

## Лекция 4

Статистические методы распознавания,  
Распознавание при заданной точности для некоторых классов,  
ROC-анализ

*Лектор – Сенько Олег Валентинович*

- 1 Методы, основанные на формуле Байеса
- 2 Логистическая регрессия
- 3 К ближайших соседей

Ранее было показано, что максимальную точность распознавания обеспечивает байесовское решающее правило, относящее распознаваемый объект, описываемый вектором  $\mathbf{x}$  переменных (признаков)  $X_1, \dots, X_n$  к классу  $K_*$ , для которого условная вероятность  $P(K_* | \mathbf{x})$  максимальна. Байесовские методы обучения основаны на аппроксимации условных вероятностей классов в точках признакового пространства с использованием формулы Байеса. Рассмотрим задачу распознавания классов  $K_1, \dots, K_L$ . Формула Байеса позволяет рассчитать Условные вероятности классов в точке признакового пространства могут быть рассчитаны с использованием формулы Байеса. В случае, если переменные  $X_1, \dots, X_n$  являются дискретными формула Байеса может быть записана в виде:

$$P(K_i | \mathbf{x}) = \frac{P(\mathbf{x} | K_i)P(K_i)}{\sum_{i=1}^L P(K_i)P(\mathbf{x} | K_i)} \quad (1)$$

где  $P(K_1), \dots, P(K_L)$  - вероятность классов  $K_1, \dots, K_L$  безотносительно к признаковым описаниям (априорная вероятность).

В качестве оценок априорных вероятностей

$$P(K_1), \dots, P(K_L)$$

могут быть взяты доли объектов соответствующих классов в обучающей выборке. Условные вероятности  $P(\mathbf{x} | K_1), \dots, P(\mathbf{x} | K_L)$  могут оцениваться на основании сделанных предположений.

Например, может быть использовано предположение о независимости переменных для каждого из классов. Предположим, что переменные  $X_1, \dots, X_n$  являются **категориальными**, а каждая переменная  $X_j$  принимает значения из конечного множества

$$\widetilde{M}_j^i = \{a_{ji}^1, \dots, a_{ji}^{r(i,j)}\}$$

Тогда вероятность  $P(\mathbf{x}_j | K_i)$  для вектора  $\mathbf{x}_k = (x_{j1}, \dots, x_{jn})$  может быть представлена в виде:

$$P(\mathbf{x}_j | K_i) = \prod_{i=1}^n P(X_i = x_{ji} | K_i). \quad (2)$$

Для того, чтобы воспользоваться формулой (2) достаточно знать вероятность выполнения равенства  $X_j = a_{ji}^k$  для произвольного класса и произвольной переменной. Для оценки вероятности  $P(X_j = a_{ji}^k | K_i)$  может использоваться  $\nu_{ij}^k$  - доля объектов из  $\tilde{S}_t \cap K_i$ , для которых  $X_j = a_{ji}^k$ . Метод использующий формулу Байеса вместе с гипотезой о независимости признаков называется наивным байесовским классификатором

**Непрерывность переменных** В случае, если переменные  $X_1, \dots, X_n$  являются непрерывными, формула Байеса может быть записана с использованием

$$P(K_i | \mathbf{x}) = \frac{p_i(\mathbf{x})P(K_i)}{\sum_{i=1}^L P(K_i)p_i(\mathbf{x})}, \quad (3)$$

где  $p_1(\mathbf{x}), \dots, p_L(\mathbf{x})$  - значения плотностей вероятностей классов  $K_1, \dots, K_L$  в пространстве  $\mathbb{R}^n$ .

лотности вероятностей

$$p_1(\mathbf{x}), \dots, p_L(\mathbf{x})$$

также могут оцениваться исходя из предположения о нормальности распределения  $X_1, \dots, X_n$ .

Принимая во внимание, что знаменатель формулы

$$P(K_i | \mathbf{x}) = \frac{p_i(\mathbf{x})P(K_i)}{\sum_{i=1}^L P(K_i)p_i(\mathbf{x})},$$

всегда больше 0 олучаем, что условием отнесения  $\mathbf{x}$  классу  $K_i$  является выполнение условия

$$p_i(\mathbf{x})P(K_i) > p_j(\mathbf{x})P(K_j)$$

$$j \neq i, j = 1, \dots, L$$

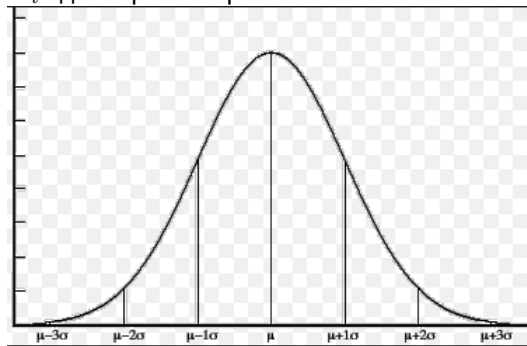
## Одномерный случай. Нормальное распределение

Плотность одномерного распределения для класса  $K_i$  задаётся формулой:

$$p_i(X) = \frac{1}{\sqrt{2\pi D_i}} e^{-\frac{(X-M_i)^2}{2D_i}},$$

$M_i$ - математическое ожидание переменной  $X$

$D_i$ - дисперсия переменной  $X$





# Одномерный случай. Задача распознавания с двумя классами $K_1$ и $K_2$

Область значения  $X$ , в которой происходит отнесение в класс  $K_1$  задаётся неравенством

$$\frac{P(K_1)}{\sqrt{2\pi D_1}} e^{-\frac{(X-M_1)^2}{2D_1}} > \frac{P(K_2)}{\sqrt{2\pi D_2}} e^{-\frac{(X-M_2)^2}{2D_2}} \quad (4)$$

которое эквивалентно неравенству

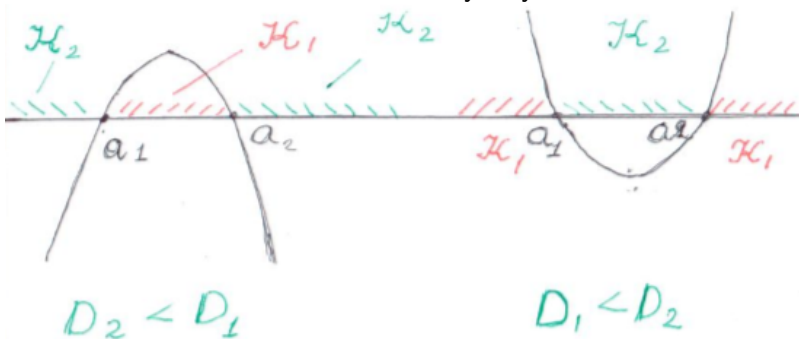
$$\ln \frac{P(K_1)D_2}{P(K_2)D_1} - \frac{(X-M_1)^2}{2D_1} + \frac{(X-M_2)^2}{2D_2} > 0 \quad (5)$$

# Одномерный случай. Задача распознавания с двумя классами $K_1$ и $K_2$

Пусть  $a_1$  и  $a_2$  - корни уравнения

$$\ln \frac{P(K_1)D_2}{P(K_2)D_1} - \frac{(X - M_1)^2}{2D_1} + \frac{(X - M_2)^2}{2D_2} = 0 \quad (6)$$

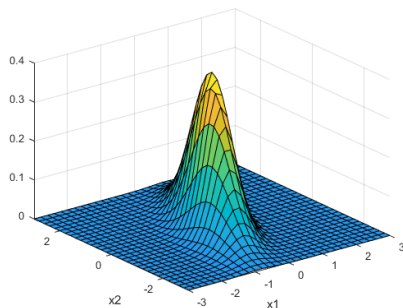
Тогда области отнесения классам соответствуют указанным на



рисунке.

# Аппроксимация плотности с помощью многомерного нормального распределения

При решении задач распознавания с помощью формулы Байеса в форме (3) могут использоваться плотности вероятности  $p_1(\mathbf{x}), \dots, p_L(\mathbf{x})$ , в которых переменные  $X_1, \dots, X_n$  не обязательно являются независимыми. Чаще всего используется многомерное



нормальное распределения.

# Аппроксимация плотности с помощью многомерного нормального распределения

Плотность данного распределения в общем виде представляется выражением

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})^t\right], \quad (7)$$

где

$\boldsymbol{\mu}$  - математическое ожидание вектора признаков  $\mathbf{x}$ ;  $\Sigma$  - матрица ковариаций признаков  $X_1, \dots, X_n$ ;  $|\Sigma|$  - детерминант матрицы  $\Sigma$ .

Для построения распознающего алгоритма достаточно оценить вектора математических ожиданий  $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_L$  и матрицы ковариаций  $\Sigma_1, \dots, \Sigma_L$  для классов  $K_1, \dots, K_L$ , соответственно.

# Аппроксимация плотности с помощью многомерного нормального распределения

Оценка вектора математических ожиданий  $\mu_i$  вычисляется как среднее значение векторов признаков по объектам обучающей выборки  $\tilde{S}_t$  из класса  $K_i$  :

$$\hat{\mu}_i = \frac{1}{m_i} \sum_{s_j \in \tilde{S}_t \cap K_i} \mathbf{x}_j$$

, где  $m_i$  - число объектов класса  $K_i$  в обучающей выборке. Элемент матрицы ковариаций для класса  $K_i$  вычисляется по формуле

$$\hat{\sigma}_{kk'}^i = \frac{1}{m_i} \sum_{s_j \in \tilde{S}_t \cap K_i} (x_{jk} - \mu_{ik})(x_{jk'} - \mu_{ik'}),$$

где  $x_{jk} - \mu_{ik}$  -  $k$ -я компонента вектора  $\mu_i$ . Матрицу ковариации, состоящую из элементов  $\hat{\sigma}_{kk'}^i$  обозначим  $\hat{\Sigma}_i$ . Очевидно, что согласно формуле Байеса максимум  $P(K_i | \mathbf{x})$  достигается для тех же самых классов для которых максимально произведение  $P(K_i)p_i(\mathbf{x})$ .

## Использование формулы Байеса. Многомерное нормальное распределение

Очевидно, что для байесовской классификации может использоваться также натуральный логарифм  $\ln[P(K_i)p_i(\mathbf{x})]$  который согласно вышеизложенному может быть оценён выражением

$$g_i(\mathbf{x}) = -\frac{1}{2}\mathbf{x}\hat{\Sigma}_i^{-1}\mathbf{x}^t + \mathbf{w}_i\mathbf{x}^t + g_i^0,$$

где  $\mathbf{w}_i = \hat{\boldsymbol{\mu}}_i\hat{\Sigma}_i^{-1} g_i^0$  - не зависящее от  $\mathbf{x}$  слагаемое:  
 $\nu_i$  - доля объектов класса  $K_i$  в обучающей выборке. Слагаемое  $g_i^0$  имеет вид

$$g_i^0 = -\frac{1}{2}\hat{\boldsymbol{\mu}}_i\hat{\Sigma}_i^{-1}\hat{\boldsymbol{\mu}}_i^t - \frac{1}{2}\ln(|\hat{\Sigma}_i|) + \ln(\nu_i) - \frac{n}{2}\ln(2\pi).$$

## Использование формулы Байеса. Многомерное нормальное распределение

Таким образом объект с признаковым описанием  $x$  будет отнесён построенной выше аппроксимацией байесовского классификатора к классу, для которого оценка  $g_i(x)$  является максимальной. Следует отметить, что построенный классификатор в общем случае является квадратичным по признакам. Однако классификатор превращается в линейный, если оценки ковариационных матриц разных классов оказываются равными.

## Многомерное нормальное распределение при справедливости гипотезы о независимости признаков

этом случае  $p_i(\mathbf{x})$  может быть представлена в виде произведения одномерных плотностей

$$p_i(\mathbf{x}) = \prod_{j=1}^n p_{ji}(X_j),$$

где  $p_{ji}(X_j)$  - плотность распределения переменной  $X_j$  для класса  $K_i$ , задаваемая формулой

$$p_{ji}(X_j) = \frac{1}{\sqrt{2\pi D_{ji}}} e^{-\frac{(X_j - M_{ji})^2}{2D_{ji}}},$$

где  $M_{ji}, D_{ji}$  являются математическим ожиданием и дисперсией переменной  $X_j$ . Данные параметры легко оцениваются по  $\tilde{S}_t$ .



этом случае  $p_i(\mathbf{x})$  может быть представлена в виде произведения одномерных плотностей

$$p_i(\mathbf{x}) = \prod_{j=1}^n p_{ji}(X_j),$$

где  $p_{ji}(X_j)$  - плотность распределения переменной  $X_j$  для класса  $K_i$ , задаваемая формулой

$$p_{ji}(X_j) = \frac{1}{\sqrt{2\pi D_{ji}}} e^{-\frac{(X_j - M_{ji})^2}{2D_{ji}}},$$

где  $M_{ji}, D_{ji}$  являются математическим ожиданием и дисперсией переменной  $X_j$ . Данные параметры легко оцениваются по  $\tilde{S}_t$ .

Для случая смешанных признаков для оценки  $p_i(\mathbf{x})$  может быть использована формула

$$p_i(\mathbf{x}_j) = \prod_{i=1}^{n_{cat}} P(X_i = x_{ji} | K_i) \prod_{j=1}^{n_{cont}} N_{ji}(X_j) \quad (8)$$

$n_{cont}$ - число непрерывных переменных;  $n_{cat}$ - число категориальных переменных.

Рассмотрим вариант метода Линейный дискриминант Фишера (ЛДФ) для распознавания двух классов  $K_1$  и  $K_2$ . В основе метода лежит поиск в многомерном признаковом пространстве такого направления  $\mathbf{w}$ , чтобы средние значения проекции на него объектов обучающей выборки из классов  $K_1$  и  $K_2$  максимально различались. Проекцией произвольного вектора  $\mathbf{x}$  на направление  $\mathbf{w}$  является отношение

$$\frac{(\mathbf{w}\mathbf{x}^t)}{|\mathbf{w}|}.$$

В качестве меры различий проекций классов на используется функционал

$$\Phi(\mathbf{w}, \tilde{S}_t) = \frac{(\hat{X}_{\mathbf{w}1} - \hat{X}_{\mathbf{w}2})^2}{\hat{d}_{\mathbf{w}1} + \hat{d}_{\mathbf{w}2}},$$

где

$$\hat{X}_{wi} = \frac{1}{m_i} \sum_{s_j \in \tilde{S}_t \cap K_i} \frac{(w x_j^t)}{|w|}$$

- среднее значение проекции векторов переменных  $X_1, \dots, X_n$ , описывающих объекты из класса  $K_i$ ;

$$\hat{d}_{wi} = \frac{1}{m_i} \sum_{s_j \in \tilde{S}_t \cap K_i} \left[ \frac{(w x_j^t)}{|w|} - \hat{X}_{wi} \right]^2$$

- дисперсия проекций векторов, описывающих объекты из класса  $K_i, i \in \{1, 2\}$ . Смысл функционала  $\Phi(w, \tilde{S}_t)$  ясен из его структуры. Он является по сути квадратом отличия между средними значениями проекций классов на направление  $w$ , нормированным на сумму внутриклассовых выборочных дисперсий.

Можно показать, что  $\Phi(\mathbf{w}, \tilde{S}_t)$  достигает максимума при

$$\mathbf{w} = \hat{\Sigma}_{12}^{-1}(\hat{\boldsymbol{\mu}}_1^t - \hat{\boldsymbol{\mu}}_2^t), \quad (9)$$

где  $\hat{\Sigma}_{12} = \hat{\Sigma}_1 + \hat{\Sigma}_2$ . Таким образом оценка направления, оптимального для распознавания  $K_1$  и  $K_2$  может быть записана в виде ( 5 )

Распознавание нового объекта  $s_*$  по векторному описанию  $\mathbf{x}_*$  производится по величине его проекции на направление  $\mathbf{w}$ :

$$\gamma(\mathbf{x}_*) = \frac{(\mathbf{w}, \mathbf{x}_*^t)}{|\mathbf{w}|}. \quad (10)$$

При этом используется простое пороговое правило: при  $\gamma(\mathbf{x}_*) > b$  объект  $s_*$  относится к классу  $K_1$  и  $s_*$  относится к классу  $K_2$  в противном случае.

Граничный параметр  $b$  подбирается по обучающей выборке таким образом, чтобы проекции объектов разных классов на оптимальное направление  $w$  оказались бы максимально разделёнными. Простой, но эффективной, стратегией является выбор в качестве порогового параметра  $b$  средней проекции объектов обучающей выборки на  $w$ . Метод ЛДФ легко обобщается на случай с несколькими классами. При этом исходная задача распознавания классов  $K_1, \dots, K_L$  сводится к последовательности задач с двумя классами  $K'_1$  и  $K'_2$ :

Зад. 1. Класс  $K'_1 = K_1$ , класс  $K'_2 = \Omega \setminus K_1$

.....

Зад.  $L$ . Класс  $K'_1 = K_L$ , класс  $K'_2 = \Omega \setminus K_L$

Для каждой из  $L$  задач ищется оптимальное направление. В результате получается набор из  $L$  направлений  $w_1, \dots, w_L$ .

В результате получается набор из  $L$  направлений  $w_1, \dots, w_L$ . При распознавании нового объекта  $s_*$  по признаковому описанию  $x_*$  вычисляются проекции на  $w_1, \dots, w_L$ :

$$\gamma_1(x_*) = \frac{(w_1, x_*^t)}{|w_1|}, \dots, \gamma_L(x_*) = \frac{(w_L, x_*^t)}{|w_L|}.$$

Распознаваемый объект относится к тому классу, соответствующему максимальной величине проекции. Распознавание может производиться также по величинам

$$[\gamma_1(x_*) - b_1], \dots, [\gamma_L(x_*) - b_L].$$

# Логистическая регрессия.

Целью логистической регрессии является аппроксимация плотности условных вероятностей классов в точках признакового пространства. При этом аппроксимация производится с использованием логистической функции.

$$g(z) = \frac{e^z}{1 + e^z} = \frac{1}{1 + e^{-z}}$$

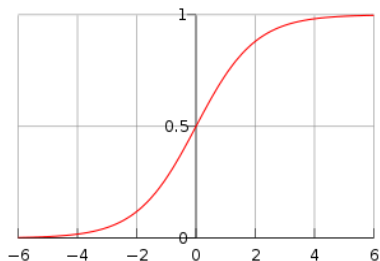


Рис 1. Логистическая функция.



В методе логистическая регрессия связь условной вероятности класса  $K$  с прогностическими признаками осуществляются через переменную  $Z$ , которая задаётся как линейная комбинация признаков:

$$z = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n.$$

Условная вероятность  $K$  в точке векторного пространства  $\mathbf{x}_* = (x_{1*}, \dots, x_{n*})$  задаётся в виде

$$P(K | \mathbf{x}) = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n}} = \frac{1}{1 + e^{-\beta_0 - \beta_1 X_1 - \dots - \beta_n X_n}} \quad (11)$$

Оценки регрессионных параметров  $\beta_0, \beta_1, \dots, \beta_n$  могут быть вычислены по обучающей выборке с помощью различных вариантов метода максимального правдоподобия. Предположим, что объекты обучающей выборки сосредоточены в точках признакового пространства из множества  $\tilde{\mathbf{x}} = \{\mathbf{x}_1, \dots, \mathbf{x}_r\}$ . При этом распределение объектов обучающей выборки по точкам задаётся с помощью набора пар  $\{(m_1, k_1), \dots, (m_r, k_r)\}$ , где  $m_i$  - общее число объектов в точке  $\mathbf{x}_i$ ,  $k_i$  - число объектов класса  $K$  в точке  $\mathbf{x}_i$ . Вероятность данной конфигурации подчиняется распределению Бернулли. Введём обозначение  $\varrho(\mathbf{x}) = P(K | \mathbf{x})$ . Оценка вектора регрессионных параметров  $\beta = (\beta_0, \dots, \beta_n)$  может быть получена с помощью метода максимального правдоподобия. Функция правдоподобия может быть записана в виде

$$L(\beta, \tilde{\mathbf{x}}) = \prod_{j=1}^r C_{m_i}^{k_i} [\varrho(\mathbf{x})_j]^{k_j} [1 - \varrho(\mathbf{x})_j]^{(m_j - k_j)} \quad (12)$$

Принимая во внимание справедливость равенств

$$\frac{\varrho(\mathbf{x})}{1 - \varrho(\mathbf{x})} = e^{\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n},$$

$$1 - \varrho(\mathbf{x}) = \frac{1}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n}},$$

приходим равенству

$$L(\boldsymbol{\beta}, \tilde{\mathbf{x}}) = \prod_{j=1}^r C_{m_i}^{k_i} e^{k_i \beta_0 + \beta_1 x_{j1} + \dots + \beta_n x_{jn}} \frac{1}{(1 + e^{\beta_0 + \beta_1 x_{j1} + \dots + \beta_n x_{jn}})^{m_i}} \quad (13)$$

Поиск оптимального значения параметров удобнее производить, решая задачу максимизации логарифма функции правдоподобия, который в нашем случае принимает вид:

$$\begin{aligned}\ln[L(\boldsymbol{\beta}, \tilde{\boldsymbol{x}})] &= \sum_{j=1}^r \ln C_{m_j}^{k_j} + \sum_{j=1}^r [k_j(\beta_0 + \beta_1 x_{j1} + \dots + \beta_n x_{jn})] + \\ &+ \sum_{j=1}^r m_j \ln\left(\frac{1}{1 + e^{\beta_0 + \beta_1 x_{j1} + \dots + \beta_n x_{jn}}}\right)\end{aligned}$$

Простым, но достаточно эффективным подходом к решению задач распознавания является метод  $k$ -ближайших соседей. Оценка условных вероятностей  $P(K_i | \mathbf{x})$  ведётся по ближайшей окрестности  $V_k$  точки  $\mathbf{x}$ , содержащей  $k$  признаковых описаний объектов обучающей выборки. В качестве оценки выступает отношение  $\frac{k_i}{k}$ , где  $k_i$  - число признаковых описаний объектов обучающей выборки из  $K_i$  внутри  $V_k$ . Окрестность  $V_k$  задаётся с помощью функции расстояния  $\rho(\mathbf{x}', \mathbf{x}'')$  заданной на декартовом произведении  $\tilde{X} \times \tilde{X}$ , где  $\tilde{X}$  - область допустимых значений признаковых описаний. В качестве функции расстояния может быть использована стандартная евклидова метрика. То есть расстояние между двумя векторами  $\mathbf{x}' = (x'_1, \dots, x'_n)$  и  $\mathbf{x}'' = (x''_1, \dots, x''_n)$

$$\rho(\mathbf{x}', \mathbf{x}'') = \sqrt{\frac{1}{n} \sum_{i=1}^n (x'_i - x''_i)^2}.$$

Для задач с бинарными признаками в качестве функции расстояния может быть использована метрика Хэмминга, равная числу совпадающих позиций в двух сравниваемых признаковых описаниях. Окрестность  $V_k$  ищется путём поиска в обучающей выборке  $\tilde{S}_t$  векторных описаний, ближайших в смысле выбранной функции расстояний, к описанию  $x_*$  распознаваемого объекта  $s_*$ . Единственным параметром, который может быть использован для настройки (обучения) алгоритмов в методе  $k$ -ближайших соседей является собственно само число ближайших соседей. Для оптимизации параметра  $k$  обычно используется метод, основанный на скользящем контроле. Оценка точности распознавания производится по обучающей выборке при различных  $k$  и выбирается значение данного параметра, при котором полученная точность максимальна.

Каждый алгоритм распознавания классов  $K_1, \dots, K_L$  может быть представлен как последовательное выполнение распознающего оператора  $\mathbf{R}$  и решающего правила :

$$A = \mathbf{R} \otimes C.$$

Оператор оценок вычисляет для распознаваемого объекта  $s$  вещественные оценки  $\gamma_1, \dots, \gamma_L$  за классы  $K_1, \dots, K_L$  соответственно.

Множество (модель) алгоритмов  $\widetilde{W} = \{A : \widetilde{X} \rightarrow \widetilde{Y}\}$  внутри которого производится поиск оптимального алгоритма прогнозирования вместе со способом решения оптимизационной задачи будем называть методом прогнозирования или методом распознавания, если прогнозируемая величина принадлежит конечному множеству. В качестве примера рассмотрим известный метод решения задачи распознавания – Линейная машина



Метод «Линейная машина» предназначен для решения задачи распознавания с классами  $K_1, \dots, K_L$ . Алгоритм распознавания имеет следующий вид. В процессе обучения классам  $K_1, \dots, K_L$  ставятся в соответствие линейные функции от переменных  $X_1, \dots, X_n$ :

$$\gamma_1(X_1, \dots, X_n) = w_0^1 + w_1^1 X_1 + \dots + w_n^1 X_n$$

.....

$$\gamma_L(X_1, \dots, X_n) = w_0^L + w_1^L X_1 + \dots + w_n^L X_n.$$

Таким образом алгоритм распознавания задаётся параметров

$$w_0^1, \dots, w_n^1$$

.....

$$w_0^L, \dots, w_n^L$$

Пусть требуется распознать объект  $s^*$ , описание которого задаётся вектором  $\mathbf{x}^*$ . Вычисляются значения функций  $\gamma_1, \dots, \gamma_L$  в точке  $\mathbf{x}^*$ . Объект  $s^*$  будет отнесён классу  $K_i$ , если выполняется набор неравенств

$$\gamma_i(\mathbf{x}^*) > \gamma_j(\mathbf{x}^*),$$

где  $j \in \{1, \dots, L\} \setminus \{i\}$ .

Максимальная точность на выборке  $\tilde{S}_t$  соответствует выполнению максимального числа блоков неравенств:

$$\gamma_{J(1)}(\mathbf{x}_1) > \gamma_i(\mathbf{x}_1), i \in \{1, \dots, L\} \setminus \{J(1)\} \quad (14)$$

.....

$$\gamma_{J(m)}(\mathbf{x}_m) > \gamma_i(\mathbf{x}_m), i \in \{1, \dots, L\} \setminus \{J(m)\}.$$

Каждый из блоков соответствует одному из объектов выборки  $\tilde{S}_t$  и включает  $L - 1$  неравенств. Таким образом суммарное число неравенств во всех блоках составляет  $m(L - 1)$ . Каждое из неравенства из системы (1) соответствует сравнению оценки вектора  $\mathbf{x}_r$  за класс  $K_{J(r)}$  с оценкой за класс  $K_i \neq K_{J(r)}$

Имеется задача распознавания с 3-я классами и 2-я признаками. Предполагается, что с использованием метода ЛМ для каждого класса найдены линейные разделяющие функции:

- $\gamma_1(X_1, X_2) = 4 + 2X_1 - X_2$ ;
- $\gamma_2(X_1, X_2) = -2 + X_1 - 3X_2$ ;
- $\gamma_3(X_1, X_2) = 1 + X_1 - 2X_2$ .

Область, где одновременно выполняются неравенства

- $\gamma_1(X_1, X_2) > \gamma_2(X_1, X_2)$ ;
- $\gamma_1(X_1, X_2) > \gamma_3(X_1, X_2)$ ;

соответствует классу 1.

Последняя система эквивалентна неравенствам

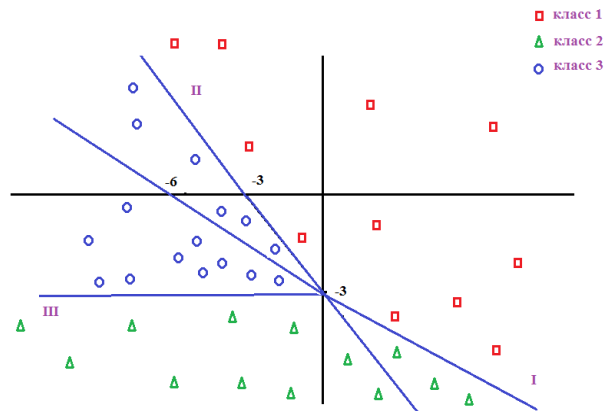
- $6 + X_1 + 2X_2 > 0$  (I),
- $3 + X_1 + X_2 > 0$  (II).

Данные неравенства задают граничные прямые на плоскости, которые обозначены римскими цифрами (I) и (II) соответственно. Область на плоскости, соответствующая классу 1, обозначена красными квадратиками. Предположим, что точка на плоскости не принадлежит классу 1. Тогда она принадлежит классу 2, если выполняется неравенство:

$$\gamma_1(X_1, X_2) > f_3(X_1, X_2),$$

которое эквивалентно неравенству  $X_2 < -3$ . Область на плоскости, соответствующая классу 2, обозначена зелёными треугольниками. Область, соответствующая классу 3 обозначена синими кружками.

# Линейная машина. Пример



Был разработан эффективный метод поиска оптимальных весов

$$w_0^1, \dots, w_n^1$$

.....

$$w_0^L, \dots, w_n^L,$$

. который называется релаксационным алгоритмом.