A} \Gamma_{si}(\omega', \Omega_m)$$

Величины $\Gamma(\omega', \Omega_1), \dots, \Gamma(\omega', \Omega_m)$ представляют собой оценки строки ω' для соответствующих классов по системе опорных множеств алгоритма S_A . Реш-е правило может принимать разл-е формы, в частности распозн-й объект м/б отнесен к классу, кот-у соответ-т макс-я оценка, либо эта оценка будет превышать оценки всех ост-х классов не меньше чем на опред-ю пороговую величину n_1 , либо значение отношения

соответ-й оценки к сумме оценок для всех остальных классов будет не менее значения порога n_2 и т. д. Параметры типа n_1 и n_2 также включаются в АВО.

37. Перцептрон и его математическая модель.

Перцептрон-устройство, моделирующее работу человеческого мозга. Состоит из : а) S_1, \dots, S_n – рецепторы (воспринимают сигналы из окруж-й среды).

б) A_1, \dots, A_n – А-элементы (каждый из них соединён с группой s-элементов, воспринимая их сигналы, выраб-т вых-й сигнал, если общее воздействие s-э превосходит нек-й заданный порог). в) R-э (воспринимает сигнал от А-э, выраб-т вых-й сигнал, кот-й явл-ся линейной комбинацией а-э).

$$R_i = \sum_{j=1}^k w_{ij} y_j \quad \text{Обучение перц-на свод-ся к установлению таких}$$

значений w_i , при кот-х маркир-я ОП правильно классиф-ся перц-м. Замечание:

различ-е модели перц-в отлич-ся друг от друга: 1)восприним-й способн-ю S-э.

2)св-ми А-э. 3)системами соедин-я элементов. 4)кол-м слоёв элем-в. 5)проц-ми обучения перцепторна.

Матмодель перц-на: S-э поставим в соответ-е вектор $x=(x_1, \dots, x_n)$; А-э –вектор $y=(y_1, \dots, y_{n1})$. $y=(y_1, \dots, y_{n1})^T = \varphi(x) = (\varphi_1(x), \dots, \varphi_{n1}(x))^T$.

Решение об отнесении x к одному из классов принимается так: 1) если $\sum_{i=1}^{n1} w_i y_i$ (в матр виде $w^T y$) $> 0 \Leftrightarrow \sum_{i=1}^{n1} w_i \varphi_i(x) > 0 \Rightarrow x$ принадл-т W_1 . 2)если $< 0 \Leftrightarrow < 0 \Rightarrow x$ прин-т w_2 . Геометричес-я интерпрет-я: если $x \in X^n$, то $w^T \varphi(x) = 0$ –определяет разделяющую гиперповерхность в пространстве X . Если X лежит по одну сторону раздел-й гиперпов-ти, то $x \in w_1$, и наоборот. $x \in X^n \Leftrightarrow y \in Y^{n1}$, то $w^T y = 0$ –ур-е плоскости, проход-й ч/з начало координат. Тогда решение: 2 полупрост-ва (для w_1 и w_2), простр-во Y^{n1} – спрямляющее пространство.

38. Алгоритм обучения перцептрона.

$y=(y^1, \dots, y^N)$ –векторы из спрямляющего пр-во Y принадлежащие w_1 или w_2

Задачи обучения перцептрона: на данной обуч. Посл. найти $w=(w_1, \dots, w_k)$ с помощью которой данная обуч. посл. Y классифицируется безошибочно.

Алг.: на k -м шаге:

если $y(k) \in w_1$ и $w^T(k) y(k) \leq 0$, то $w(k+1) = w(k) + c y(k)$

если $y(k) \in w_2$ и $w^T(k) y(k) \geq 0$, то $w(k+1) = w(k) - c y(k)$

иначе $w(k+1) = w(k)$

c –корректирующее приращение

Останов. когда вся обуч. последовательность распознана правильно при неизменном в-ре весов.

Замечания: 1) приведенный алгоритм реализует принцип подкрепления и наказания.

2) при построении модели предполагалось, что распрямляющая плоскость проходит через 0, но реально м-т оказ-ся иначе. Это испр-ляется путем ввода доп. координаты $y=(y_1, y_2, \dots, y_{n1}, 1)$.

3) преобразуем обуч. посл-ть Y в $\bar{Y} = (\bar{y}^1, \dots, \bar{y}^N)$, где $\bar{y}^i = \begin{cases} y^i, & y^i \in \omega_1 \\ -y^i, & y^i \in \omega_2 \end{cases}$

тогда алгоритм проще: если $\bar{y}(k) \in w_1$ и $w^T(k) \bar{y}(k) \leq 0$, то $w(k+1) = w(k) + c \bar{y}(k)$ иначе $w(k+1) = w(k)$

39. Сходимость алгоритма обучения перцептрона. Теорема Новикова.

Выпуклой оболочкой мн-ва наз-ся наименьшее вып-е мн-во, содержащее данное множество. Выпуклая оболочка- U -е всех выпуклых множеств, содержащих данное множество. Выпук-й оболоч-й конечного мн-ва точек будет выпуклый многогранник (n -мерный симплекс), содерж-й все эти точки, вершины кот-го совпадают с нек-ми (возможно со всеми) точками исходного мн-ва. Точки ОП м\б разделены перцептроном на классы w_1 и w_2 , если в спрямл-м прост-ве существует раздел-я гиперплоскость, т.е. $\exists W$,

что $w^T \tilde{y}^i > \rho_0 > 0$ для любого $i=1..N$. **Теорема Новикова:** пусть 1)имеется ∞ преобраз-я ОП

$\tilde{Y} = \{\tilde{y}^1, \dots, \tilde{y}^i, \dots\}$, элементы кот-й относятся как к классу w_1 , так и w_2 . 2)в СП $Y \exists$ разд-я

гиперплоскость, т.е. \exists такой един-й век-р $|w^*|=1$, что $w^{*T} \tilde{y}^i > \rho_0$ для $\forall i$. 3) величина $D < \infty$ (конечна).

Тогда при нач. в-ре $w(1)=0$ и корр. прир. $C=1$ алгоритм обучения перцептрона сходится, причем число

$$\text{исправлений вектора весов } k \leq \left\lceil \frac{D^2}{\rho_0^2} \right\rceil$$

40. Итеративные процедуры распознавания на основе градиентных методов: минимизация одной функции.

$Z=(Z^1,...,Z^n)$ – вектор признаков. Даны 2 класса: W_1 и W_2 ($m=2$). Функция $f(\alpha, Z) > 0$, если $Z \in W_1$, < 0 – $Z \in W_2$, $\alpha=(\alpha_1,...,\alpha_k)$ -вектор нек-х параметров. ОП $Z=(Z^1,..., Z^N)$. Найти вектор α^* , для кот-го $\Delta: f(\alpha^*, Z^i) > 0$, если $Z^i \in W_1$, для $i=1..N$; < 0 – $Z^i \in W_2$ для $i=1..N$. Возможна ситуация: \exists одна функция $\Phi(\alpha; Z^1,..., Z^N)$ такая, что $\Phi(\alpha^*; Z^1,..., Z^N) = \min_{\alpha} \Phi(\alpha; Z^1,..., Z^N) \Leftrightarrow \Delta$. Алгоритм поиска \min одной функции: $\alpha(K)$ -значение вектора α на k -том шаге работы алгоритма. $\alpha(K+1) = \alpha(K) - c \text{grad } f(\alpha)$ по $\alpha = \alpha(K)$, c -величина шага grad . Зададим $\alpha(0)$. Его сходимость зависит от вида функции f , величины шага c . Далее решить задачу мин-и с помощью подходящей модификации метода град-го спуска, т.е. решить задачу Δ .

41. Итеративные процедуры распознавания на основе градиентных методов: совместная минимизация нескольких функций.

Пусть изображение хар-ся вектором признаков $Z=(Z^1,...,Z^n)$. Рассмотрим случай 2-х классов W_1 и W_2 ($m=2$). Предположим, что \exists функ-я $f(\alpha, Z) > 0$, если $Z \in W_1$, < 0 – $Z \in W_2$, где $\alpha=(\alpha_1,...,\alpha_k)$ -вектор нек-х параметров. ОП $Z=(Z^1,..., Z^N)$, элементы кот-й как из w_1 , так из w_2 . Найти вектор α^* , для кот-го $\Delta: f(\alpha^*, Z^i) > 0$, если $Z^i \in W_1$, для $i=1..N$; < 0 – $Z^i \in W_2$ для $i=1..N$. Возможна ситуация: $\exists N$ функций $F(\alpha, Z^i)$, $i=1..N$, таких что каждая из них достигает минимума в т. α^* , так что $F(\alpha^*, Z^i) = \min_{\alpha} F(\alpha, Z^i)$ для всех $i=1..N$ тогда и только тогда, когда выполнено Δ . Предположим, что $F(\alpha, Z^i)$, $i=1..N$ со св-ми: 1) они непрерывны и дифф-мы по $\alpha_1,...,\alpha_k$; 2) имеют каждая единственный минимум (но не обяз-но единств-ую тчк минимума). Тогда используем понятие градиента для отыскания тчк экстремума функции. Нужно определить стратегию применения алгоритмов град-го спуска к набору функций $F(\alpha, Z^i)$, $i=1..N$. Стратегия1: провести град-й спуск для $F(\alpha, Z^1)$, начав с нек-й тчк $\alpha^0 \Rightarrow$ получили тчк мин-ма этой функ. α^1 (далее эта тчк начальная) \Rightarrow для $F(\alpha, Z^2)$, не выходя из области наим-го зн-я $F(\alpha, Z^1) \Rightarrow$ тчк совместного экстремума для $F(\alpha, Z^1)$, $F(\alpha, Z^2)$ и т.д. $\Rightarrow \alpha^{N-1}$ (т. мин всех предыдущих функций) $\Rightarrow F(\alpha, Z^N) \Rightarrow \alpha^*$ -решение. Стратегия2: провести град-й спуск для $F(\alpha, Z^1)$, начав с нек-й тчк $\alpha(0) \Rightarrow$ вектор $\alpha(1) \rightarrow \text{grad } F(\alpha, Z^2), \dots, \alpha(N-1) \rightarrow \text{grad } F(\alpha, Z^N) \Rightarrow \alpha(N) \rightarrow \alpha(0)$ (в качестве нач тчк), если не все $\text{grad}=0$. Схема-стоп, когда найдена т. α^* , в кот-й $\text{grad } F(\alpha^*, Z^i) = 0$ для всех $i=1..N$. -: отыскание точки α^* не гарантировано, даже если она \exists . Но для нек-х видов функ-и F применение этих стратегий оказывается успешным.

42. Алгоритм обучения перцептрона как реализация спецстратегии совместной минимизации неск-х функций с помощью градиентных методов.

Установим соответствия: $Z=(Z^1,...,Z^n) \Rightarrow \tilde{y}=(\tilde{y}^1,...,\tilde{y}^n)$; $\alpha=(\alpha_1,...,\alpha_k) \Rightarrow (k=n) W=(W^1,...,W^{n-1},W^{n+1})$; $f(\alpha, Z) \Leftrightarrow W^T \tilde{y}$. ОП $Z=\{Z^1,..., Z^N\} \Rightarrow Y=\tilde{y}=(\tilde{y}^1,...,\tilde{y}^N)$. Задача заключается в отыскании такого набора значений w^* , для кот-го $w^{*T} \tilde{y}^i > 0$, $i=1,...,N$ (**). Рассмотрим функции: $F(w^T, \tilde{y}^i) = 1/2(|w^T \tilde{y}^i| - w^T \tilde{y}^i) = 1/2(|\sum_{j=1}^{n+1} W_j \tilde{y}_j^i| - \sum_{j=1}^{n+1} W_j \tilde{y}_j^i)$, $i=1..N$. Их св-ва: 1) $F(w^T, \tilde{y}^i) \geq 0$; 2) $\min F(w^T, \tilde{y}^i) = 0$ и дост-ся титт, когда $(w^T, \tilde{y}^i) > 0$, $i=1..N$, если только \exists раздел-я гиперплоск-ть $w^T \tilde{y}^i = 0$. Для каждой ф-и $F(w, \tilde{y}^i)$ АГС имеет вид: $W(k+1) = W(k) - C/2(\text{sign}(w^T, \tilde{y}^i) \tilde{y}^i - \tilde{y}^i)$, т.к. $\text{grad}_w F(w, \tilde{y}^i) = 1/2(\text{sign}(w^T, \tilde{y}^i) \tilde{y}^i - \tilde{y}^i)$, где $\text{sign}(w^T, \tilde{y}^i) = 1$, $w^T \tilde{y}^i > 0$; -1 , $w^T \tilde{y}^i \leq 0$. Поэтому $W(k+1) = W(k) - C/2(\tilde{y}^i - \tilde{y}^i) = W(k)$, если $w^T(k) \tilde{y}^i > 0$; $W(k+1) = W(k) - C/2(-\tilde{y}^i - \tilde{y}^i) = W(k) + c \tilde{y}^i$, если $w^T(k) \tilde{y}^i \leq 0$. Таков АГС для данной ф-и $F(w, \tilde{y}^i)$. Пусть для поиска совместного минимума функций $F(w, \tilde{y}^i)$, $i=1..N$, используется стратегия, когда очередной шаг данного алгоритма исп-ся для след-й в последовательности функции, а W_k -значение вектора весов, полученное для предшествующей функции. Обозначим ч/з $\tilde{y}(k)$ то значение \tilde{y}^i , кот-е используется на шаге $k+1$, получим $W(k+1) = W(k) + \tilde{y}(k)$, если $w^T(k) \tilde{y}(k) \leq 0$; иначе $W(k+1) = W(k)$. Алгоритм-стоп, если найден набор w^* , для кот-го все ф-и $F(w, \tilde{y}^i)$ принимают мин-е значение. При $c=1$ имеем АО перцептрона, сходимость кот-го гарантируется теоремой Новикова.

43. Физическая интерпретация метода потенциальных функций.

Пусть нужно разделить два класса ω_1 и ω_2 . Выборочные образы, представлены векторами или точками в n -мерном пространстве образов. Если ввести аналогию м/у точками, представл-ми выборочные образы, и нек-м источником энергии, то в любой из этих точек потенциал достигает макс-го значения и быстро уменьшается при переходе во всякую точку, отстоящую от точки, предст-й выборочный образ x_k . На основе этой аналогии можно допустить существование эквипотенциальных контуров, которые

описываются потенциальной функцией $K(x, x_k)$. Можно считать, что кластер, образ-й выбор-ми образами, принадл-ми классу ω_1 , образует «плато», приче-е образы размещаются на вершинах нек-й группы холмов. Подобную геометрическую интерпретацию можно ввести и для образов класса ω_2 . Эти два «плато» разделены «долиной», в кот-й, как считается, потенциал падает до нуля. На основе таких интуитивных доводов создан метод потенциальных функций, позвол-й при проведении классификации определять решающие функции.

44. Кумулятивный потенциал. Алгоритм итеративного вычисления кумулятивного потенциала.

В начале этапа обучения исходное значение к. п. $K_0(x)$ полагается=0. При предъявлении первого образа x_1 из обуч. выборки значение к. п. просто = зн-ию потенц. ф-ии для выбороч. образа x_1 . Потенциал предполагается положительным для образов, принадлежащих классу ω_1 , и отрицательным для образов, принадлежащих классу ω_2 .

На $k+1$ шаге $K_{k+1}(x) = K_k(x) + r_{k+1}K(x, x_{k+1})$

где коэфф-ты r_{k+1} при корректирующем члене опр-ся соотношениями

$$r_{k+1} = \begin{cases} 0 & \text{при } x_{k+1} \in \omega_1 \text{ и } K_k(x_{k+1}) > 0 \\ 0 & \text{при } x_{k+1} \in \omega_2 \text{ и } K_k(x_{k+1}) < 0 \\ 1 & \text{при } x_{k+1} \in \omega_1 \text{ и } K_k(x_{k+1}) \leq 0 \\ -1 & \text{при } x_{k+1} \in \omega_2 \text{ и } K_k(x_{k+1}) \geq 0 \end{cases}$$

45. Теоремы о сходимости обучения классификации методом потенциальных функций.

Теорема 1. (О свойствах сходимости алгоритма.) Пусть векторы образов x удовлетворяют в пространстве образов следующим условиям.

1. Потенциальная функция

$$K(x, x_j) = \sum_{i=1}^{\infty} \lambda_i^2 \varphi_i(x) \varphi_i(x_j)$$

ограничена для $x \in T_1 \cup T_2$.

2. Существует решающая функция, представимая в виде

$$d(x) = \sum_{i=1}^m c_i \varphi_i(x), \quad (*)$$

такая, что

$$d(x) \begin{cases} > \varepsilon, x \in \omega_1, \\ < -\varepsilon, x \in \omega_2, \end{cases} \quad (**)$$

где $\varepsilon > 0$.

3. Обучающая выборка образов обладает следующими статическими свойствами: (а) в обучающей последовательности выборочные образы появляются независимо; (б) если на k -м шаге алгоритма обучения решающая функция $d_k(x)$ не обеспечивает правильной классификации всех образов x_1, x_2, \dots, x_k , то с положительной вероятностью будет предъявлен образ x_{k+1} , корректирующий ошибку.

Тогда с вероятностью 1 можно определить конечное число шагов R , таких, что кумулятивный потенциал

$$K_R(x) \begin{cases} > 0, x \in \omega_1, \\ < 0, x \in \omega_2. \end{cases}$$

Другими словами, последовательная аппроксимация решающей функции $d_k(x)$ с вероятностью 1 сходится к решающей функции $d(x)$ за конечное число предъявлений образов обучающей выборки. Это означает, что разделение классов ω_1 и ω_2 осуществляется за конечное число шагов с вероятностью 1.

Теорема (условия прекращения выполнения алгоритма): пусть процесс обучения прекращается, если после осуществления k коррекций неправильной классификации при предъявлении L_0 следующих выборочных образов никакие коррекции больше не производятся. Т.е. процесс обучения прекращается после предъявления L_k выборочных образов, где L_k определяется выражением $L_k = L_0 + k$. Общее число предъявлений образов, необходимое для прекращения работы алгоритма, увеличится на 1 после каждой коррекции неправильной классификации. Задача заключается в определении числа контрольных выборочных образов L_0 , необходимых для обеспечения заданного качества процедуры обучения.

Обозначим через $P_{L_k}(e)$ вероятность совершения ошибки после предъявления системе L_k выборочных образов. Тогда для любых $\varepsilon > 0$ и $\delta > 0$ вероятность того, что $P_{L_k}(e) < \varepsilon$, будет больше, чем $1 - \delta$, если

$$L_0 > \frac{\log \varepsilon \delta}{\log(1 - \varepsilon)}.$$

Отметим, что выбор числа контрольных выборочных образов зависит только от заданных значений ε и δ , характеризующих качество обучаемой процедуры. Выбор величины L_0 не зависит от свойств классов ω_1 и ω_2 и статистических характеристик образов.

46. Достоинства и недостатки распознающих процедур перцептронного типа.

+: возможность обучения при определенных условиях безошибочной классификации объектов; универсальность; ОП МВ в достаточной степени произвольной; результат обучения не зависит от начального состояния перцептрона.

-: для решения сложных ЗР может потребоваться перцептрон с большим числом элементов; длина ОП может оказаться очень большой; для сложных задач невозможно проверить условия, при которых ЗР с помощью перцептрона разрешена, а также условия сходимости АОП, которые описаны в тн Новикова.

47. Элементы задачи решения.

Введем в рассмотрение следующие элементы задачи принятия решения:

$\Omega = \{w_1, \dots, w_m\}$ – конечное множество m состояний природы (ситуаций)

$D = \{d_1, \dots, d_p\}$ – конечное множество p возможных решений (действий)

$L(d_i, w_j)$ – функция потерь, которую интерпретируется следующим образом: потери от принятия решения d_i , когда реализуется состояние w_j .

$X = (x_1, \dots, x_n)$ – n -мерный признак. Будем считать, что он является реализацией системы СВ $X = (X_1, \dots, X_n)$, которая для состояний природы w_j характеризуется:

- Плотностью вероятности $p(x/w_j)$, если X – система абстрактных непрерывных СВ

- Законом распределения (набором вероятностей) $P(x/w_j) = P(X=x/w_j)$, если X – система дискретных СВ.

Пусть $P(w_j)$ – вероятность появления состояния природы w_j и пусть $P(w_j/x)$ – условная вероятность появления состояния природы w_j при условии, что будет наблюдаться значение случайного вектора $X=x$.

$$P(w_j | X=x) = \begin{cases} \frac{p(x/w_j) * P(w_j)}{p(x)}, & \text{если } X \text{ - САНСВ} \\ \frac{P(X=x/w_j) * P(w_j)}{P(X=x)}, & \text{если } X \text{ - СДСВ} \end{cases}$$

$$p(x) = \sum_{j=1}^m p(x/w_j) * P(w_j)$$

$$P(x) = \sum_{j=1}^m P(X=x/w_j) * P(w_j)$$

48. Условный риск. Общий риск.

Введем вел-ну

$$R(d_i / X = x) = \sum_{j=1}^m L(d_i, w_j) P(w_j |_{X=x}) - \text{усл-е мат. ожидание потерь от принятия реш-я } d_i \text{ при усл-и, что } X=x.$$

-усл. риск

Введем в рассм-е функцию $d(x)$ – каждому в-ру x ставит в соотв-е некот реш-е из мн-ва реш-й D .

$$R(d(x) |_{X=x}) = \sum_{j=1}^m L(d_i, w_j) P(w_j |_{X=x}) - \text{усл. мат. ожид-е потерь при усл-и, что исп. реш. ф-я } d(x)$$

Безусловное мат ожид-е потерь при использовании решающего правила $d(x)$ есть

$$R(d) = \begin{cases} \sum_x R(d(x) |_{X=x}) P(X=x), \text{ если } X - \text{ДДСВ} \\ \int_x R(d(x) |_{X=x}) p(x) dx, \text{ если } X - \text{САНСВ} \end{cases}$$

$R(d)$ назыв. общим риском.

49. Постановка байесовской задачи решения. Оптимальное решающее правило. Связь с задачей распознавания.

Поставим следующую задачу: найти решающее правило $d(x)$ такое, которое доставляет минимум $R(d)$. Эта задача решается очень просто. Т.к. $R(d)$ будет минимально, если $d(x)$ выбрано так, что $R(d(x)/x)$ имеет минимальное значение для каждого возможного x , то оптимальное решающее правило можно описать следующим образом: для данного x вычислить $R(d_i/x)$ для $i=1, \dots, p$ и выбрать то d_i , при котором $R(d_i/x)$ минимально, т.е.

$$d_{\text{opt}}(x) = \arg \min_{d_i \in D} R(d_i / X = x)$$

Такое решающее правило называется байесовским решающим правилом, а соответствующее минимальное значение общего риска $R(d_{\text{opt}})$ – байесовским риском.

Задача распознавания образов получается, если между элементами множества Ω и D установим взаимнооднозначное соответствие

Решение d_i заключается в отнесении объекта, имеющего изображение x к одному из классов w_1, w_2, \dots, w_m . (в этом случае $p=m$)

50. Классификация с минимальной вероятностью ошибки.

Рассм частный случай байесовской процедуры распознавания, когда потери от принятия любого неверного решения одинаковы (любое неверное решение одинаково нежелательно). В этом случае потери можно представить в виде

$$L(d_i = w_i, w_j) = \begin{cases} 0, & w_i = w_j \\ 1, & w_i \neq w_j \end{cases}$$

Для заданной L байесовское решающее правило будет обеспечивать минимальную вероятность ошибки классификации.

51. Минимаксное решающее правило.

$$\begin{aligned} d_{\text{opt}}(x) &= w_{\text{opt}}(x) = \\ &= \arg \min_{w_i \in w_1, w_m} R(d_i = w_i / X = x) = \\ &= \arg \min_{w_i \in w_1, w_m} [1 - P(w_i / X = x)] = \\ &= \arg \max_{w_i \in w_1, w_m} P(w_i / X = x) \end{aligned}$$

52. Процедура Неймана-Пирсона.

$f(x) - 3P, H_0: x \in w_1$ или $f(x) \cdot g(x) - 3P, H_1: x \in w_2$ или $g(x)$. Если в результате проведения наблюдения $\frac{g(x)}{f(x)} \gg c$, то H_0 отвергаем, т.е. $x \rightarrow w_2$ (в случае X-С ДСВ); Если $\frac{1}{n} \sum_{i=1}^n (x_i - a_0) \geq \frac{S_\alpha}{\sqrt{n}}$, то гипотезу H_0 отвергаем, т.е. $x \rightarrow w_2$ (в случае X-С НСВ).

54. Обучение байесовской процедуры распознавания: оценка неизвестных параметров.

1) вид любого УЗР известен, но ЗР содержат неизвестные параметры. Тогда это параметр-я задача обучения, решение кот-й сводится к оценке неизвестных параметров с помощью статист-х методов (максимального правдоподобия, метод моментов). 2) вид УЗР неизвестен. Обучение: а) представление неизвестного ЗР в виде ряда разложения по системе определённых функций с последующим определением по данным наблюдения оценок коэф-в ряда, отбрасывания ряда. б) минимизация общего риска: $R(d) \rightarrow \tilde{R}(d)$ решается заданием $\tilde{R}(d) \rightarrow \min$

55. Оценка неизвестной плотности вероятности по априорной информации.

Если плотность распределения СВ неизвестна, то, согласно принципу максимума энтропии следует выбирать такую плотность распределения, которая обеспечивает максимизацию энтропии СВ с учетом всех известных ограничений. Энтропия совокупности образов с плотностью распредел-я $p(x)$:

$H = \int_x p(x) \ln p(x) dx$ под $p(x)$ имеется в виду $p(x|w_i)$. Априорная информация об СВ задается в виде:

$\int_x p(x) dx = 1$ (1а) и $\int_x b_k(x) p(x) dx = a_k, k = 1, \dots, Q$ (1б) Задача состоит в том, чтобы задать такое

распределение $p(x)$, чтобы величина энтропии при условиях (1а) и (1б) была максим-й. Используя

множители Лагранжа $\lambda_0, \dots, \lambda_Q$ построим функцию $H_1 = \int_x p(x) [\ln p(x) - \sum_{k=0}^Q \lambda_k b_k(x)] dx - \sum_{k=0}^Q \lambda_k a_k$, где

$a_0=1, b_0(x)=1$ для всех образов x . Взяв частные производные от функции H_1 по плотности

распределения $p(x)$ имеем: $\frac{\partial H_1}{\partial p(x)} = -\int_x \{[\ln p(x) - \sum_{k=0}^Q \lambda_k b_k(x)] + 1\} dx$ приравняв подынтегральное

выражение 0 и выразив $p(x)$ получим: $p(x) = \exp[\sum_{k=0}^Q \lambda_k b_k(x) - 1]$ Здесь параметры $\lambda_0, \dots, \lambda_Q$ следует

выбирать так, чтобы они соответствовали априорной информации об образах x , содержащейся в соотношениях (1а) и (1б). Равномерное распределение. Выбираем, когда известно, что СВ отлична от нуля только в конечном интервале. Нормальное распределение. Выбираем, когда известно, что СВ может принимать любое действительное значение.

56. Оценка неизвестной плотности вероятности с использованием экспериментальных данных.

При неизвестном типе распределения СВ требуется оценить плотность распределения. Пусть $p'(x)$ – оценка плотности вероятности $p(x)$, причем под $p(x)$ принимаем $p(x|w_i)$. Найдём такую оценку, которая обеспечила бы минимизацию среднеквадратичной ошибки R , определяемой как

$R = \int_x u(x) [p(x) - p'(x)]^2 dx$, где $u(x)$ – весовая функция (1). Разложим оценки $p'(x)$ в ряд:

$p'(x) = \sum_{j=1}^m c_j \varphi_j(x)$ (2) c_j – коэффициенты, подлежащие определению, $\{\varphi_j(x)\}$ – множество заданных

базисных функций, m – число членов разложения. (2)->(1): $R = \int_x u(x) [p(x) - \sum_{j=1}^m c_j \varphi_j(x)]^2 dx$ Найти

такие коэффициенты c_j , которые обеспечат миним-ю интеграла вероятности ошибки. Необходимое

условие минимальности: $\frac{\partial R}{\partial c_k} = 0, k = 1, \dots, m$ $\sum_{j=1}^m c_j \int_x u(x) \varphi_j(x) \varphi_k(x) dx = \int_x u(x) \varphi_k(x) p(x) dx$ (3)

Правая часть – мат. ожидание функции $u(x)\varphi_k(x)$. Матожидание аппроксимируем выборочным средним:

$\int_x u(x) \varphi_k(x) p(x) dx \approx \frac{1}{N} \sum_{i=1}^N u(x_i) \varphi_k(x_i)$ Подставим в (3): $\sum_{j=1}^m c_j \int_x u(x) \varphi_j(x) \varphi_k(x) dx =$

$\frac{1}{N} \sum_{i=1}^N u(x_i) \varphi_k(x_i)$ (4) Если базисные функции $\{\varphi(x)\}$ выбраны так, что они ортогональны весовой

функции $u(x)$, то из определения ортогональности : $\int_x u(x) \varphi_k(x) p(x) dx = \begin{cases} A_k, j = k \\ 0, j \neq k \end{cases}$ (5). (5) в (4)

$c_k = \frac{1}{NA_k} \sum_{i=1}^N u(x_i) \varphi_k(x_i), k = 1, 2, \dots, m$ Если базисные функции ортонормированны, то $A_k=1$ для всех k .

Кроме того, т.к. члены $u(x_i)$ не зависят от k , и одинаковы для всех коэффициентов, то их можно исключить из аппроксимирующего выражения. Результирующая формула для коэффициентов

следующая: $c_k = \frac{1}{N} \sum_{i=1}^N \varphi_k(x_i), k = 1, \dots, m$, где N - количество элементов в выборке. После вычисления

коэффициентов можно вычислить оценку плотности распределения по формуле (2). Замеч1: качество аппроксимации с помощью выбранной системы базисных функций зависит от числа m членов разложения. 2: Выбор базисных функций сильно влияет на простоту решения. При отсутствии априорных сведений о характере плотности распределения $p(x)$ базисные функции в первую очередь должны выбираться исходя из простоты реализации.

57. Правило ближайшего соседа как пример непараметрического метода распознавания: оценка вероятности ошибки классификации.

Вероятность ошибки классификации БС - правила не более чем вдвое больше минимальной вероятности ошибки классификации, обеспечиваемой байесовским решающим правилом:

$$P_e^* \leq P_e \leq P_e^* (2 - \frac{m}{m-1} P_e^*) \leq 2P_e^*$$

58. Основы МГУА: опорные функции, обуч-я и проверочная послед-ти, крит. селекции, схема.

Главной характеристикой алгоритмов МГУА является вид элементарной функции y : линейный полином

$$y_i = a_0 + a_1 x_j + a_2 x_k$$

сокращенный полином второй степени $y_i = a_0 + a_1 x_j + a_2 x_k + a_3 x_j x_k$.

квадратичный элементарный полином $y_i = a_0 + a_1 x_j + a_2 x_k + a_3 x_j x_k + a_4 x_j^2 + a_5 x_k^2$ элементарный полином смешанного вида $z_i = a_0 + a_1 x_j + a_2 y_k + a_3 x_j y_k$.

Вся выборка делится на обучающую и проверочную: $N_{выб} = N_{обуч} + N_{пров}$. Вх-й вектор имеет

размерность N : $(X = (x_1, \dots, x_N))$. 1-ый ряд - на основе обучающей последовательности

строятся частные описания от всех попарных комбинаций исходных аргументов, приближающие по

методу наим-х квадратов выходную переменную: $y_1 = f_1(x_1, x_2), y_2 = f_2(x_1, x_3), \dots, y_k = f_k(x_{n-1}, x_n)$. Из этих

моделей выбирается нек-е число лучших по критерию селекции (используя проверочную послед-ть). 2-ой

ряд - полученные переменные принимаются в качестве аргументов второго ряда, и снова строятся все

частные описания от двух аргументов: $z_1 = \varphi_1(y_1, y_2), z_2 = \varphi_2(y_1, y_3), \dots, z_1 = \varphi_1(y_{F-1}, y_F)$. Из них по

критерию селекции отбирается $F2$ лучших моделей в качестве переменных следующего ряда и т.д.

Критерий остановки МГУА: ряды наращиваются до тех пор, пока снижается значение критерия

селекции. Заключительный этап состоит в восстановлении искомой функции по полученным

промежуточным функциям. **Критерии селекции:**

$$1) \text{ критерий регулярности (точности) } \bar{\epsilon}^{-2} = \frac{1}{N_{np}} \cdot \sum_{i=1}^{N_{np}} (y_i - y_i^*)^2 \quad \bar{\Delta}_{np}^{-2} = \frac{\sum_{i=1}^{N_{np}} (y_i - y_i^*(x))^2}{\sum_{i=1}^{N_{np}} (y_i - \bar{y})^2}$$

2) критерий несмещенности. Всю выборку делим на две части $R=R_1+R_2$. Первый эксперимент: R_1 -обучающая выборка, R_2 -проверочная; определяем выходы модели y_i^* , $i=1..R$. Второй эксперимент: R_2 -обучающая выборка, R_1 -проверочная; определяем выходы модели y_i^{**} , $i=1..R$ и сравниваем. Критерий несмещенности: $n_{см} = \frac{1}{N} \cdot \sum_{i=1}^R (y_i^* - y_i^{**})^2$ Процесс селекции осуществляется до тех пор, пока этот критерий не перестанет уменьшаться, т.е. до достижения условия $n_{см} \rightarrow \min$.

59. Применение МГУА для решения задач распознавания.

Задачу РО будем решать с помощью выявления нек-го инвариантного св-ва для совокупности изображений, объединяемых в один класс. Т.е. данное свойство для образов из одного класса примерно одинаково, а для образов из различных классов имеет отличающ-ся значения. Пусть инвариантным св-вом образов класса будет нек-я фун-я $f(X)$. Инвариантность - значит, что функ-я остается почти одинаковой для всей совокупности образов одного класса, т.е. $f(X) \equiv C_1$ для класса V_1 ;

$f(X) \equiv C_2$ для класса V_2 и т.д. Функ-ю $f(X)$ можно представить как поверх-ть в многомерном простр-ве, построенную так, что образам разных классов соответствуют несвязные отрезки на оси f . Всегда можно построить такую поверхность. В качестве такой функции можно использовать полином

$$\text{Колмогорова-Габора : } f(x) = a_0 + \sum_{i=1}^n a_i x_i + \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j + \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n a_{ijk} x_i x_j x_k + \dots$$

Здесь x_i, x_j, x_k - признаки. Решающее правило, основанное на построенной функции, в случае

двух классов записывается так: $\begin{cases} X \in V_1 \text{ если } |f(X) - \overline{f_1(X)}| < |f(X) - \overline{f_2(X)}| \\ X \in V_2 \text{ если } |f(X) - \overline{f_1(X)}| \geq |f(X) - \overline{f_2(X)}| \end{cases}$ где $\overline{f_1(X)}$ и $\overline{f_2(X)}$ -

средняя высота поверхности над изображениями первого и второго классов. Если же классов больше двух, то решающее правило принимает вид $X \in V_k$, где $k = \arg \min_j |f(X) - \overline{f_j(X)}|$.

60. Достоинства и недостатки стохастического подхода в распознавании образов.

+ : универсальность, простота реализации процедур распознавания, проверяемость предположений о обучающей послед-ти с помощью методов матстатистики. - : для достаточно хорошей оцекни ЗР ССВ X может понадобиться большая обучающая послед-ть (особенно при рассмотрении большого кол-ва вводимых признаков).

61. Постановка задачи распознавания лингвистических образов.

Пусть изображение объекта представляет собой цепочки в нек-м алфавите V_T и всего имеется m возможных образов. Тогда каждый из этих образов представляет собой нек-й язык $L(G_i)$ над алфавитом V_T , порождаемый грам-й $G_i = \{V_N^i, V_T, P_i, S_i\}$, $i=1..m$. Задача классиф-и заключается в определении принадлежности любой цепочки x одному из языков L_1, \dots, L_m . Процедура классиф-и:

$$x \Rightarrow w_i \Leftrightarrow x \in L_i = L(G_i)$$

62. Три способа решения задачи распознавания лингвистических образов.

1) сопоставление вх-й цепочки x с эталонными цепочками z_1, \dots, z_m , $z_i \in L(G_i)$, $i=1..m$. Цепочка x относится к тому классу w_i , с эталоном z_i кот-го x согласуется наилучшим образом по критерию меры близости м\у цепочками $D(x, z_i)$ (применить правило ближ-го соседа по критерию расстояния). +: быстрота и эффективность. -: потеря части информации о языке, проблема с выбором эталона и меры близости. 2) классиф-я вх цепочки x с помощью определённого для каждого языка набора синтаксических правил. Цепочка $x \rightarrow w_i$, если x не противоречит его набору синтакс-х правил S_i . +: простота алгоритма. -: если язык достаточно сложный, то кол-во синтаксических правил м\б чрезмерно большим (проблема в построении этих правил). 3) Классиф-я вх цепочки x с помощью синтаксического анализа (грам-го разбора). Проверяемая цепочка $x \rightarrow w_i$, для соответствующей грам-ки которого G_i грам-й разбор x оказался успешным.

63. Задача восстановления грамматики по обучающей выборке.

Пусть имеется нек-й алфавит символов V_T ; нек-я последовательность цепочек в этом алфавите $St = \{x_1, \dots, x_t\}$ – выборка. В St содержатся цепочки двух видов: 1) цепочки из языка $L(G^*)$ нек-й грам-ки G^* , т.е. $St \subseteq L(G^*)$ или $St^+ = L(G^*) \cap St \subseteq L(G^*)$; 2) цепочки из дополнения к этому языку, т.е. $St \subseteq V_T^* \setminus L(G^*)$ или $St = V_T^* \setminus L(G^*) \cap St \subseteq V_T^* \setminus L(G^*)$. Определим $St+1 = St \cup \{Xt+1\}$. Пусть A – алгоритм, формирующий для каждого St грам-ку Gt , в кот-й выводимы все цепочки из St^+ (грам-ка допускает St^+) и не выводимы все цепочки из St (грам-ка отвергает St), т.е. $St^+ \subseteq L(Gt)$, $St \cap L(Gt) = \emptyset$. Пусть Γ – мн-во грам-к, кот-е может сформировать алгоритм A и пусть $Gt \in \Gamma$. ЗВГ G^* назовём разрешимой, если существует алгоритм A , формируй за конечное число шагов такую грам-ку $G \in \Gamma$, что $L(G) = L(G^*)$, т.е. существует такое t^* , что для всех $t > t^*$ грам-ка $Gt \sim G^*$.

64. Информаторное представление задачи восстановления грамматики.

ЗВГ G^* дана в информативном представлении (или язык $L(G^*)$ дан в информаторном представлении), если $St = St^+ \cup St^-$, $St^- = \emptyset$, $St^+ \neq \emptyset$ и для любого $x \in L(G^*)$ существует такое t' , что $x \in St$ при $t \geq t'$, а для любого $y \in V_T^* \setminus L(G^*)$ существует t'' : при $t \geq t''$ $y \notin St$.

65. Текстуальное представление задачи восстановления грамматики.

Задача ВГ G^* дана в текстуальном представлении (или язык $L(G^*)$ дан в текстуальном представлении), если $St = St^+$ и для любого $x \in L(G^*)$ существует такое t' , $x \in St$.

66. Теорема о разрешимости для информаторного представления.

Th.: пусть Γ – мн-во разрешимых грамматик. Тогда при информаторном представлении ЗВГ разрешима для $\forall G^* \in \Gamma$. Док-во: пусть A перечисляет Γ в порядке G^1, \dots, G^k, \dots . Пусть G^k – первая из грамматик, эквив-ная G^* . На любом шаге t алгоритм A выбирает в качестве Gt первую грам-ку в послед-ти G^1, \dots, G^k, \dots , кот-я допускает St^+ и отвергает St^- . Пусть G^i ($i < k$) – первая такая грам-ка, т.е. A полагает $Gt = G^i$. Т.к. $i < k$, то G^i не эквив-на G^* , поэтому с увеличением t (ростом числа цепочек в St) в St найдётся цепочка x , ошибочно допускаемая или ошибочно отвергаемая грам-кой G^i . Тогда G^i будет отброшена и выбрана следующая допуст-я выборкой грам-ка G^j , $j > i$. На нек-м шаге t^* будет выбрана грам-ка G^k . Т.к. $G^k \sim G^*$, то в дальнейшем (при $t > t^*$) она никогда не будет отброшена. Это означает, что ЗВГ G^* решена.

67. Теорема о неразрешимости для текстуального представления.

Th.: пусть Γ – мн-во грамматик, содержащее все грамматики, порождающие все конечные языки и грам-ку, пород-т ∞ язык. Тогда при текстуальном представлении ЗВГ не разрешима для произвольной грамматики из мн-ва Γ .

Док-во: пусть G^* и G' порождают конечные языки и $L(G^*) \subsetneq L(G')$. Пусть алгоритм A перечисляет Γ в порядке: $\dots, G', \dots, G^*, \dots$. Тогда для любого $St = St^+$ имеем $St \subseteq L(G^*) \subsetneq L(G')$ для любого конечного t , т.е. G' всегда в качестве допустимой грам-ки. Т.к. $L(G^*)$ конечен, то для нек-го t нельзя построить $St+1$, т.е. G' принимается алгоритмом A в качестве грам-ки, эквив-й G^* , хотя условие $L(G') = L(G^*)$ не выполнено. В этом случае A разрешает ЗВГ, порождающей ∞ язык. Пусть теперь A перечисляет Γ так: по данной конечной выборке St строит грам-ку G' , пород-ю конечный язык $L(G') = St$; здесь A разрешает задачу восстановления любой грам-ки $G^* \in \Gamma$, пород-й конечный язык. Возьмём в качестве G^* , пород-ю ∞ язык. Тогда число цепочек в St с ростом t неограниченно растёт, но для любого конечного t существует грам-ка G'' с конечным языком $L(G'')$, кот-я отыскивается алгоритмом A . Т.к. таких грамматик бесконечно много, алгоритм A не разрешает ЗВГ G^* .

68. Два класса алгоритмов восстановления грамматики.

Существует 2 класса таких алгоритмов: 1) алгоритмы восстановления грамматики индукцией (напр, алгоритм Фельдмана); 2) восстановление перечислением.

69. Алгоритм Фельдмана.

Параметры работы алгоритма: k -длина правой части остаточного правила;

$S_t^+ = \{x_1, \dots, x_t\}$. Замечание: список цепочек в S_t^+ отсортирован по убыванию длины. 1) *Формирование основы грамматики.*

Введем в рассмотрение множество $I^0 = \{x \mid x \in S_t^+, |x| = n\}$, где n – длина самой длинной цепочки в S_t^+ .

На данном этапе формируем мн-ва продукций P^0 и нетерминалов V_N^0 по следующему правилу:

$P^0 = \emptyset$, $V_N^0 = \{S\}$; Для каждой цепочки $a_1 \dots a_n \in I^0$:

$$P^0 = P^0 \cup \{S \rightarrow a_1 A_1, A_1 \rightarrow a_2 A_2, \dots, A_{n-k} \rightarrow a_{n-k} A_{n-k},$$

$$A_{n-k} \rightarrow a_{n-k+1} \dots a_n\}$$

$$V_N^0 = V_N^0 \cup \{A_1, A_2, \dots, A_{n-k}\}.$$

Продукции вида $A_{n-k} \rightarrow a_{n-k+1} \dots a_n$ наз-ся *остаточными правилами*. Длина остаточного правила равна k . 2) *Дополнение грамматики (до нерекурсивной)*. На данном этапе формируем мн-ва продукций P^1 и нетерминалов V_N^1 по правилу: $P^1 = \emptyset$, $V_N^1 = \emptyset$,

Для каждой цепочки $a_1 \dots a_l \in S_t^+ \setminus I^0$:

$$\text{если } \exists p < l \text{ и } \exists A' \in V_N^1 \cup V_N^0 \text{ такие, что } S \xRightarrow[p^0 \cup P^1]{+} a_1 \dots a_p A', \text{ то}$$

$$P^1 = P^1 \cup \{A' \rightarrow a_{p+1} A_{p+1}, A_{p+1} \rightarrow a_{p+2} A_{p+2}, \dots,$$

$$A_{l-2} \rightarrow a_{l-1} A_{l-1}, A_{l-1} \rightarrow a_l\}$$

$$V_N^1 = V_N^1 \cup \{A_{p+1}, A_{p+2}, \dots, A_{l-1}\}$$

иначе

$$P^1 = P^1 \cup \{S \rightarrow a_1 A_1, A_1 \rightarrow a_2 A_2, \dots,$$

$$A_{l-2} \rightarrow a_{l-1} A_{l-1}, A_{l-1} \rightarrow a_l\}$$

$$V_N^1 = V_N^1 \cup \{A_1, A_2, \dots, A_{l-1}\}.$$

После первых двух этапов работы алгоритма получаем грамматику

$$G^0 = \langle V_N = V_N^0 \cup V_N^1, V_T, P = P^0 \cup P^1, S \rangle \text{ кот-я порождает только те цепочки, которые входят в } S_t^+. 3)$$

Получение рекурсивной автоматной грамматики. Для каждого остаточного правила $A \rightarrow a_1 \dots a_k \in P$:

если $\exists A' \in V_N \setminus \{A \in V_N : A \rightarrow a_1 \dots a_k \in P\}$ такой, что $A' \xRightarrow{P}{+} a_1 \dots a_k$, то:

$$\text{a) } P = P \setminus \{A \rightarrow a_1 \dots a_k\}, \text{ b) } V_N = V_N \setminus \{A\},$$

с) во всех продукциях из P заменяем A на A' . 4) *Объединение эквивалентных нетерминалов.* Для

каждой пары нетерминалов $A_i, A_j \in V_N$: если $\{x \mid A_i \xRightarrow{P}{+} x\} = \{x \mid A_j \xRightarrow{P}{+} x\}$, то:

$$\text{a) } V_N = V_N \setminus \{A_j\}, \text{ b) во всех продукциях из } P \text{ заменяем } A_j \text{ на } A_i.$$

70. Требования к вектору признаков.

1°) ВП не должен содержать компонент, кот-е для данного набора объектов не определены; 2°) не должен содержать признаков, кот-е для всех рассматриваемых объектов принимают одно и тоже значение; 3°) не должен содержать компонент, кот-е не могут быть измерены (получены); 4°) \neg , кот-е статистически или функционально зависимына всём мн-ве объектов; 5°) желательно, чтобы компоненты в ВП для объектов из одного класса были близки, а из разных - сильно отличались; 6°) кол-во компонент (n) д\б как можно меньше. Замечание: выполнение 6° способствует выполнению 1°-4°.

71. Среднеквадратичные расстояния м\у различными типами объектов в евклидовом пространстве.

$D^2(a, b) = \sum_{k=1}^n (a_k - b_k)^2$ - расстояние м\у различными типами объектов в евклид-м простр-ве. Пусть есть

тчк x в ЕП и мн-во точек $\{a^i\}_{i=1}^k$. Расстояние м\у ними : $\sum_{k=1}^n (x_k - a_k^i)^2$

Среднеквадратичное расст-е м\у x и $\{a^i\}_{i=1}^k$:

$D^2(x, \{a^i\}_{i=1}^k) = \frac{1}{k} \sum_{i=1}^k \sum_{k=1}^n (x_k - a_k^i)^2$ Среднекв-е внутримножеств-е расстояние :

$$D^2(a^j, \{a^i\}_{i=1}^k) = \frac{1}{k-1} \sum_{i=1, i \neq j}^k \sum_{k=1}^n (a_k^j - a_k^i)^2, i \neq j \quad D^2(\{a^j\}_{j=1}^k, \{a^i\}_{i=1}^k) = \frac{1}{k} \sum_{j=1}^k D^2(a^j, \{a^i\}_{i=1, i \neq j}^k) =$$

Рассмотри

$$\frac{1}{k(k-1)} \sum_{j=1}^k \sum_{i=1}^k \sum_{k=1}^n (a_k^j - a_k^i)^2$$

м $\{a^j\}_{j=1}^{ka}$ и $\{b^i\}_{i=1}^{kb}$. Среднеквадрат-е расст-е м\у ними:

$$D^2(\{a^j\}_{j=1}^{ka}, \{b^i\}_{i=1}^{kb}) = \frac{1}{ka \cdot kb} \sum_{j=1}^{ka} \sum_{i=1}^{kb} \sum_{k=1}^n (a_k^j - b_k^i)^2$$

$$D^2(\{a^j\}_{j=1}^{ka}, \{b^i\}_{i=1}^{kb}) = \frac{1}{ka \cdot kb} \sum_{j=1}^{ka} \sum_{i=1}^{kb} \sum_{k=1}^n (a_k^j - b_k^i)^2$$

72. Преобразование детерминированного вектора признаков, минимизирующее внутриклассовое расстояние.

$$D_p^2 = D^2(\{z_p^j\}_{j=1}^{Np}, \{z_p^i\}_{i=1}^{Np}) \Rightarrow D_p^* =$$

$$2 \sum_{k=1}^n (\sigma_{pk}^*)^2 = 2 \sum_{k=1}^n w_{kk}^2 (\sigma_{pk})^2 \rightarrow \min \quad \text{при ограничениях: } \sum_{k=1}^n w_{kk} = 1 \text{ и } w_{kk} \geq 0, k=1, n.$$

73. Способы учета изменения межклассовых расстояний при минимизации внутриклассовых расстояний.

1) внутриклассовое расстояние должно быть $\leq A$, а межклассовое $\geq B$; Но для такой задачи решения может не оказаться. 2) перейти от диагональной матрицы преобраз-я к матрице общего вида. 3) перейти от нелинейного к линейному виду.

74. Сравнение стохастических признаков на основе "апостериорных" вероятностей.

Имеется m классов: $\Omega = \{w_1, \dots, w_m\}$. Известны $\{p(w_i)\}_{i=1}^m$. Изображение: $x = (x_1, \dots, x_n) \in X = (X_1, \dots, X_n)$ (ССВ), для каждой из СВ известны условные законы распределения этой СВ, при условии реализации каждого из классов w_1, \dots, w_m ; т.е. $P(X_k = x_k / w_i)$ -ДСВ, $p(X_k / w_i)$ -АНСВ (для $i=1..m$). Выделим один из признаков X_k . $\Delta 1$ -U-е всех интервалов, на кот-х отличен от 0 ровно один из условных ЗР для СВ X_k . $\Delta 2$ - \setminus - два УЗР СВ X_k Δm - \setminus -, где все m УЗР "+"-ны для СВ X_k . Предположим, что решение о принадлежности к какому-либо классу выносится только на основании реализации x_k СВ X_k . Тогда для

любого разумного правила для вынесения решения справедливо: если реализуется событие $\{X_k \in \Delta_l\}$, то выносится решение об отнесении объекта к одному из l классов. $P(X_k \in \Delta_l) = P_{kl} = \sum_{i=1}^m P(X_k \in \Delta_l / w_i) P(w_i)$. $P(X_k \in \Delta_l / w_i) = 1$ \sum по $X_k \in \Delta_l$ $P(X_k = x_k / w_i)$ - ДСВ; 2) \int по Δ_l $P_k(X_k / w_i)$ - АНСВ. Рассмотрим СВ:

Y_k (ДСВ): 1 2 ... l ... m

P : $P_{k1} P_{k2} \dots P_{kl} \dots P_{km}$ (это ЗР). Вычислим матожидание: $M_k = M(Y_k) = \sum_{i=1}^m l * P_{kl}$. Возьмём другой признак: X_h . Для него сделаем всё то же, что и для X_k . Y_h (ДСВ): 1 2 ... l ... m

P : $P_{h1} P_{h2} \dots P_{hl} \dots P_{hm}$ (это ЗР), $M_h = M(Y_h) = \sum_{i=1}^m l * P_{hl}$. СВ Y_k можно трактовать как нек-й неопределённости классификации при классиф-и по признаку X_k . СВ Y_h - \- при классиф-и только по признаку X_h . Введённые матожидания M_k и M_h – меры неопр-ти соответственно при классиф-и по признаку X_k и X_h . Процедура сравнения: если $M_k < M_h \Rightarrow X_k$ – полезнее, чем X_h , т.е. $X_k > X_h$.

75. Сравнение стохастических признаков на основе их условных матожиданий и условных дисперсий.

Имеется m классов: $\Omega = \{w_1, \dots, w_p\}$, $\{p(w_i)\}_{i=1}^p$. Изображение явл-ся реализацией системы СВ: $x = (x_1, \dots, x_n) \leq X = (X_1, \dots, X_n)$, для кот-х известны их условные матожидание и дисперсия. $M[X_j / w_{ij}] = m_{ij}$;

$D[X_j / w_{ij}] = \sigma_{ij}^2$ ($i=1..p$, $j=1..n$). Выделим признак X_j . Введём две СВ:

μ_j : $m_{1j}, m_{2j}, \dots, m_{pj}$ | δ_j : $\sigma_{1j}^2, \sigma_{2j}^2, \dots, \sigma_{pj}^2$

p : $p(w_1), p(w_2), \dots, p(w_p)$ | p : $p(w_1), p(w_2), \dots, p(w_p)$

$M[\mu_j] = \sum_{i=1}^p m_{ij} P(w_i) = M_j$, $M[\delta_j] = \sum_{i=1}^p \sigma_{ij}^2 P(w_i) = D_j$. σ_{ij}^2 – нек-я мера рассеяния СВ X_j вокруг величины m_{ij} , поэтому, если σ_{ij}^2 – мала, то это интерпрет-ся как малая изменчивость признака X_j внутри класса w_i ;

если это имеет место для всех классов, то это означает малую изменчивость признака X_j для всех классов. Если σ_{ij}^2 мала для всех i , то D_j – тоже мала. Если D_j – большая, то для каких-то i признак X_j сильно изменяется для всех классов. D_j – показатель внутриклассовой изменчивости. Если $X_j > X_k$, $D_j < D_k \Rightarrow X_j$ – полезнее. Ориен-сь на D_j можно удовлетв-ть 5° частично. $D[\mu_j] = M[(\mu_j - M_j)^2] = \sum_{i=1}^p (m_{ij} - M_j)^2 P(w_i) = DJ$ – показывает меру рассеяния возможных значений m_{ij} вокруг M_j . Если DJ – мала, то СВ X_j будет принимать значения, близкие к M_j для различных классов, т.е. признак X_j будет обладать малой межклассовой изменчивостью. Если $X_j > X_k$, $DJ > DK \Rightarrow X_j$ – предпочтит-е (удовл-ет 2-ю часть 5°). Рассмотрим показатель: $K_j = D_j / DJ$. Если $X_j > X_k$, $K_j > K_k \Rightarrow X_j$ – полезнее.