

Байесовский классификатор. Оценка плотности распределения.

Евгений Борисов

Байесовский классификатор

X - объекты, Y - метки классов

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

выборка: $(X' \times Y') \subset (X \times Y)$

Задача: построить классификатор с минимальной ошибкой

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

принцип максимума апостериорной вероятности

$$a(x) = \operatorname{argmax}_{y \in Y} P(y|x)$$

Байесовский классификатор

принцип максимума апостериорной вероятности

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

формула Байеса :

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

байесовский классификатор

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

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

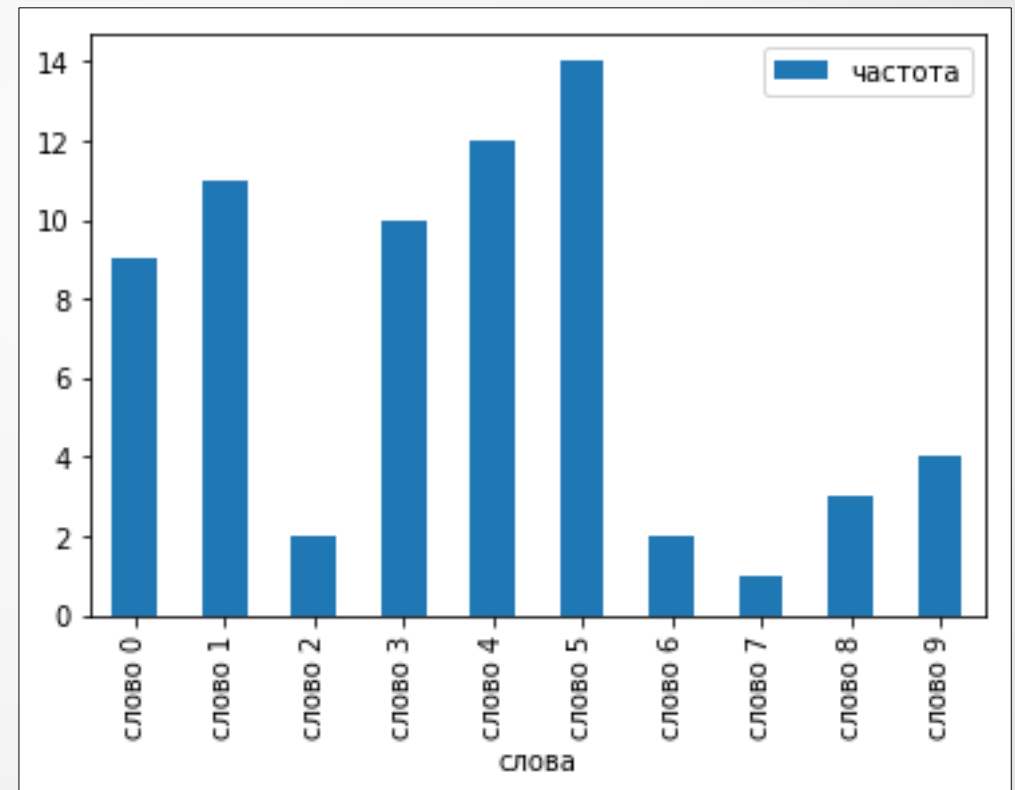
$$p(x|y) = p_1(x_1|y) \dots p_n(x_n|y)$$

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)*

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

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

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

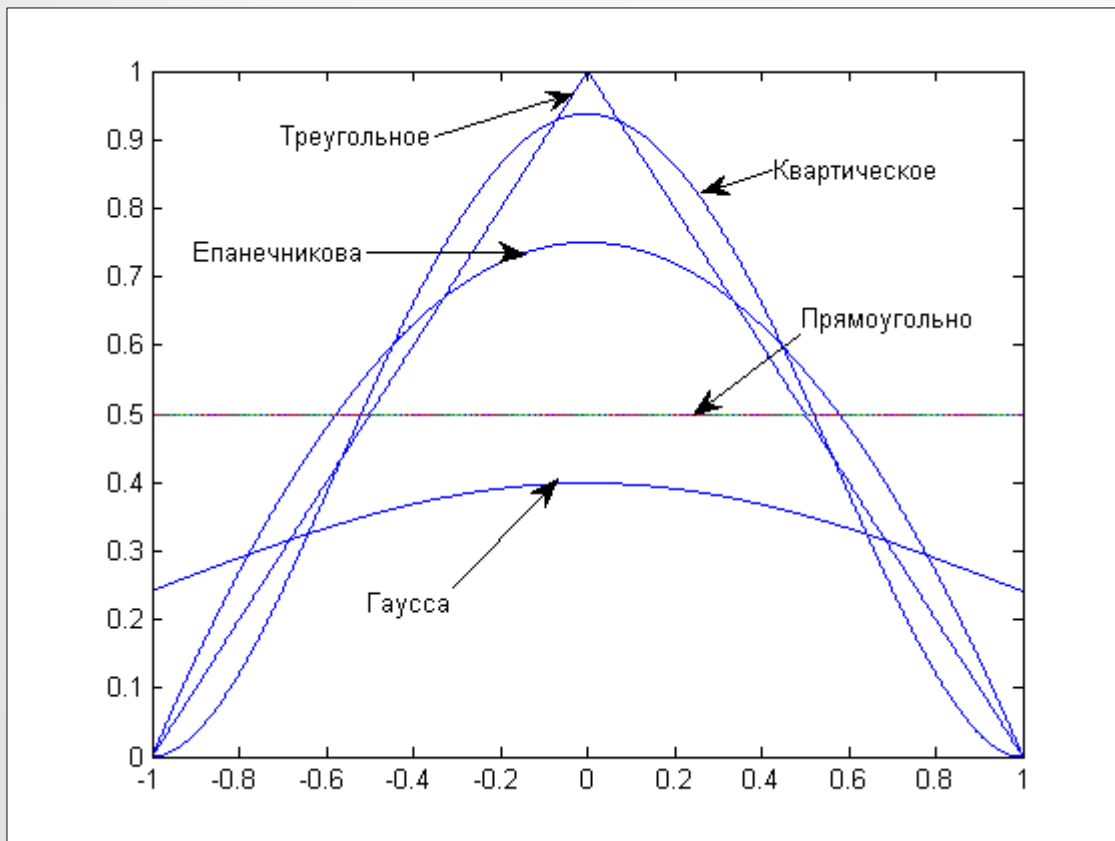
$\rho(x_i, x_j)$ - мера на X

$V(h)$ - 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:

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

Восстановление плотности

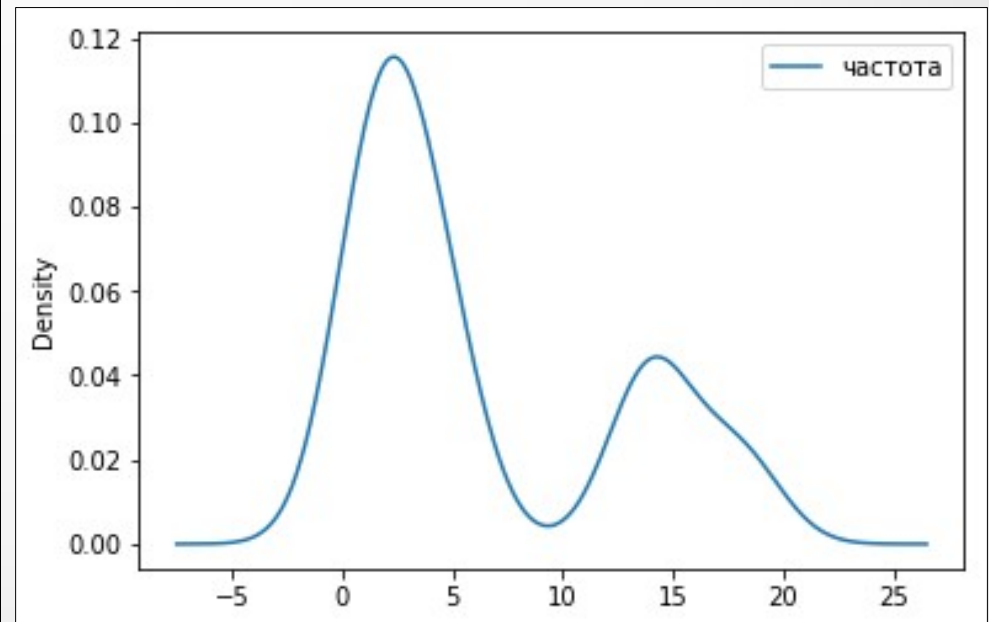
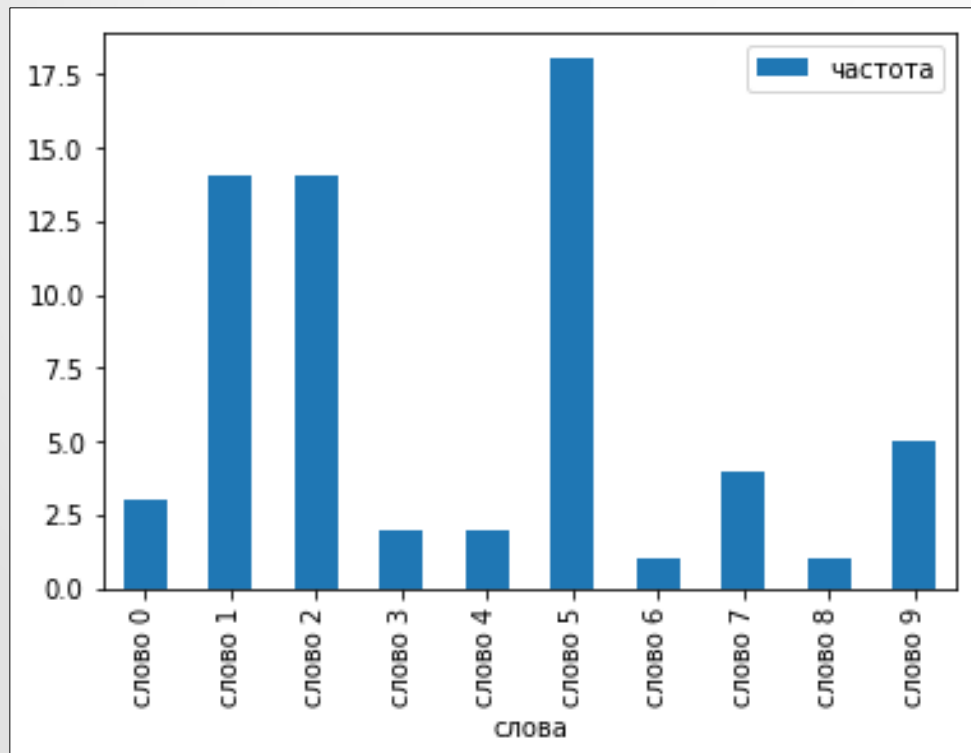
непараметрические методы:

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

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

Байесовский классификатор, метод Парзеневского окна

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

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

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

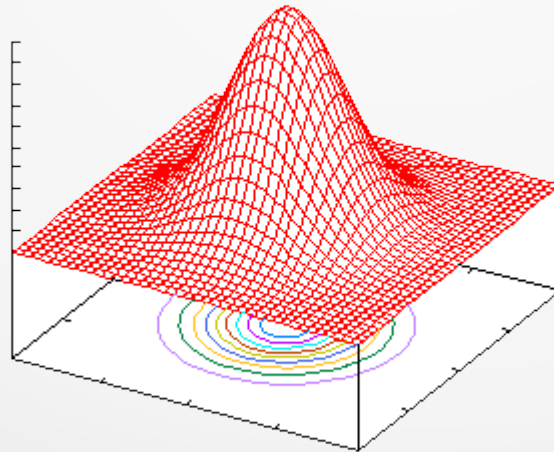
условие оптимума

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

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

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

Байесовский классификатор

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

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

Байесовский классификатор: квадратичный дискриминант

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

Байесовский классификатор

Дополнение:

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

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

Байесовский классификатор

Дополнение:

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

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

Байесовский классификатор: линейный дискриминант Фишера

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

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