

Лекция 10: классификатор логистическая регрессия

Евгений Борисов

логистическая регрессия

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} & y_1 \\ x_{21} & x_{22} & \cdots & x_{2n} & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} & y_m \end{bmatrix}$$

x - вектор-признак

y - метка класса

n - размер пространства признаков

m - количество примеров

логистическая регрессия

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1n} \\ 1 & x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

x - вектор-признак

y - метка класса (0 или 1)

n - размер пространства признаков

m - количество примеров

θ - параметры

логистическая регрессия

$$z(x, \theta) = \theta_0 + \theta_1 \cdot x_1 + \dots + \theta_n \cdot x_n$$

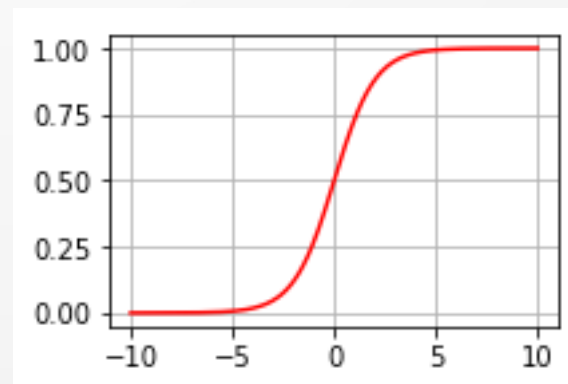
x - вектор-признак

n - размер пространства признаков

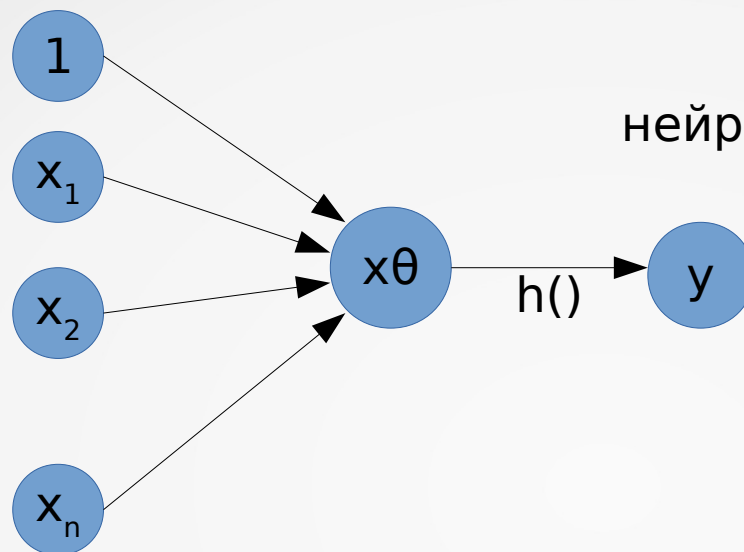
θ - параметры

функция- сигмоид

$$h(z) = \frac{1}{1 + \exp(-z)}$$



логистическая регрессия



нейросеть из одного нейрона

x - вектор-признак

n - размер пространства признаков

θ - параметры

h - функция сигмоид

y - выход

логистическая регрессия

функция потерь - перекрёстная энтропия (cross entropy)

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m \left[y_i \cdot \log_2(h(x_i, \theta)) + (1 - y_i) \cdot \log_2(1 - h(x_i, \theta)) \right]$$

оценка близости распределений правильных и фактических ответов

y - номер класса объекта x

$o = h(x, \theta)$ - ответ классификатора

if (y==1) log(o) else log(1-o)

задача оптимизации

$$\min_{\theta} J(\theta)$$

логистическая регрессия

метод градиентного
спуска

инициализировать θ

вычисляем потерю $J(\theta)$

корректируем
параметры
 $\theta := \theta - \eta \cdot \nabla J(\theta)$

$J < \text{порога}$

конец
работы

значение
градиента
 $\nabla J(\theta)$

логистическая регрессия

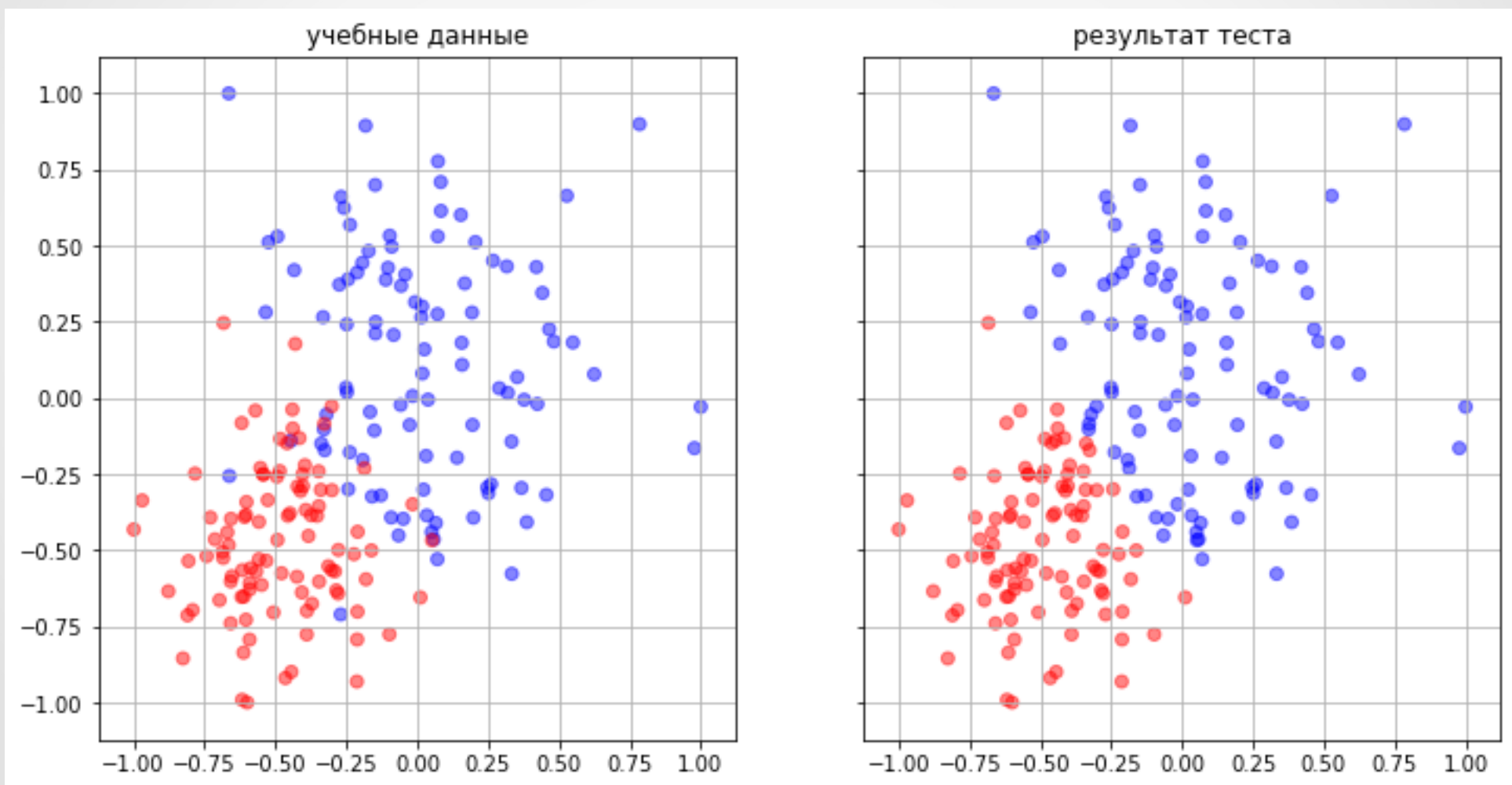
градиент и изменение параметров

$$\theta_j := \theta_j - \alpha \cdot \sum_{i=1}^m (h(x_i, \theta) - y_i) \cdot x_{ij}$$

$$\theta := \theta - \alpha \cdot X^T \cdot (h(x, \theta) - y)$$

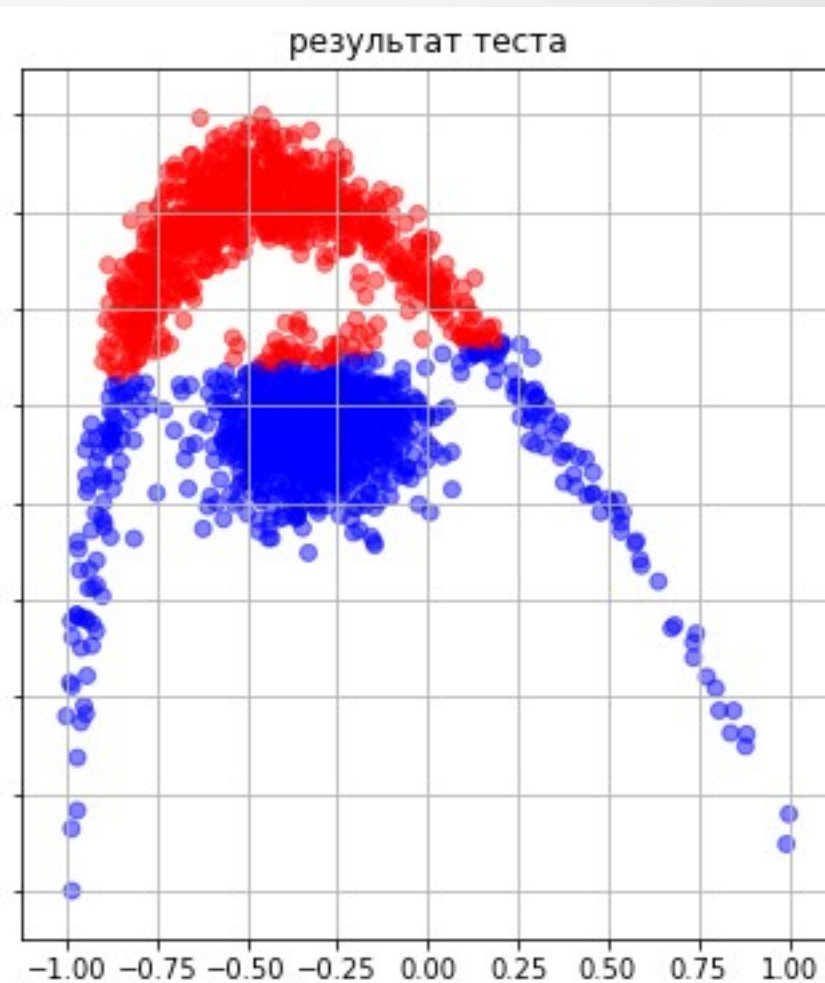
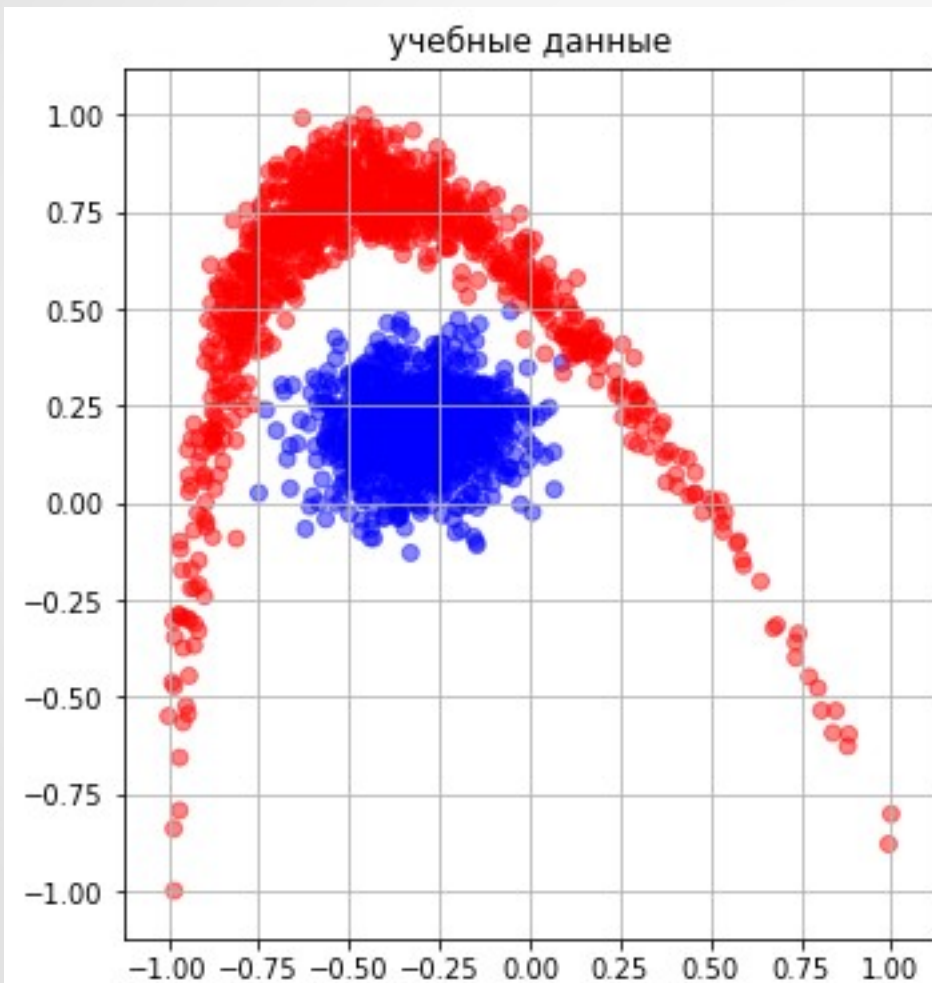
логистическая регрессия

линейный классификатор и линейно разделимые данные



логистическая регрессия

линейный классификатор и линейно **неразделимые** данные



логистическая регрессия

$$z(x, \theta) = \theta_0 + \theta_1 \cdot x_1 + \theta_2 \cdot x_2 + \theta_3 \cdot x_1 \cdot x_2 + \theta_4 \cdot x_1^2 + \theta_5 \cdot x_2^2$$

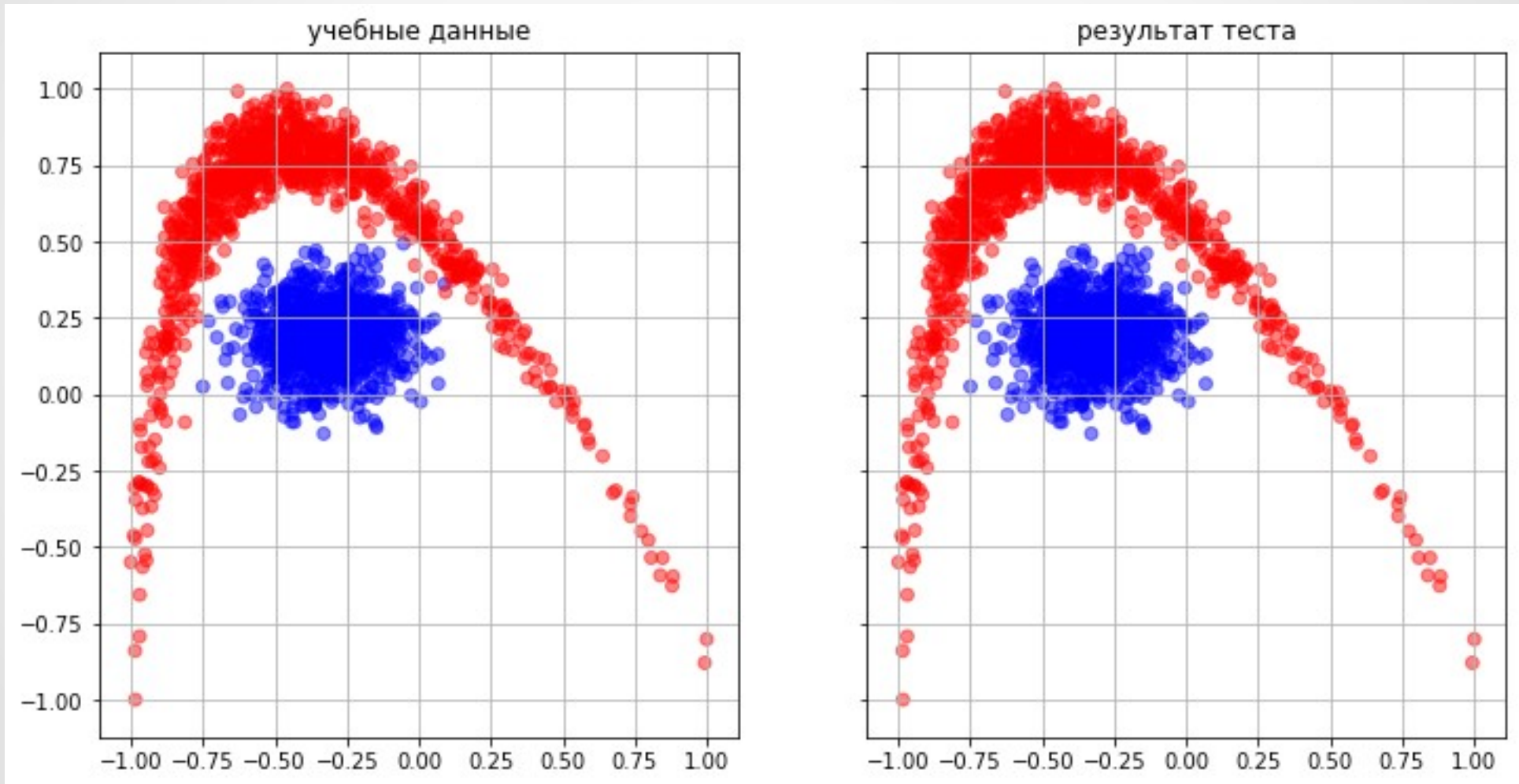
строим полином степени k на признаках X

повышаем размерность пространства признаков

увеличиваем число параметров θ

логистическая регрессия

нелинейный классификатор и линейно **н**еразделимые данные



логистическая регрессия



Вопросы ?

логистическая регрессия

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

*Борисов Е.С. Классификатор на основе логистической регрессии.
<http://mechanoid.kiev.ua/ml-regression-class.html>*

логистическая регрессия

источники данных для экспериментов



sklearn.datasets
UCI Repository
kaggle



задание

выбрать данные в репозитории UCI
(<https://archive.ics.uci.edu/ml>)

реализовать для них логрессию

посчитать оценки качества классификации