

Московский физико-технический институт
Физтех-школа прикладной математики и информатики

ВВЕДЕНИЕ В АНАЛИЗ ДАННЫХ
IV СЕМЕСТР

Лектор:



Автор: *Киселев Николай*
Репозиторий на Github

весна 2025

Содержание

1	Лекция 2	2
1.1	Линейная Регрессия	2

1 Лекция 2

1.1 Линейная Регрессия

Определение 1.1. $\mathcal{M} = \{y : \mathbb{R}^d \rightarrow \mathbb{R} | y(x) = x^T \theta, \theta \in \mathbb{R}^d\}$ — множество рассматриваемых нами моделей.

Наша цель — получить наилучшую модель, то есть оценить θ .

Пусть $\hat{\theta}$ — оценка θ . Тогда $\hat{y}(x) = x^T \hat{\theta}$ — предсказание для x .

Пусть x_1, \dots, x_n — объекты, Y_1, \dots, Y_n — таргеты. Пусть $\hat{Y}_i = \hat{y}(x_i)$

Введем функционал ошибки:

$$\mathcal{L}(y, z) = (y - z)^2$$

И тогда получаем, что

$$F(\theta) = \sum_{i=1}^n \mathcal{L}(Y_i, \hat{Y}_i) = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n (Y_i - x_i^T \hat{\theta})^2 = \|Y - X\hat{\theta}\|^2$$

Где

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, X = \begin{pmatrix} x_{11} & \dots & x_{1d} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{nd} \end{pmatrix}$$

Мы хотим минимизировать $F(\theta)$

Утверждение 1.1. Если XX^T не вырождена, то $\hat{\theta} =$

Доказательство.

$$F(\theta) = \|Y - X\theta\|^2 = (Y - X\theta)^T (Y - X\theta) = Y^T Y - 2Y^T X\theta + \theta^T X^T X\theta$$

□