Лекция 4:
Байесовский классификатор

Евгений Борисов

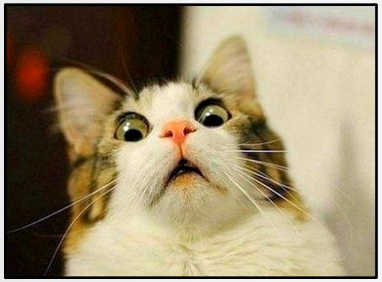
четверг, 11 октября 2018 г.

Классификатор: с чего все начинается?

хорошие и плохие коты

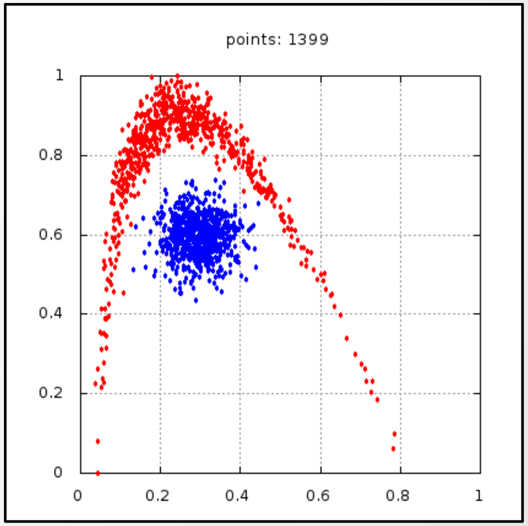
извлекаем признаки

один кот — одна точка



→ [0.14, 12, ..., 345]

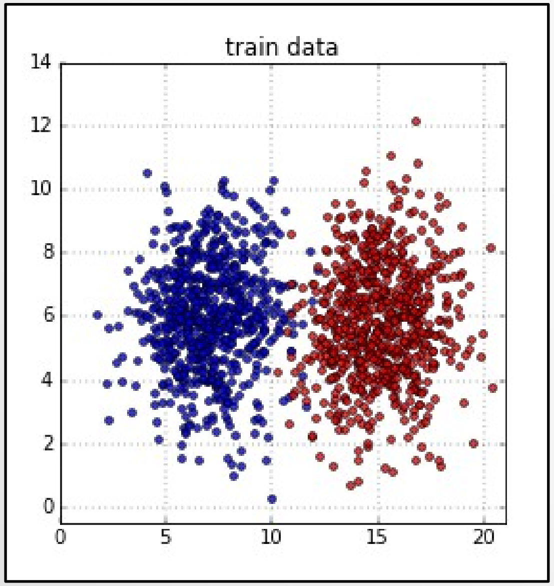
→ [78.0, 20, ..., 177]

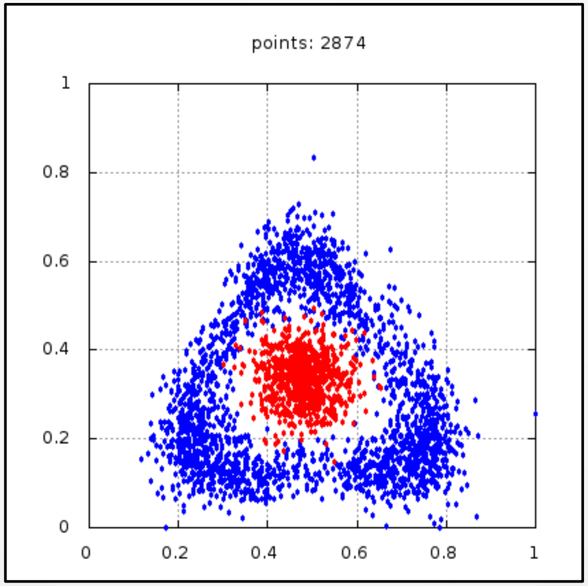


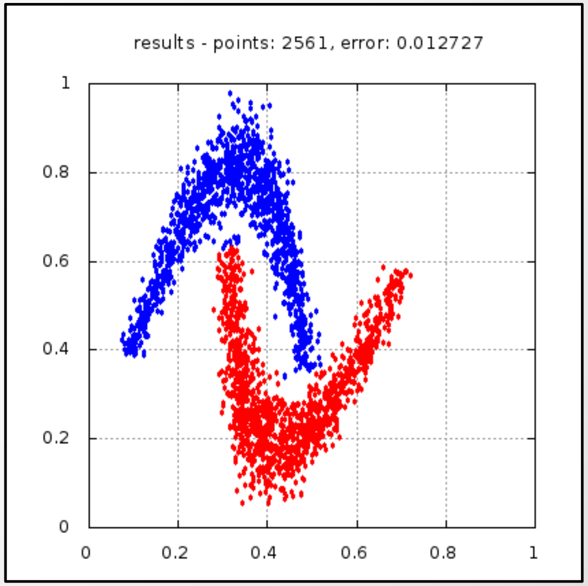
Классификатор

разделения объектов на классы

Детектор котов:  → вектор-признак → есть/нет







Классификатор: о задаче

разделение данных на части (классы)
обучение «с учителем»

Учебный набор: [объект, ответ]

Задача: классификатор
объект → вектор-признак → результат

Обучение: минимизация ошибки
ошибка = результат - правильный ответ

Критерий остановки:
достигнут порог значения ошибки,
и/или порог количества циклов

Классификатор: данные

$$\begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_n^{(1)} & y^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \dots & x_n^{(2)} & y^{(2)} \\ \vdots & \vdots & & \vdots & \vdots \\ x_1^{(m)} & x_2^{(m)} & \dots & x_n^{(m)} & y^{(m)} \end{bmatrix}$$

x - вектор-признак

y - метка класса

n - размер пространства признаков

m - количество примеров

Байесовский классификатор

X - объекты   Y - ответы

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

(x_i, y_i) - выборка

Задача:
найти ф-цию (классификатор)
$a: X \rightarrow Y$ с минимальной ошибкой

Байесовский классификатор

X - объекты Y - ответы

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

(x_i, y_i) - выборка

Задача:
найти ф-цию (классификатор)
a: X → Y с минимальной ошибкой

принцип максимума апостериорной вероятности

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

Байесовский классификатор

принцип максимума апостериорной вероятности

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

P(y) - априорная вероятность класса y

p(x|y) - ф-ция правдоподобия класса y

P(y|x) - апостериорная вероятность класса y

формула Байеса : $P(A|B) = \dfrac{P(A)\, P(B|A)}{P(B)}$

Байесовский классификатор

о функционале среднего риска:

$a: X \to Y$ - классификатор

$A_y = \{\, x \in X \mid a(x) = y \,\},\ y \in Y$ - разбиение **X** на части

Ошибка: объект ***x*** класса ***y*** попал в класс **s**

$$A_s,\ s \neq y$$

Байесовский классификатор

о функционале среднего риска:

$a: X \to Y$ - классификатор

$A_y = \{ x \in X \mid a(x) = y \},\ y \in Y$ - разбиение **X** на части

Ошибка: объект ***x*** класса ***y*** попал в класс **s**

$A_s,\ s \neq y$

Вероятность ошибки: $$P(A_s, y) = \int_{A_s} p(x, y)\, dx$$

Байесовский классификатор

про функционал среднего риска:

Потеря от ошибки: зададим $\lambda_{ys} \geq 0$ для всех пар $(y,s) \in Y \times Y$

Средний риск: мат.ожидание потери классификатора

$$R(a) = \sum_{y \in Y} \sum_{s \in Y} \lambda_{ys} P(A_s, y)$$

Байесовский классификатор

Теорема про оптимальный байесовский классификатор

пусть заданы:
- априорные вероятности классов P(y),
- плотности их распределений p(x|y)
- $\lambda_{ys} \geq 0$ потери от ошибки

тогда минимум среднего риска ***R(a)*** достигается классификатором

$$a(x) = \underset{s \in Y}{\arg\min} \sum_{y \in Y} \lambda_{ys} P(y) p(x|y)$$

Байесовский классификатор

Теорема про оптимальный байесовский классификатор

пусть заданы:
- априорные вероятности классов P(y),
- плотности их распределений p(x|y)
- $\lambda_{ys} \geq 0$ потери от ошибки

тогда минимум среднего риска R(a) достигается классификатором

$$a(x) = \underset{s \in Y}{\operatorname{argmin}} \sum_{y \in Y} \lambda_{ys} P(y) p(x|y)$$

Дополнение:

если $\lambda_{yy}=0$ и $\lambda_{ys} = \lambda_y$ *для всех* $y,s \in Y$ то

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

Байесовский классификатор

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

λ_y - потеря для объектов y

P(y) - априорная вероятность класса y
(доля примеров класса y,
пропорция классов должна соответствовать)

p(x|y) - ф-ция правдоподобия класса y (плотность)

Байесовский классификатор

подходы к оценке плотности распределения:

- непараметрический
- параметрический
- смеси распределений

Байесовский классификатор

параметрический подход к оцениванию плотности

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

Байесовский классификатор

параметрический подход к оценке плотности

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

смеси распределений

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

Байесовский классификатор

параметрический подход к оцениванию плотности

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

смеси распределений

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

НЕпараметрический подход к оцениванию плотности

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

Байесовский классификатор

«наивный Байес»

допущение: признаки X - независимы друг от друга

тогда многомерную плотность
можно представить как произведение одномерных плотностей

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

Байесовский классификатор

непараметрические методы

оценка плотности распределения

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

оценка плотности распределения

непрерывный случай: эмпирическая оценка, окно ширины h

(доля объектов попавших в отрезок)

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

Байесовский классификатор

непараметрические методы

оценка 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 плотности Парзона-Розенблата

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

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

K(r) - ядро

чётная ф-ция K(r)=K(-r)

нормированная $\int K(r)dr=1$

невозрастающая при r>0, неотрицательная ф-ция

Байесовский классификатор

оценка Парзона-Розенблата для класса y

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

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

K(r) - ядро

l_y - количество объектов y

ρ() - мера на X

V(h) - нормирующий множитель

Байесовский классификатор

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

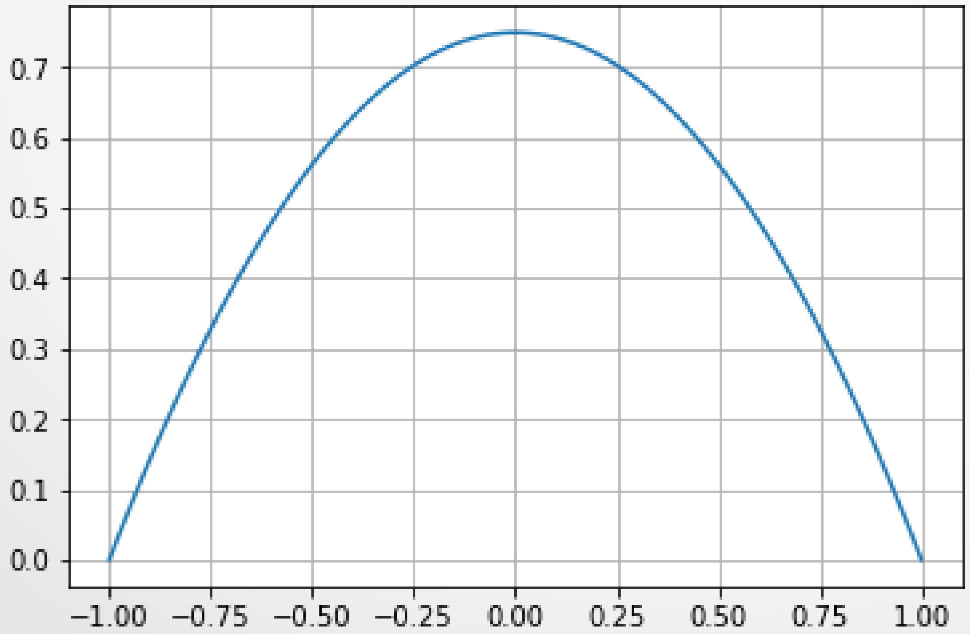
метод Парзоновского окна

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

Байесовский классификатор

ядро Епанечникова

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



Байесовский классификатор

выбор оптимального размера окна h

метод скользящего контроля (Leave One Out, LOO)

параметр h выбираем перебором

проверяем суммарную ошибку на учебном множестве

из учебного набора удаляется текущий (проверяемый) пример.

$$LOO(h, X) = \sum_{i=1}^{l} \left[a\left(x_i, \{X \setminus x_i\}, h\right) \neq y_i \right] \rightarrow \min_h$$

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

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

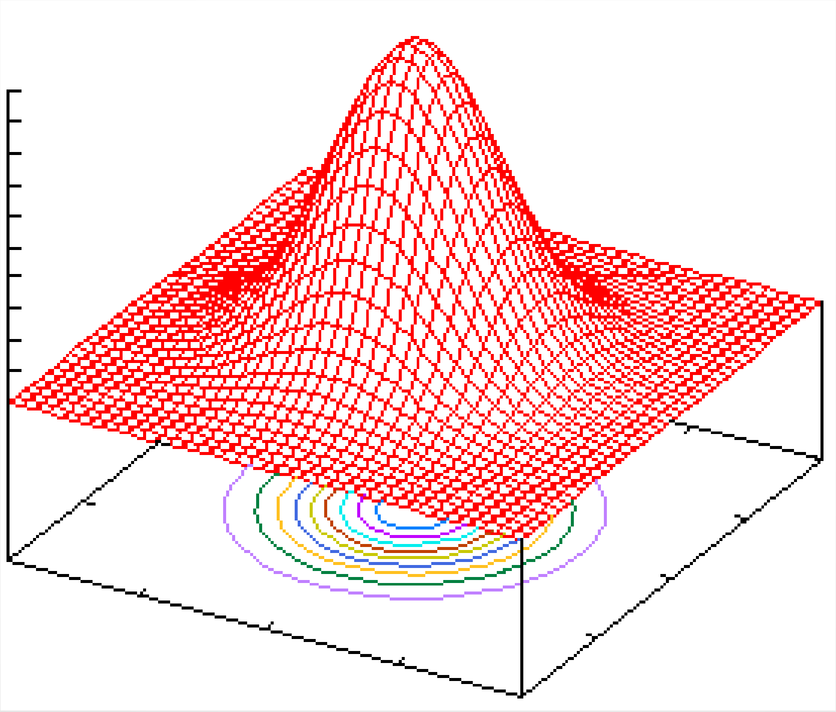
условие оптимума

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

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

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

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

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

Байесовский классификатор

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

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

классификатор: линейный дискриминант Фишера

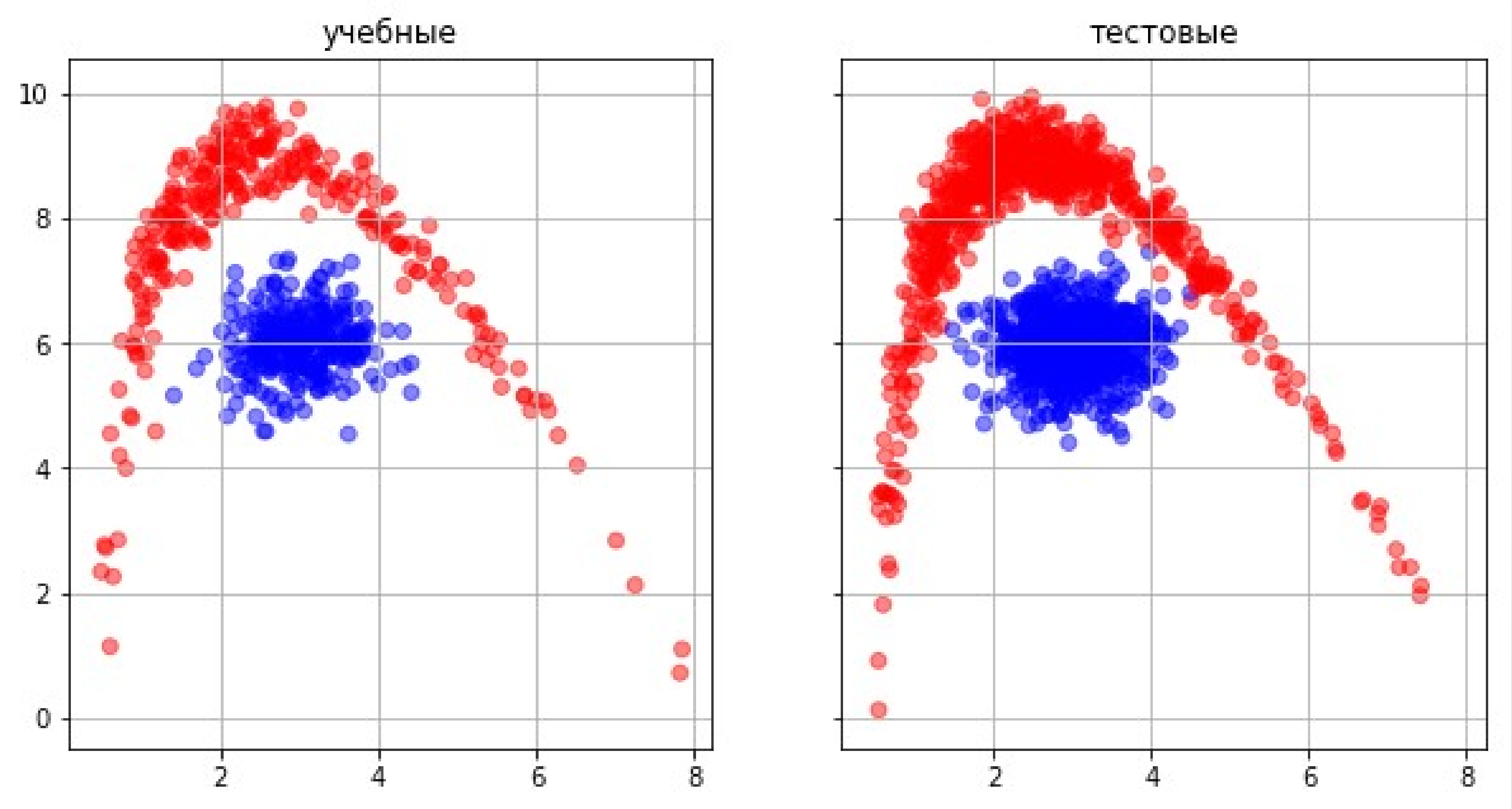
$$a(x) = \underset{y \in Y}{argmax} \left(\ln(\lambda_y P_y) - \frac{1}{2} \hat{\mu}_y^T \hat{\Sigma}^{-1} \hat{\mu}_y + x^T \hat{\Sigma}^{-1} \hat{\mu}_y \right)$$

Классификатор: оценка результата 1

разделяем набор данных
- учебный
- тестовый

недообучение (underfitting)
большая ошибка на учебном наборе

переобучение (overfitting)
малая ошибка на учебном наборе
большая ошибка на тестовом наборе



Классификатор: оценка результата 2

метрики качества на тестовом наборе

- погрешность (accuracy)

- матрица ошибок (confusion matrix)

- точность (precision)

- полнота (recall)

- F-мера

- ROC/AUC

Классификатор: оценка результата 3

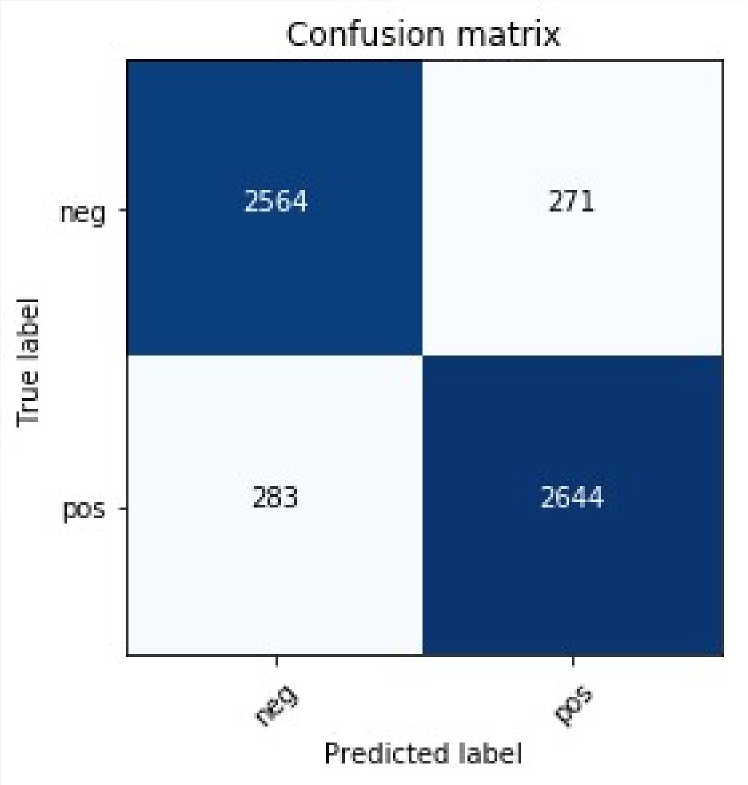
погрешность (accuracy)

правильные ответы / всего примеров

оценка для сбалансированного набора, т.е.
количество примеров в классах +- одинаковое

Классификатор: оценка результата 8

матрица ошибок (confusion matrix)



два класса — четыре группы

- TP истинно положительные
- TN истинно отрицательные
- FP ложно положительные
- FN ложно отрицательные

Классификатор: оценка результата 9

точность (precision)

TP / (TP + FP)

(метрики для отдельного класса)

доля объектов действительно принадлежащих данному классу относительно всех объектов, которые классификатор отнес к этому классу

полнота (recall)

TP / (TP + FN)

доля объектов, найденных классификатором, относительно всех объектов этого класса

F-мера

*(precision*recall) / (precision+recall)*

усреднение точности и полноты

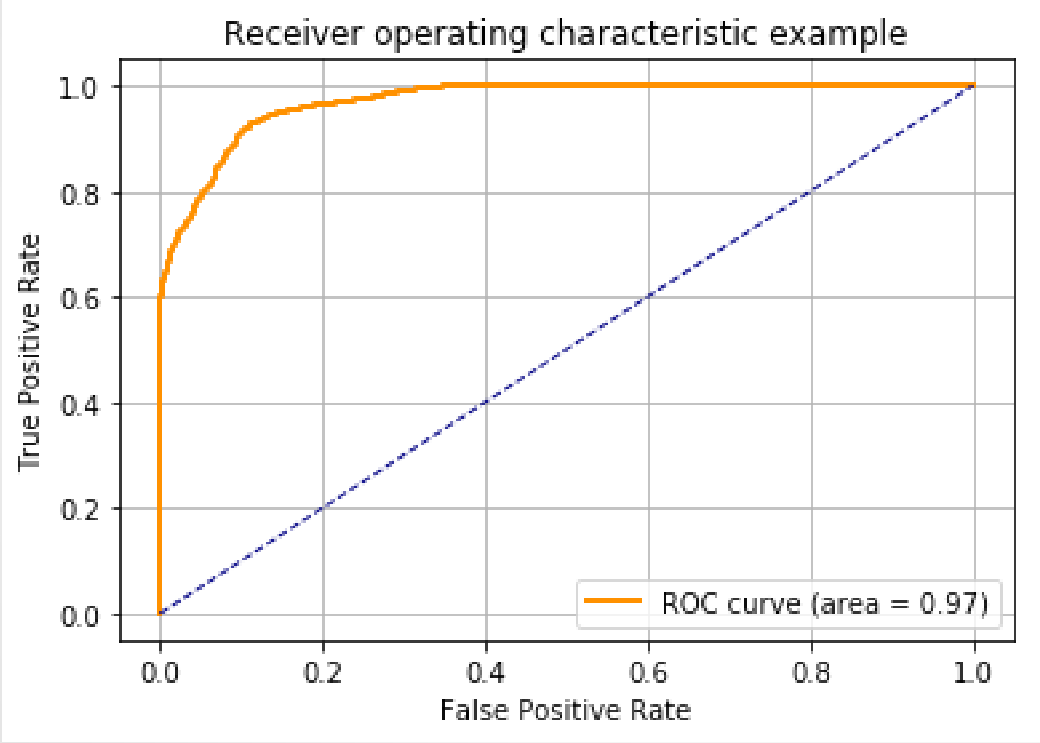
Классификатор: оценка результата 10

Пример *classification_report*

	precision	recall	f1-score	support
0	0.90	0.90	0.90	2835
1	0.91	0.90	0.91	2927
avg / total	0.90	0.90	0.90	5762

Классификатор: оценка результата 11

ROC - receiver operating characteristic,
рабочая характеристика приёмника



TPR=TP/(TP+FN)
полнота(recall), доля объектов, найденных классификатором, относительно всех объектов этого класса

FPR=FP/(FP+TN)
доля объектов negative класса алгоритм предсказал неверно

ROC - показывает зависимость полноты **TPR**
от доли ложно-негативных **FPR**
при изменении порога скора

AUC - area under ROC curve,
площадь под ROC-кривой
характеристика качества классификации

Классификатор: литература

К.В. Воронцов Байесовская теория классификации и методы восстановления плотности. - Курс "Машинное обучение" ШАД Яндекс 2014

Борисов Е.С. Байесовский классификатор.
http://mechanoid.kiev.ua/ml-bayes.html

git clone https://github.com/mechanoid5/ml_lectorium.git

Классификатор: почти последний слайд...

Вопросы ?

Классификатор: практика

источники данных для экспериментов

sklearn.datasets

UCI Repository

kaggle