Байесовский классификатор. Оценка плотности распределения.

Евгений Борисов

Байесовский классификатор

Вероятностное пространство математическая модель случайного эксперимента (опыта)

$$(\Omega, A, P)$$

Ω - элементарные исходы эксперимента, множество объектов ω.

A - случайные события, набор подмножеств **Ω**.

$A \ni \Omega$ - достоверное событие

$A \ni \varnothing$ - невозможное событие

P - функция вероятности $P: A \longrightarrow [0,1]$

$0 \leqslant P(a) \leqslant 1$ вероятность события **a** из **A**

$P(\varnothing)=0$; $P(\Omega)=1$

Случайная величина в пространстве **(Ω,A,P)** это числовая функция

$$X: A \longrightarrow \mathbb{R}$$

типы случайных величин:

дискретные (discrete) - принимающая конечное или счетное число значений
(*Пример*: *частота слов в тексте, количество детей в семье*)

непрерывные (continuous) - принимают значение в определённом интервале
(*Пример*: *рост людей*)

Байесовский классификатор

Вероятностное пространство **(Ω,A,P)**
Ω - элементарные исходы эксперимента
A - случайные события, набор подмножеств **Ω**
P - функция вероятности **P: A ⟶ [0,1]**
X - случайная величина **X: A ⟶** $\mathbb{R}$

случайная величина **X** задаётся распределением вероятностей **F** своих значений

F(x) = P(X ⩽ x)

Рассмотрим интервалы **(x,x+Δx)**, где **Δx** - бесконечно малые приращения **x** для **F(x)**

$$\varphi(x) = \lim_{\Delta x \to 0} \frac{F(x+\Delta x) - F(x)}{\Delta x}$$

Плотностью распределения (вероятности) **φ(x)** непрерывной случайной величины **X** назовём первую производную функции распределения **F(x)**

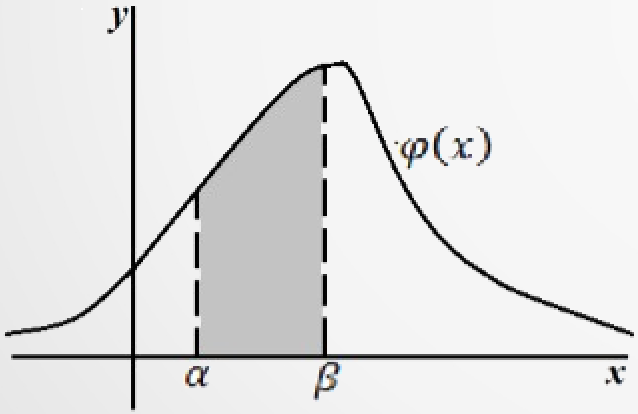
Байесовский классификатор

<u>Вероятностное пространство</u> **(Ω,A,P)**
Ω - элементарные исходы эксперимента
A - случайные события, набор подмножеств **Ω**
P - функция вероятности **P: A ⟶ [0,1]**
X - случайная величина **X: A ⟶ ℝ**
F(x)=P(X⩽x) - распределение вероятностей **P**
φ(x)=F`(x) - плотность распределения **F**

график функции **φ(x)**
плотности распределения **X**

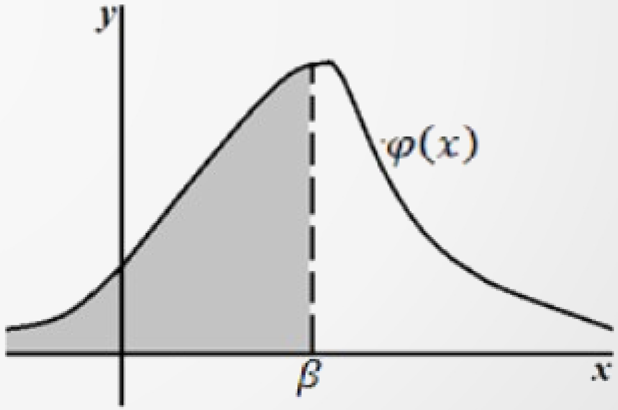
$$P(\Omega)=\int_{-\infty}^{+\infty}\varphi(x)\,dx=1$$

площадь криволинейной трапеции, ограниченной графиком **φ(x)** и прямыми **x=a, x=b, y=0** это вероятность **P(a⩽X⩽b)** попадания **X** в интервал **[a,b]**



$$P(\alpha\leqslant X\leqslant\beta)=\int_{\alpha}^{\beta}\varphi(x)\,dx$$

площадь бесконечной криволинейной трапеции, ограниченной графиком **φ(x)**, прямой **x=b, y=0** это функция распределения **F(b)=P(X⩽b)**



$$F(\beta)=P(X\leqslant\beta)=\int_{-\infty}^{\beta}\varphi(x)\,dx$$

https://spravochnick.ru/matematika/plotnost_raspredeleniya_veroyatnostey_nepreryvnoy_sluchaynoy_velichiny/

Байесовский классификатор

X - объекты,   Y - метки классов

$X \times Y$ - вероятностное пространство с плотностью $p(x, y)$

<u>выборка:</u> $(X' \times Y') \subset (X \times Y)$

<u>Задача</u>**:** построить классификатор с минимальной ошибкой

$$a : X' \rightarrow Y'$$

принцип максимума апостериорной вероятности

$$a(x) = \underset{y \in Y}{argmax}\, P(y|x)$$

Байесовский классификатор

принцип максимума апостериорной вероятности

$$a(x) = \arg\max_{y \in Y} P(y|x)$$

формула Байеса

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

байесовский классификатор

$$a(x) = \arg\max_{y \in Y} \lambda_y P(y) p(x|y)$$

λ_y - потеря для объектов y

$P(y)$ - доля примеров класса y (априорная вероятность)

$p(x|y)$ - плотность класса y

Байесовский классификатор

подходы к оценке плотности распределения

- непараметрический
- параметрический
- смеси распределений

Восстановление плотности

подходы к оценке плотности распределения:

параметрический подход

$$\hat{p}(x) = \varphi(x, \theta)$$

НЕпараметрический подход

$$\hat{p}(x) = \frac{1}{m V(h)} \sum_{j=1}^{m} K\left(\frac{\rho(x, x_j)}{h}\right)$$

смеси распределений

$$\hat{p}(x) = \sum_{j=1}^{k} w_j \varphi_j(x, \theta_j)$$

Байесовский классификатор

«наивный Байес»

допущение: признаки *X* - независимы друг от друга

тогда многомерную плотность можно представить как произведение одномерных плотностей

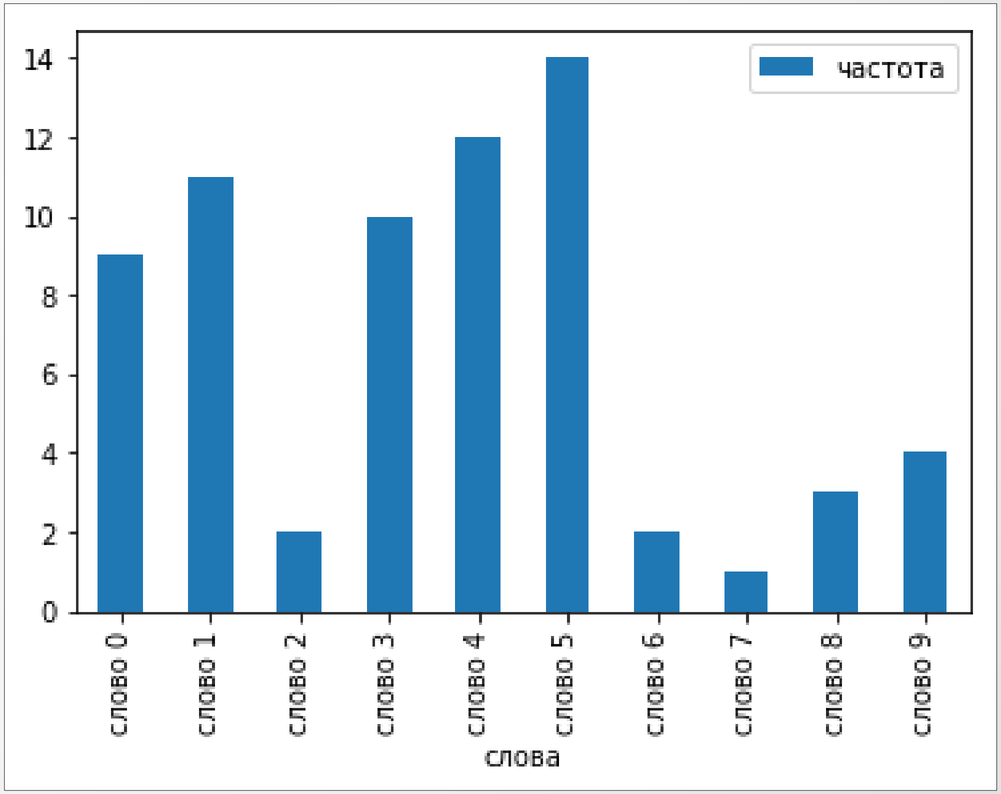
$$p(x|y) = p_1(x_1|y)\dots p_n(x_n|y)$$

Восстановление плотности

НЕпараметрический подход

дискретный случай: гистограмма

$$\hat{p}(x)=\frac{1}{m}\sum_{i=1}^{m}[x=x_i]$$



пример: распределение повторов слов в тексте

Байесовский классификатор

непараметрические методы
оценки плотности распределения

непрерывный случай: эмпирическая оценка, окно ширины h

(доля объектов попавших в отрезок)

$$\hat{p}(x) = \frac{1}{2hm} \sum_{i=1}^{m} \left[|x - x_i| < h \right]$$

$$\hat{p}(x) = \frac{1}{mh} \sum_{i=1}^{m} \frac{1}{2} \left[\frac{|x - x_i|}{h} < 1 \right]$$

Байесовский классификатор

оценка Парзена-Розенблата для класса y

$$\hat{p}(x|y)=\frac{1}{L_y V(h)}\sum_{i:y=y_i} K\left(\frac{\rho(x,x_i)}{h}\right)$$

$K(r)$ - ядро

L_y - количество объектов y

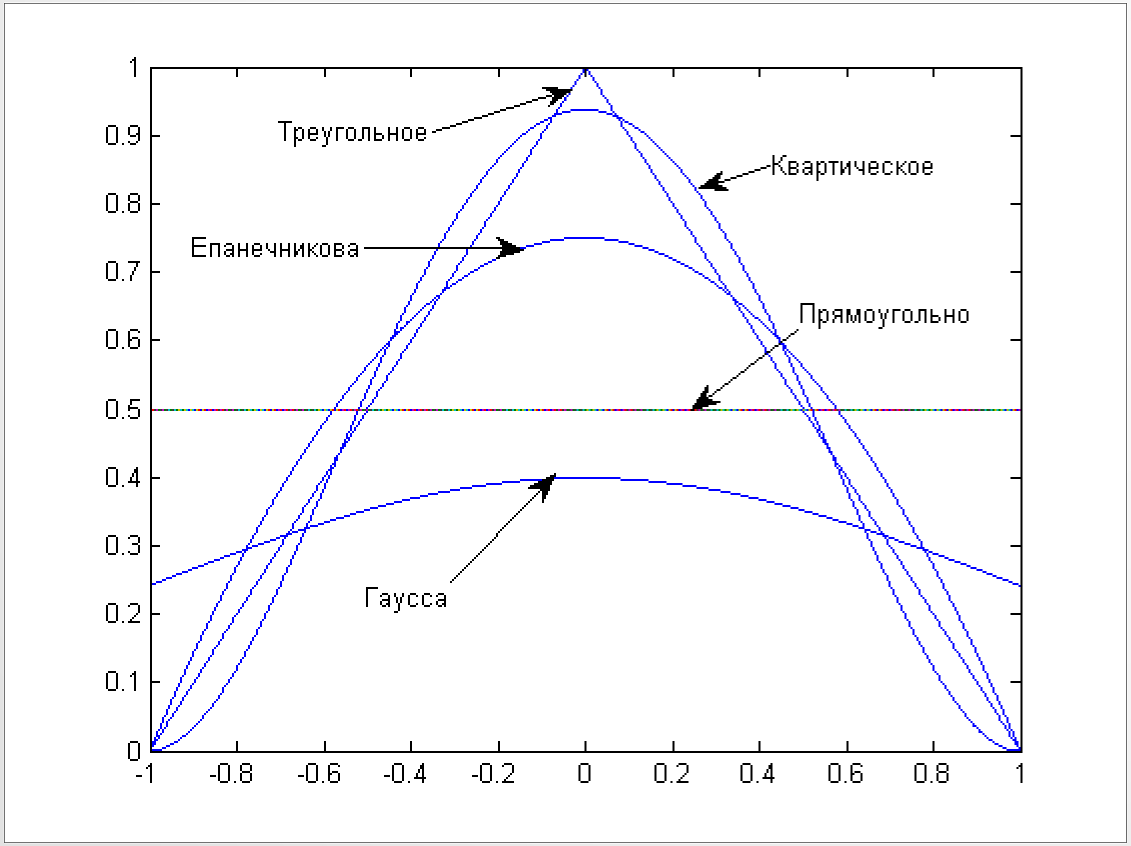
$\rho(x_i,x_j)$ - мера на X

$V(h)$ - нормирующий множитель

Восстановление плотности

функции ядра для сглаживания гистограммы

KDE, Kernel Density Estimation



чётная ф-ция

$$K(r) = K(-r)$$

нормированная

$$\int K(r)\,dr = 1$$

невозрастающая при *r>0*, неотрицательная ф-ция

ядро Епанечникова:

$$K(r) = \frac{3}{4}(1 - r^2);|r| \le 1$$

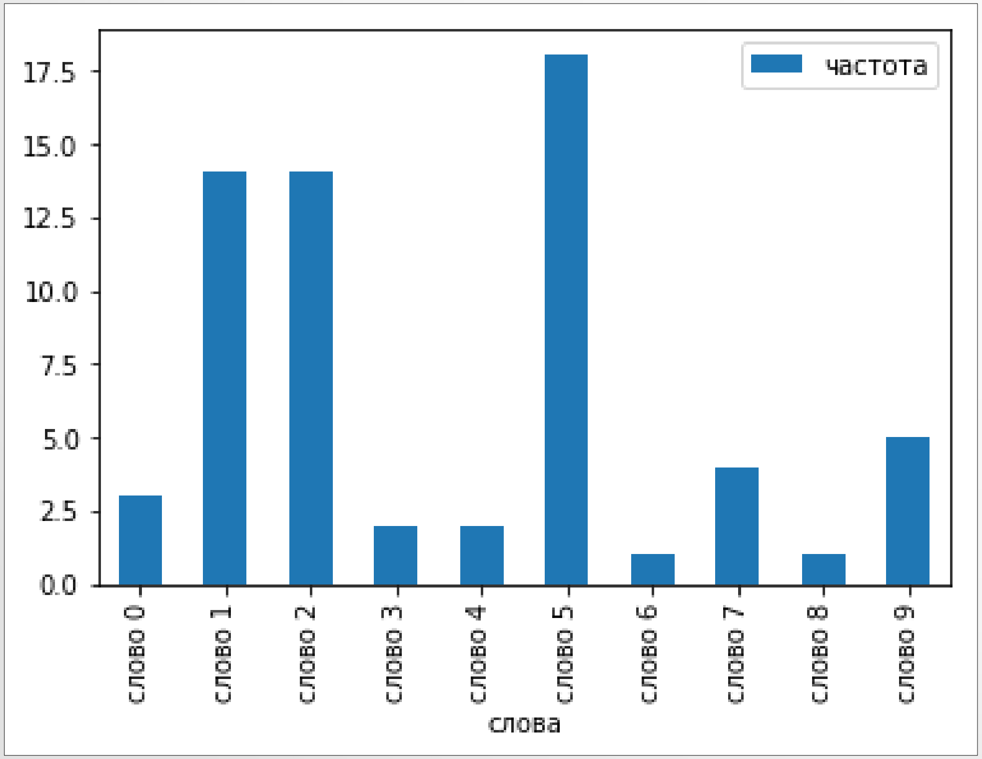
Восстановление плотности

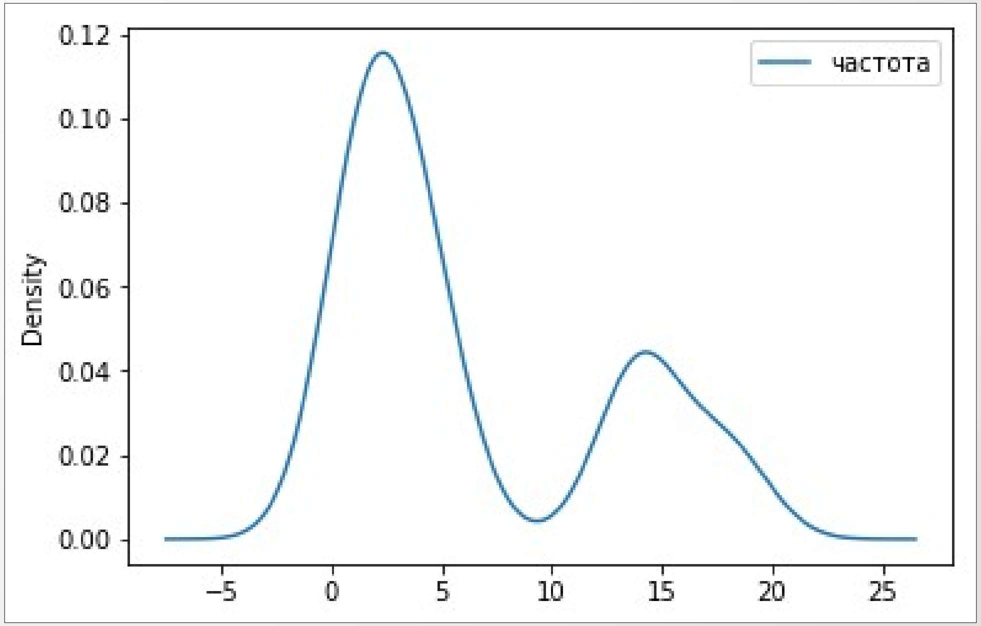
непараметрические методы:
оценка плотности Парзена-Розенблата

ядерное сглаживание (гистограммы)

KDE, Kernel Density Estimation

$$\hat{p}(x) = \frac{1}{m V(h)} \sum_{i=1}^{m} K\left(\frac{\rho(x, x_i)}{h}\right)$$





Байесовский классификатор

$$a(x) = \underset{y \in Y}{argmax}\, \lambda_y P(y)\, p(x|y)$$

Байесовский классификатор, метод Парзеновского окна

$$a(x, X^L, h) = \underset{y \in Y}{argmax}\, \lambda_y P(y) \frac{1}{L_y} \sum_{i:\, y = y_i} K\left(\frac{\rho(x, x_i)}{h}\right)$$

Байесовский классификатор

оценка плотности - параметрический подход

$$\hat{p}(x) = \varphi(x, \theta)$$

Байесовский классификатор

оценка плотности - параметрический подход

$$\hat{p}(x) = \varphi(x, \theta)$$

принцип максимума правдоподобия

$$L(\theta, X) = \sum_{i=1}^{m} \ln \varphi(x_i, \theta) \rightarrow \max_{\theta}$$

Байесовский классификатор

оценка плотности - параметрический подход

$$\hat{p}(x) = \varphi(x, \theta)$$

принцип максимума правдоподобия

$$L(\theta, X) = \sum_{i=1}^{m} \ln \varphi(x_i, \theta) \rightarrow \max_{\theta}$$

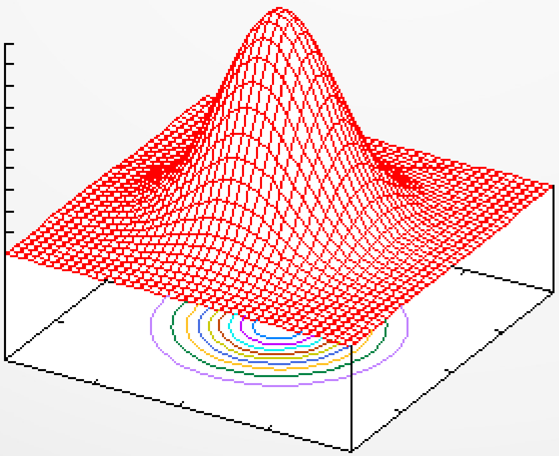
условие оптимума

$$\frac{\partial}{\partial \theta} L(\theta, X) = \sum_{i=1}^{m} \frac{\partial}{\partial \theta} \ln \varphi(x_i, \theta) = 0$$

Байесовский классификатор

допущение: классы имеют n-мерную нормальную плотность

$$p(x|y) = \frac{\exp\left(-\frac{1}{2}(x-\mu_y)^T \Sigma_y^{-1}(x-\mu_y)\right)}{\sqrt{(2\pi)^n \det \Sigma_y}}; y \in Y$$



Байесовский классификатор

Теорема: параметры оценки максимального правдоподобия для n-мерных гауссовских плотностей классов y имеют следующий вид

$$\hat{\mu}_y = \frac{1}{l_y} \sum_{i:y=y_i} x_i \qquad \hat{\Sigma}_y = \frac{1}{l_y} \sum_{i:y=y_i} (x_i - \hat{\mu}_y)(x_i - \hat{\mu}_y)^T$$

Байесовский классификатор

Теорема: параметры оценки максимального правдоподобия для n-мерных гауссовских плотностей классов y имеют следующий вид

$$\hat{\mu}_y = \frac{1}{l_y} \sum_{i:y=y_i} x_i \qquad \hat{\Sigma}_y = \frac{1}{l_y} \sum_{i:y=y_i} (x_i - \hat{\mu}_y)(x_i - \hat{\mu}_y)^T$$

Байесовский классификатор: квадратичный дискриминант

$$a(x) = \underset{y \in Y}{argmax} \left(\ln(\lambda_y P_y) - (x - \hat{\mu}_y)^T \hat{\Sigma}_y^{-1} (x - \hat{\mu}_y) - \frac{1}{2} \ln(det \hat{\Sigma}_y) \right)$$

Байесовский классификатор

Дополнение:

если матрицы ковариаций классов равны

то параметры оценки плотности имеют следующий вид

$$\hat{\mu}_y = \frac{1}{l_y} \sum_{i:y=y_i} x_i \qquad \hat{\Sigma} = \frac{1}{l} \sum_{i=1}^{l} \left(x_i - \hat{\mu}_{yi}\right)\left(x_i - \hat{\mu}_{yi}\right)^T$$

Байесовский классификатор

Дополнение:
если матрицы ковариаций классов равны
то параметры оценки плотности имеют следующий вид

$$\hat{\mu}_y = \frac{1}{l_y} \sum_{i:y=y_i} x_i \qquad \hat{\Sigma} = \frac{1}{l} \sum_{i=1}^{l} (x_i - \hat{\mu}_{y_i})(x_i - \hat{\mu}_{y_i})^T$$

Байесовский классификатор: линейный дискриминант Фишера

$$a(x) = \underset{y \in Y}{argmax} \left(\ln(\lambda_y P_y) - \frac{1}{2} \hat{\mu}_y^T \hat{\Sigma}^{-1} \hat{\mu}_y + x^T \hat{\Sigma}^{-1} \hat{\mu}_y \right)$$

Литература

Борисов Е.С. Методы машинного обучения. 2024
https://github.com/mechanoid5/ml_lectorium_2024_I

Константин Воронцов - Машинное обучение. ШАД Яндекс
https://www.youtube.com/playlist?list=PLJOzdkh8T5kp99tGTEFjH_b9zqEQiiBtC

SciKit-Learn : Naive Bayes
https://scikit-learn.org/stable/modules/naive_bayes.html