



ТЕХНОСФЕРА

Лекция 6 Линейные модели для классификации и регрессии

Николай Анохин

28 октября 2014 г.

План занятия

Линейная регрессия

Логистическая регрессия

Обобщенные линейные модели

Постановка задачи

Пусть дан набор объектов $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}$, $\mathbf{x}_i \in \mathcal{X}$, $y_i \in \mathcal{Y}$, $i \in 1, \dots, N$, полученный из неизвестной закономерности $y = f(\mathbf{x})$. Необходимо выбрать из семейства параметрических функций

$$H = \{h(\mathbf{x}, \theta) : \mathcal{X} \times \Theta \rightarrow \mathcal{Y}\}$$

такую $h^*(\mathbf{x}) = h(\mathbf{x}, \theta^*)$, которая наиболее точно аппроксимирует $f(\mathbf{x})$.

Задачи

- ▶ Регрессия: $\mathcal{Y} = [a, b] \subset \mathbb{R}$
- ▶ Классификация: $|\mathcal{Y}| < C$

Линейная регрессия

Модель

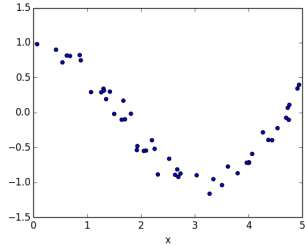
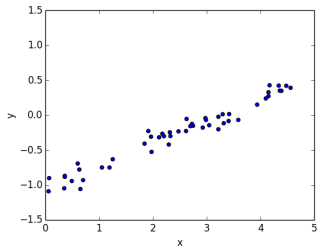
$$y = h(\mathbf{x}, \theta) + \epsilon,$$

где ϵ – гауссовский шум

$$\epsilon = \mathcal{N}(0, \beta^{-1}),$$

откуда

$$p(y|\mathbf{x}, \theta, \beta) = \mathcal{N}(h(\mathbf{x}, \theta), \beta^{-1})$$



Линейная модель

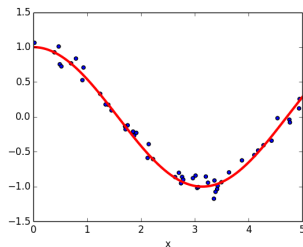
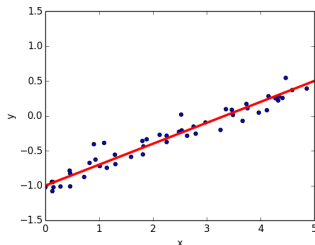
простейшая модель

$$y = w_0 + w_1x_1 + \dots + w_Mx_M = \sum_{j=0}^M w_jx_j$$

улучшенная модель

$$y = \sum_{j=0}^M w_j\phi_j(\mathbf{x}),$$

$\phi_j(\mathbf{x})$ – базисная функция, $\phi_0(\mathbf{x}) = 1$



Логистическая регрессия

Обобщенные линейные модели

Вопросы

