Лекция 4

Задачи прогнозирования, Линейная машина, Теоретические методы оценки обобщающей способности,

Лектор – *Сенько Олег Валентинович*

Курс «Математические основы теории прогнозирования»
4-й курс, III поток

Содержание лекции

Использование формулы Байеса

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

$$P(K_i \mid \boldsymbol{x}) = \frac{P(\boldsymbol{x} \mid K_i)P(K_i)}{\sum_{i=1}^{L} P(K_i)P(\boldsymbol{x} \mid K_i)} \tag{1}$$

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

Использование формулы Байеса

В качестве оценок априорных вероятностей

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

могут быть взяты доли объектов соответствующих классов в обучающей выборке. Условные вероятности $P(\boldsymbol{x} \mid K_1), \ldots, P(\boldsymbol{x} \mid K_L)$ могут оцениваться на основании сделанных предположений. Например, может быть использовано предположение о независимости переменных для каждого из классов. В последнем случае вероятность $P(\boldsymbol{x}_j \mid K_i)$ для вектора $\boldsymbol{x}_k = (x_{j1}, \ldots, x_{jn})$ может быть представлена в виде:

$$P(\boldsymbol{x}_j \mid K_i) = \prod_{i=1}^{n} P(X_j = x_{ki} \mid K_i). \tag{2}$$

Предположим, переменная X_j принимает значения из конечного множества $\widetilde{M}_j^i$ на объектах из класса K_i при $j = 1, \ldots, n$ и $i = 1, \ldots, L$. Предположим, что

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

Наивный байесовский классификатор

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 доля объектов из $\widetilde{S}_t \bigcap K_i$, для которых $X_j = a_{ji}^k$. В случае, если переменные $X_1, \ldots, X_n$ являются непрерывными, формула Байеса может быть записана с использованием

$$P(K_i \mid \boldsymbol{x}) = \frac{p_i(\boldsymbol{x})P(K_i)}{\sum_{i=1}^{L} P(K_i)p_i(\boldsymbol{x})}, \tag{3}$$

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

лотности вероятностей

$$p_1(\boldsymbol{x}), \ldots, p_L(\boldsymbol{x})$$

также могут оцениваться исходя из предположения взаимной независимости переменных $X_1, \ldots, X_n$.

Наивный байесовский классификатор

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

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

где $p_{ji}(X_j)$ - плотность распределения переменной X_j для класса K_i. Плотности $p_{ji}(X_j)$ могут оцениваться в рамках предположения о типе распределения. Например, может использоваться гипотеза о нормальности распределений

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

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

Методы, основанные на формуле Байеса

Методы распознавания, основанные на использовании формулы Байеса в форме (1) и (3) и гипотезе о независимости переменных обычно называют наивными байесовскими классификаторами. Отметим, что знаменатели в правых частях формул (1) и (3) тождественны для всех классов. Поэтому при решении задач распознавания достаточно использовать только числители.

Аппроксимация плотности с помощь многомерного нормального распределения

При решении задач распознавания с помощью формулы Байеса в форме (3) могут использоваться плотности вероятности $p_1(\boldsymbol{x}), \ldots, p_L(\boldsymbol{x})$, в которых переменные $X_1, \ldots, X_n$ не обязательно являются независимыми. Чаще всего используется многомерное нормальное распределения. Плотность данного распределения в общем виде представляется выражением

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} \mid \Sigma \mid^{\frac{1}{2}}} exp[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})^t], \tag{4}$$

где

$\boldsymbol{\mu}$ - математическое ожидание вектора признаков $\boldsymbol{x}$; Σ - матрица ковариаций признаков $X_1, \ldots, X_n$; $\mid \Sigma \mid$ -детерминант матрицы Σ.

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

Аппроксимация плотности с помощь многомерного нормального распределения

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

$$\hat{\boldsymbol{\mu}}_i = \frac{1}{m_i} \sum_{s_j \in \widetilde{S}_t \bigcap K_i} \boldsymbol{x}_j$$

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

$$\hat{\sigma}^i_{kk'} = \frac{1}{m_i} \sum_{s_j \in \widetilde{S}_t \bigcap K_i} (x_{jk} - \mu_{ik})(x_{jk'} - \mu_{ik'}),$$

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

Использование формулы Байеса. Многомерное нормальное распределение

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

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

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

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

Использование формулы Байеса. Многомерное нормальное распределение

Таким образом объект с признаковым описанием $\boldsymbol{x}$ будет отнесён построенной выше аппроксимацией байесовского классификатора к классу, для которого оценка $g_i(\boldsymbol{x})$ 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.

Использование формулы Байеса. Многомерное нормальное распределение

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 $\boldsymbol{w}$, чтобы средние значения проекции на него объектов обучающей выборки из классов K_1 и K_2 максимально различались. Проекцией произвольного вектора $\boldsymbol{x}$ на направление $\boldsymbol{w}$ является отношение

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

В качестве меры различий проекций классов на используется функционал

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

где

$$\hat{X}_{\boldsymbol{w}i} = \frac{1}{m_i} \sum_{s_j \in \widetilde{S}_t \bigcap K_i} \frac{(\boldsymbol{w}\boldsymbol{x}_j^t)}{\mid \boldsymbol{w} \mid}$$

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

$$\hat{d}_{\boldsymbol{w}i} = \frac{1}{m_i} \sum_{s_j \in \widetilde{S}_t \bigcap K_i} [\frac{(\boldsymbol{w}\boldsymbol{x}_j^t)}{\mid \boldsymbol{w} \mid} - \hat{X}_{\boldsymbol{w}i}]^2$$

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

Линейный дискриминант Фишера.

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

$$\boldsymbol{w} = \hat{\Sigma}_{12}^{-1}(\hat{\boldsymbol{\mu}}_1^t - \hat{\boldsymbol{\mu}}_2^t), \tag{5}$$

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 $\boldsymbol{x}_*$ производится по величине его проекции на направление $\boldsymbol{w}$:

$$\gamma(\boldsymbol{x}_*) = \frac{(\boldsymbol{w}, \boldsymbol{x}_*^t)}{\mid \boldsymbol{w} \mid}. \tag{6}$$

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

Линейный дискриминант Фишера.

Граничный параметр b подбирается по обучающей выборке таким образом, чтобы проекции объектов разных классов на оптимальное направление $\boldsymbol{w}$ оказались бы максимально разделёнными. Простой, но эффективной, стратегией является выбор в качестве порогового параметра b средней проекции объектов обучающей выборки на $\boldsymbol{w}$.
Метод ЛДФ легко обобщается на случай с несколькими классами. При этом исходная задача распознавания классов $K_1, \ldots, K_L$ 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 $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_L$.

Линейный дискриминант Фишера.

В результате получается набор из L направлений $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_L$. При распознавании нового объекта s_* по признаковому описанию $\boldsymbol{x}_*$ вычисляются проекции на $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_L$:

$$\gamma_1(\boldsymbol{x}_*) = \frac{(\boldsymbol{w}_1, \boldsymbol{x}_*^t)}{|\boldsymbol{w}_1|}, \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \gamma_L(\boldsymbol{x}_*) = \frac{(\boldsymbol{w}_L, \boldsymbol{x}_*^t)}{|\boldsymbol{w}_L|}.$$

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

$$[\gamma_1(\boldsymbol{x}_*) - b_1], \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, [\gamma_L(\boldsymbol{x}_*) - b_L].$$

Логистическая регрессия.

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

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

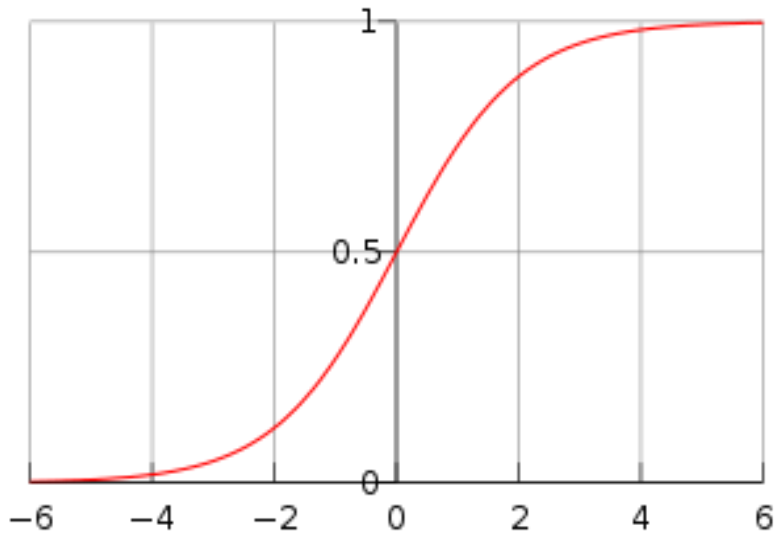

Рис 1. Логистическая функция.

Логистическая регрессия.

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

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

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

$$P(K \mid \boldsymbol{x}) = \frac{e^{\beta_0+\beta_1 X_1+\ldots+\beta_n X_n}}{1 + e^{\beta_0+\beta_1 X_1+\ldots+\beta_n X_n}} = \frac{1}{1 + e^{-\beta_0-\beta_1 X_1-\ldots-\beta_n X_n}} \qquad (7)$$

Логистическая регрессия.

Оценки регрессионных параметров $\beta_0, \beta_1, \ldots, \beta_n$ 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 $\widetilde{\boldsymbol{x}} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r\}$. При этом распределение объектов обучающей выборка по точкам задаётся с помощью набора пар $\{(m_1, k_1), \ldots, (m_r, k_r)\}$, где m_i - общее число объектов в точке $\boldsymbol{x}_i$, k_i - число объектов класса K в точке $\boldsymbol{x}_i$. Вероятность данной конфигурации подчиняется распределению Бернулли. Введём обозначение $\varrho(\boldsymbol{x}) = P(K \mid \boldsymbol{x})$. Оценка вектора регрессионных параметров $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_n)$ может быть получена с помощью метода максимального правдоподобия. Функция правдоподобия может быть записана в виде

$$L(\boldsymbol{\beta}, \widetilde{\boldsymbol{x}}) = \prod_{j=1}^{r} C_{m_i}^{k_i} [\varrho(\boldsymbol{x})_j]^{k_j} [1 - \varrho(\boldsymbol{x})_j]^{(m_j - k_j)} \tag{8}$$

Принимая во внимание справедливость равенств

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

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

приходим равенству

$$L(\boldsymbol{\beta},\widetilde{\boldsymbol{x}}) = \prod_{j=1}^{r} C_{m_i}^{k_i} e^{k_i\beta_0+\beta_1 x_{j1}+\ldots+\beta_n x_{jn}} \frac{1}{\left(1+e^{\beta_0+\beta_1 x_{j1}+\ldots+\beta_n x_j n}\right)^{m_i}} \quad (9)$$

Логистическая регрессия.

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

$$\ln[L(\boldsymbol{\beta}, \widetilde{\boldsymbol{x}})] = \sum_{j=1}^{r} \ln C_{m_j}^{k_j} + \sum_{j=1}^{r} [k_j(\beta_0 + \beta_1 x_{j1} + \ldots + \beta_n x_{jn})] +$$

$$+ \sum_{j=1}^{r} m_j \ln\left(\frac{1}{1 + e^{\beta_0 + \beta_1 x_{j1} + \ldots + \beta_n x_{jn}}}\right)$$

K-ближайших соседей.

Простым, но достаточно эффективным подходом к решению задач распознавания является метод k-ближайших соседей. Оценка условных вероятностей $P(K_i \mid \boldsymbol{x})$ ведётся по ближайшей окрестности V_k точки $\boldsymbol{x}$, 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 $\rho(\boldsymbol{x}', \boldsymbol{x}'')$ заданной на декартовом произведении $\widetilde{X} \times \widetilde{X}$, где $\widetilde{X}$ - область допустимых значений признаковых описаний. В качестве функции расстояния может быть использована стандартная эвклидова метрика. То есть расстояние между двумя векторами $\boldsymbol{x}' = (x'_1, \ldots, x'_n)$ и $\boldsymbol{x}'' = (x''_1, \ldots, x''_n)$

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

K-ближайших соседей.

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 $\widetilde{S}_t$ векторных описаний, ближайших в смысле выбранной функции расстояний, к описанию $\boldsymbol{x}_*$ 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.

Распознавание при заданной точности распознавания некоторых классов

Байесовский классификатор обеспечивает максимальную общую точность распознавания. Однако при решении конкретных практических задач потери, связанные с неправильной классификацией объектов, принадлежащих к одному из классов, значительно превышают потери, связанные с неправильной классификацией объектов других классов. Для оптимизации потерь необходимо использование методов распознавания с учётом предпочтительной точности распознавания для некоторых классов. Одним из возможных подходов является фиксирование порога для точности распознавания одного из классов. Оптимальное решающее правило в задаче распознавания с двумя классами K_1 и K_2 , обеспечивающее максимальную точность распознавания K_2 при фиксированной точности распознавания K_1 , описывается критерием Неймана-Пирсона.

Распознавание при заданной точности распознавания некоторых классов

Критерий Неймана-Пирсона Максимальная точность распознавания K_2 при точности распознавания K_1 равной α обеспечивается правилом: Объект с описанием $\boldsymbol{x}$ относится в класс K_1 , если

$$P(K_1 \mid \boldsymbol{x}) \geq \eta P(K_2 \mid \boldsymbol{x})$$

где параметр η определяется из условия

$$\int_{\Theta} P(K_1 \mid \boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x} = \alpha$$

$$\Theta = \{\boldsymbol{x} \mid P(K_1 \mid \boldsymbol{x}) \geq \eta P(K_2 \mid \boldsymbol{x})\}$$

$p(\boldsymbol{x})$ - плотность распределения $K_1 \bigcup K_2$ в точке $\boldsymbol{x}$. Критерий Неймана-Пирсона может быть использован, если известны плотности распределения распознаваемых классов. Плотности могут быть восстановлены в рамках Байесовских методов обучения на основе гипотез о виде распределений. ,

Распознавание при заданной точности распознавания некоторых классов

Критерий Неймана-Пирсона может быть использован, если известны плотности распределения распознаваемых классов. Плотности могут быть восстановлены в рамках Байесовских методов обучения на основе гипотез о виде распределений. Однако существуют эффективные средства регулирования точности распознавания при предпочтительности одного из классов, которые не требуют гипотез о виде распределения. Данные средства основаны на структуре распознающего алгоритма. Каждый алгоритм распознавания классов $K_1, \ldots, K_l$ может быть представлен как последовательное выполнение распознающего оператора $\mathbf{R}$ и решающего правила :

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

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

Распознавание при заданной точности распознавания некоторых классов

Решающее правило производит отнесение объекта s по вектору оценок $\gamma_1, \ldots, \gamma_L$ к одному из классов. Распространённым решающим правилом является простая процедура, относящая объект в тот класс, оценка за который максимальна.
В случае распознавания двух классов K_1 и K_2 распознаваемый объект s будет отнесён к классу K_1 , если $\gamma_1(s) - \gamma_2(s) > 0$ и классу K_2 в противном случае. Назовём приведённое выше правило правилом $\boldsymbol{C}(0)$. Однако точность распознавания правила $\boldsymbol{C}(0)$ может оказаться слишком низкой для того, чтобы обеспечить требуемую величину потерь, связанных с неправильной классификацией объектов, на самом деле принадлежащих классу K_1 . Для достижения необходимой величины потерь может быть использовано пороговое решающее правило $\mathbf{C}(\delta)$.

Распознавание при заданной точности распознавания некоторых классов

Прравило $\mathbf{C}(\delta)$: распознаваемый объект s будет отнесён к классу K_1 , если $\gamma_1(s) - \gamma_2(s) > \delta$ и классу K_2 в противном случае. Обозначим через $p_{ci}(\delta, s)$ вероятность правильной классификации правилом $\mathbf{C}(\delta)$ объекта s , на самом деле принадлежащего K_i , $i \in \{1, 2\}$. При $\delta < 0$ $p_{c1}(\delta, s) \geq p_{c1}(0, s)$, но $p_{c2}(\delta, s) \leq p_{c2}(0, s)$. Уменьшая δ , мы увеличиваем $p_{c1}(\delta, s)$ и уменьшаем $p_{c2}(\delta, s)$. Напротив, увеличивая δ , мы уменьшаем $p_{c1}(\delta, s)$ и увеличиваем $p_{c2}(\delta, s)$. Зависимость между $p_{c1}(\delta, s)$ и $p_{c2}(\delta, s)$ может быть приближённо восстановлена по обучающей выборке $\widetilde{S}_t$, включающей описания объектов $\{s_1, \ldots, s_m\}$.

Распознавание при заданной точности распознавания некоторых классов

Пусть
$$\begin{pmatrix} \gamma_1(s_1) & \dots & \gamma_1(s_m) \\ \gamma_2(s_1) & \dots & \gamma_2(s_m) \end{pmatrix}$$
является матрицей оценок за классы K_1 и K_2 объектов из $\widetilde{S}_t$. Пусть

$$\gamma(s_1) = \gamma_1(s1) - \gamma_2(s_1), \dots, \gamma(s_m) = \gamma_1(s_m) - \gamma_2(s_m).$$

Предположим, что величины $[\gamma(s_1), \dots, \gamma(s_m)]$ принимают r значений $\Gamma_1, \dots, \Gamma_r$. Рассмотрим r пороговых решающих правил $[\boldsymbol{C}(\Gamma_1), \dots, \boldsymbol{C}(\Gamma_r)]$ Для каждого из правил $\mathbf{C}(\Gamma_i)$ обозначим через $\nu_{c1}(\Gamma_i)$ долю K_1 среди объектов обучающей выборки, удовлетворяющих условию $\gamma(s_*) \geq \Gamma_i$, а через $\nu_{c2}(\Gamma_i)$ обозначим долю K_2 среди объектов обучающей выборки, удовлетворяющих условию $\gamma(s_*) < \Gamma_i$.

Распознавание при заданной точности распознавания некоторых классов. ROC анализ.

Отобразим результаты расчётов

$$\{[\nu_{c1}(\Gamma_1), \nu_{c2}(\Gamma_1)] \ldots, [\nu_{c1}(\Gamma_r), \nu_{c2}(\Gamma_r)]\}$$

как точки на в декартовой системе координат. Соединив точки отрезками прямых, получим ломаную линию (l), соединяющую точки (1,0) и (0,1). Данная линия графически отображает аппроксимацию по обучающей выборке взаимозависимости между $p_{c1}(\delta, s)$ и $p_{c2}(\delta, s)$ при всевозможных значениях δ . Соответствующий пример представлен на рисунке 2.

Распознавание при заданной точности распознавания некоторых классов. ROC анализ.

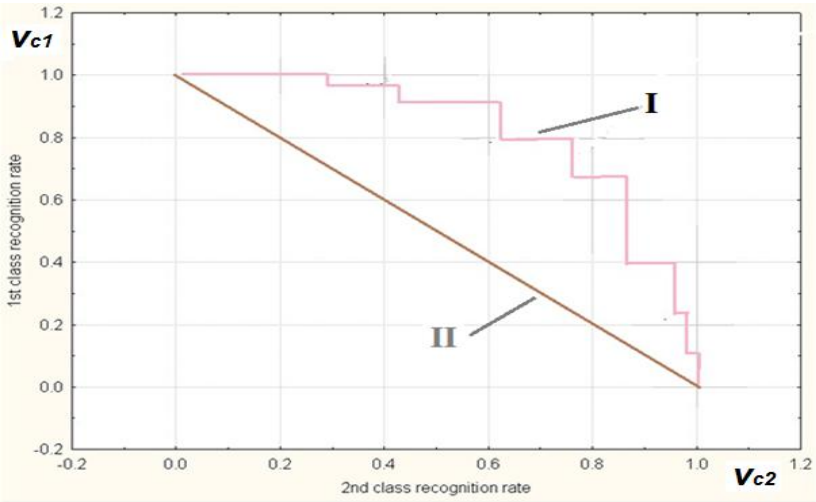

Рис.2

Взаимозависимость между ν_{c1} и ν_{c2} наиболее полно оценивает эффективность распознающего оператора $\mathbf{R}$. Отметим, что ν_{c1} постепенно убывает по мере роста ν_{c2} . .

Распознавание при заданной точности распознавания некоторых классов. ROC анализ.

Однако сохранение высокого значения ν_{c1} при высоких значениях ν_{c2} соответствует существованию решающего правила, при котором точность распознавания обоих классов высока. Таким образом эффективному распознающему оператору соответствует близость линии I к прямой, связывающей точки (0,1) и (1,1). То есть наиболее высокой эффективности соответствует максимально большая площадь под линией I. Отсутствию распознающей способности соответствует близость к прямой II, связывающей точки (0, 1) и (1,0).На рисунке 3 сравниваются линии, характеризующие эффективность распознающих операторов, принадлежащих к трём методам распознавания, при решении задачи распознавания двух видов аутизма по психометрическим показателям .

Распознавание при заданной точности распознавания некоторых классов. ROC анализ.

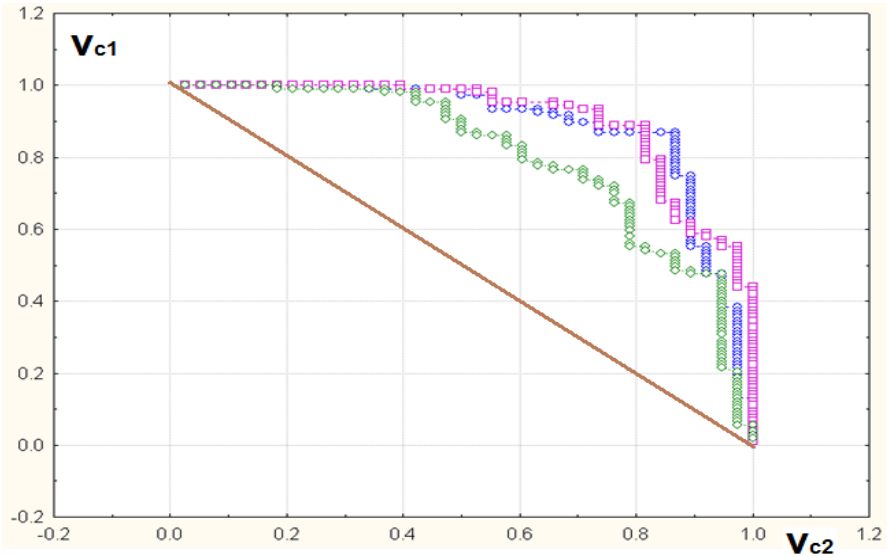

Рис.3

- - линейный дискриминант Фишера;
- - метод опорных вектор;
- - статистически взвешенные синдромы.

ROC анализ.

Методы распознавания используются при решении многих задач идентификации объектов, представляющих важность для пользователя. Эффективность идентификации для таких задач удобно описывать в терминах: «Чувствительность» - доля правильно распознанных объектов целевого класса «Ложная тревога» - доля объектов ошибочно отнесённых в целевой класс. Пример кривой, связывающей параметры «Чувствительность» и «Ложная тревога» представлен на рисунке 4. Анализ, основанный на построении и анализе линий, связывающих параметры «Чувствительность» и «Ложная тревога» принято называть анализом Receiver Operating Characteristic или ROC-анализом. Линии, связывающих параметры «Чувствительность» и «Ложная тревога» принято называть ROC-кривыми.

ROC анализ.

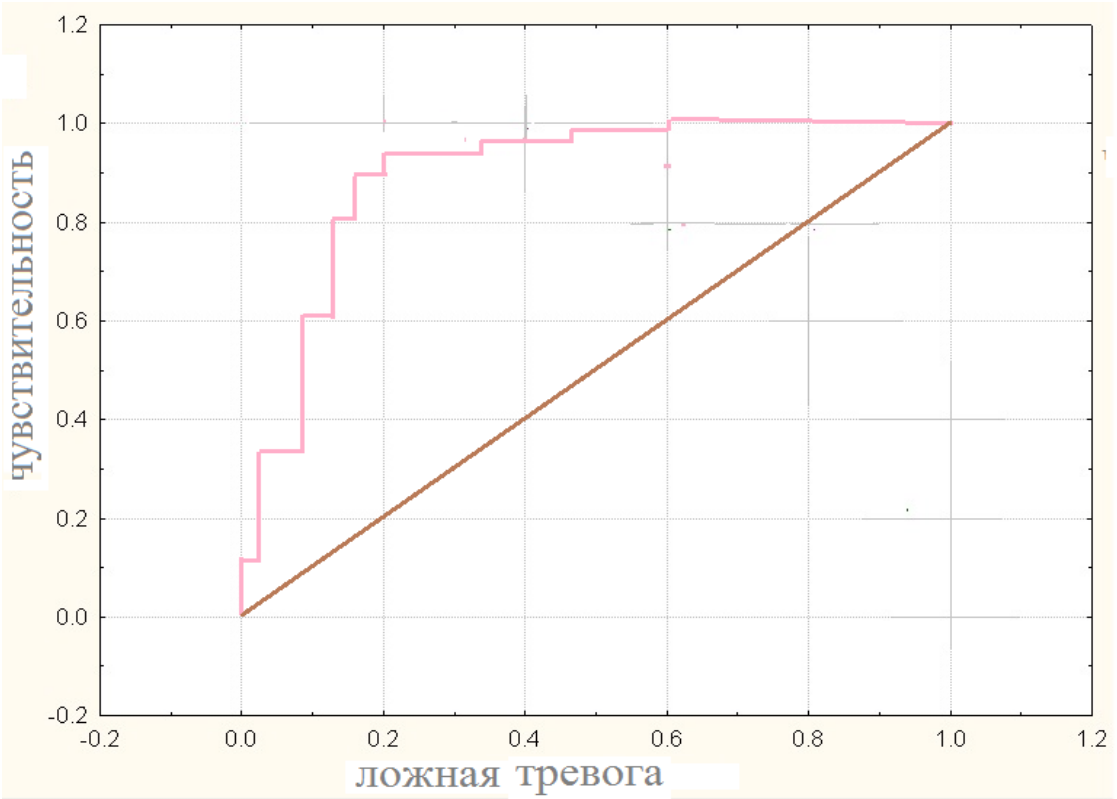

Рис.4. Пример ROC кривой