Лекция 3. Линейная регрессия

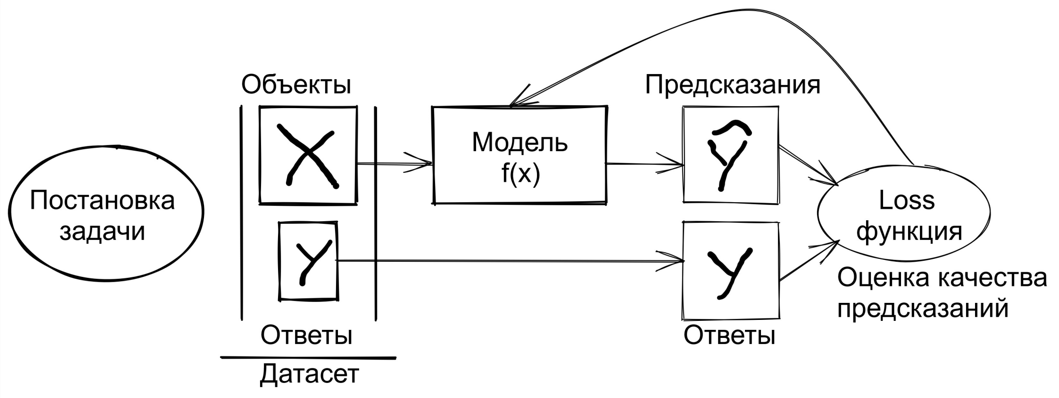
Введение в нейронные сети | 17.10.2023

Линейные модели

- **Линейная регрессия**
- Линейные классификаторы

Сегодня будем разбираться с линейной регрессией.

Как формализовывать задачу?



Формализация задачи

- Постановка задачи
- Данные
- Модель (семейство алгоритмов)
- Оценка качества (функционал ошибки)
- Метод обучения

Постановка задачи

Перенесемся в начало XIX века.

- Вы инженер-физик, у вас на столе лежит кусок неизвестного гибкого материала, который вы хотите использовать в своем устройстве
- Однако, вы не знаете его коэффициента растяжения, и не можете рассчитать возникающие силы
- Вы хотите построить зависимость силы упругости от растяжения для этого материала (и узнать коэффициент растяжения)

Формализация задачи

- **Постановка задачи: регрессия, предсказать F по x**
- Данные
- Модель (семейство алгоритмов)
- Оценка качества (функционал ошибки)
- Метод обучения

Формализация задачи

- Постановка задачи: регрессия, предсказать F по x
- **Данные**
- Модель (семейство алгоритмов)
- Оценка качества (функционал ошибки)
- Метод обучения

Данные

Вы проводите некоторое количество измерений и заносите все в табличку подобного вида:

Растяжение (м)	Сила упругости (Н)
0.182654	0.328200
0.482319	0.870494
0.177778	0.326217
0.742819	1.334474
0.117160	0.212199
...	...
0.074164	0.142423

Формализация задачи

- Постановка задачи: регрессия, предсказать F по x
- **Данные: одномерные (единственный признак x)**
- Модель (семейство алгоритмов)
- Оценка качества (функционал ошибки)
- Метод обучения

Формализация задачи

- Постановка задачи: регрессия, предсказать F по x
- Данные: одномерные (единственный признак x)
- **Модель (семейство алгоритмов)**
- Оценка качества (функционал ошибки)
- Метод обучения

Модель

Благодаря Роберту Гуку вы знаете, что сила упругости зависит от растяжения линейно:

$$F = kx$$

Модель

Благодаря Роберту Гуку вы знаете, что сила упругости зависит от растяжения линейно:

$$F = kx$$

Общий вид линейной функции одной переменной:

$$y(x) = w_1 x + w_0$$

Модель

Благодаря Роберту Гуку вы знаете, что сила упругости зависит от растяжения линейно:

$$F = kx$$

Общий вид линейной функции одной переменной:

$$y(x) = w_1 x + w_0$$

Таким образом, ваша задача среди всех таких функций (семейства) найти ту конкретную, которая лучше всего вам подходит (определить параметры)

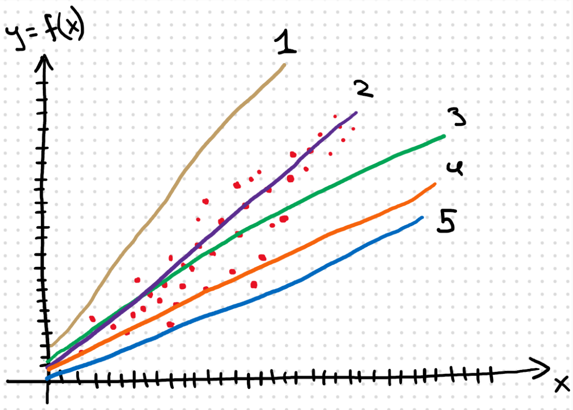
Формализация задачи

- Постановка задачи: регрессия, предсказать F по x
- Данные: одномерные (единственный признак x)
- **Модель: линейная функция одного аргумента**
- Оценка качества (функционал ошибки)
- Метод обучения

Формализация задачи

Что значит “лучше всего вам подходит”?

Формализация задачи



Какая из функций лучше?

Формализация задачи

- Постановка задачи: регрессия, предсказать F по x
- Данные: одномерные (единственный признак x)
- Модель: линейная функция одного аргумента
- **Оценка качества (функционал ошибки)**
- Метод обучения

Функционалы ошибок

Функционал ошибки — функция $\mathcal{L}(y, \hat{y})$, значение которой показывает, насколько хороши предсказания модели относительно правильных ответов.

Функционалы ошибок

Функционал ошибки — функция $\mathcal{L}(y, \hat{y})$, значение которой показывает, насколько хороши предсказания модели относительно правильных ответов.

Наиболее популярные для задачи регрессии:

$$MSE(y, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$MAE(y, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

Ещё есть MSLE, MAPE, SMAPE, и т.д.

Мы в ближайшее время будем использовать MSE, поскольку она имеет хорошие математические свойства.

Формализация задачи

- Постановка задачи: регрессия, предсказать F по x
- Данные: одномерные (единственный признак x)
- Модель: линейная функция одного аргумента
- **Функционал ошибки: MSE**
- Метод обучения

Формализация задачи

- Постановка задачи: регрессия, предсказать F по x
- Данные: одномерные (единственный признак x)
- Модель: линейная функция одного аргумента
- Функционал ошибки: MSE
- **Метод обучения**

Метод обучения

Обучение линейной регрессии: подбор таких параметров w_0, w_1, чтобы результирующая функция давала наименьшее значение MSE, то есть

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \to \min_{w_0, w_1}$$

Метод обучения

Обучение линейной регрессии: подбор таких параметров w_0, w_1, чтобы результирующая функция давала наименьшее значение MSE, то есть

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \to \min_{w_0, w_1}$$

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - w_1 x_i - w_0)^2 \to \min_{w_0, w_1}$$

Как это сделать?

Матанализ: в точке экстремума функции производная обращается в ноль

$$\begin{cases} \frac{d}{dw_0}\frac{1}{n}\sum_{i=1}^{n}(y_i - w_1x_i - w_0)^2 = 0 \\ \frac{d}{dw_1}\frac{1}{n}\sum_{i=1}^{n}(y_i - w_1x_i - w_0)^2 = 0 \end{cases}$$

Метод обучения: (не)много матана

$$\frac{d}{dw_0}\frac{1}{n}\sum_{i=1}^{n}(y_i - w_1x_i - w_0)^2 = \frac{d}{dw_0}\sum_{i=1}^{n}(y_i - w_1x_i - w_0)^2 =$$

Метод обучения: (не)много матана

$$\frac{d}{dw_0}\frac{1}{n}\sum_{i=1}^{n}(y_i - w_1x_i - w_0)^2 = \frac{d}{dw_0}\sum_{i=1}^{n}(y_i - w_1x_i - w_0)^2 =$$

$$= \sum_{i=1}^{n}\frac{d}{dw_0}(y_i - w_1x_i - w_0)^2 = \sum_{i=1}^{n}2(y_i - w_1x_i - w_0)(-1) = 0$$

Метод обучения: (не)много матана

$$\frac{d}{dw_0}\frac{1}{n}\sum_{i=1}^{n}(y_i - w_1x_i - w_0)^2 = \frac{d}{dw_0}\sum_{i=1}^{n}(y_i - w_1x_i - w_0)^2 =$$

$$= \sum_{i=1}^{n}\frac{d}{dw_0}(y_i - w_1x_i - w_0)^2 = \sum_{i=1}^{n}2(y_i - w_1x_i - w_0)(-1) = 0$$

$$\Rightarrow \sum_{i=1}^{n}(y_i - w_1x_i - w_0) = \sum_{i=1}^{n}y_i - w_1\sum_{i=1}^{n}x_i - nw_0 = 0$$

Метод обучения: (не)много матана

$$\frac{d}{dw_0}\frac{1}{n}\sum_{i=1}^{n}(y_i - w_1x_i - w_0)^2 = \frac{d}{dw_0}\sum_{i=1}^{n}(y_i - w_1x_i - w_0)^2 =$$

$$= \sum_{i=1}^{n}\frac{d}{dw_0}(y_i - w_1x_i - w_0)^2 = \sum_{i=1}^{n}2(y_i - w_1x_i - w_0)(-1) = 0$$

$$\Rightarrow \sum_{i=1}^{n}(y_i - w_1x_i - w_0) = \sum_{i=1}^{n}y_i - w_1\sum_{i=1}^{n}x_i - nw_0 = 0$$

$$\Rightarrow nw_0 = \sum_{i=1}^{n}y_i - w_1\sum_{i=1}^{n}x_i$$

Метод обучения: (не)много матана

$$\frac{d}{dw_1}\frac{1}{n}\sum_{i=1}^{n}(y_i - w_1x_i - w_0)^2 =$$

Метод обучения: (не)много матана

$$\frac{d}{dw_1}\frac{1}{n}\sum_{i=1}^{n}(y_i - w_1 x_i - w_0)^2 =$$

$$= \sum_{i=1}^{n}\frac{d}{dw_1}(y_i - w_1 x_i - w_0)^2 = \sum_{i=1}^{n}2(y_i - w_1 x_i - w_0)(-x_i) = 0$$

Метод обучения: (не)много матана

$$\frac{d}{dw_1}\frac{1}{n}\sum_{i=1}^{n}(y_i - w_1x_i - w_0)^2 =$$

$$= \sum_{i=1}^{n}\frac{d}{dw_1}(y_i - w_1x_i - w_0)^2 = \sum_{i=1}^{n}2(y_i - w_1x_i - w_0)(-x_i) = 0$$

$$\Rightarrow \sum_{i=1}^{n}x_i(y_i - w_1x_i - w_0) = 0$$

Метод обучения: (не)много матана

$$\frac{d}{dw_1}\frac{1}{n}\sum_{i=1}^{n}(y_i - w_1x_i - w_0)^2 =$$

$$= \sum_{i=1}^{n}\frac{d}{dw_1}(y_i - w_1x_i - w_0)^2 = \sum_{i=1}^{n}2(y_i - w_1x_i - w_0)(-x_i) = 0$$

$$\Rightarrow \sum_{i=1}^{n}x_i(y_i - w_1x_i - w_0) = 0$$

$$\Rightarrow w_1\sum_{i=1}^{n}x_i^2 = \sum_{i=1}^{n}x_iy_i - w_0\sum_{i=1}^{n}x_i$$

Метод обучения: немного матана

$$\begin{cases} nw_0 = \sum_{i=1}^{n} y_i - w_1 \sum_{i=1}^{n} x_i \\ w_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i - w_0 \sum_{i=1}^{n} x_i \end{cases}$$

Запишите эту систему, она нам еще понадобится на семинаре

Формализация задачи

- Постановка задачи: регрессия, предсказать F по x
- Данные: одномерные (единственный признак x)
- Модель: линейная функция одного аргумента
- Функционал ошибки: MSE
- **Метод обучения: аналитическое решение системы уравнений**

Формализация задачи

- Постановка задачи: регрессия, предсказать F по x
- Данные: одномерные (единственный признак x)
- Модель: линейная функция одного аргумента
- Функционал ошибки: MSE
- Метод обучения: аналитическое решение системы уравнений

Численно рассчитаем этот пример на семинаре, а сейчас разберем многомерную регрессию

Многомерная регрессия

Рассмотрим случай многомерной (множественной) регрессии — когда функция зависит больше чем от одного входного признака.

Каждый пример x теперь вектор в d-мерном пространстве

$$\vec{x} = (x_1, x_2, \ldots, x_d)$$

Как будет выглядеть линейная модель в данном случае?

$$f(\vec{x}) = f(x_1, x_2, \ldots) =$$

Многомерная регрессия

Рассмотрим случай многомерной (множественной) регрессии — когда функция зависит больше чем от одного входного признака.

Каждый пример x теперь вектор в d-мерном пространстве

$$\vec{x} = (x_1, x_2, \ldots, x_d)$$

Как будет выглядеть линейная модель в данном случае?

$$f(\vec{x}) = f(x_1, x_2, \ldots) =$$

$$= w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_d x_d = w_0 + \sum_{j=1}^{d} w_j x_j$$

Bias trick

$$w_0 + \sum_{j=1}^{d} w_j x_j =$$

Bias trick

$$w_0 + \sum_{j=1}^{d} w_j x_j = w_0 + \langle \vec{w}, \vec{x} \rangle =$$

Bias trick

$$w_0 + \sum_{j=1}^{d} w_j x_j = w_0 + \langle \vec{w}, \vec{x} \rangle =$$

$$= \langle \vec{w}, \vec{x} \rangle + w_0 \cdot 1 = \langle \vec{w'}, \vec{x'} \rangle$$

$$\vec{x'} = (1, x_1, x_2, \ldots, x_d) \quad \vec{w'} = (w_0, w_1, \ldots, w_d)$$

Многомерная регрессия

Для всей выборки получаем

$$\vec{y} = X\vec{w}$$

$$\vec{y} = (y_1, y_2, \dots, y_N) \quad \vec{w} = (w_0, w_1, \dots, w_d)$$

$$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1d} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & \cdots & x_{Nd} \end{bmatrix}$$

Аналитическое решение

$$\frac{1}{N}\|X\vec{w} - \vec{y}\|^2 \to \min_{\vec{w}}$$

Аналитическое решение

$$\frac{1}{N}\|X\vec{w} - \vec{y}\|^2 \to \min_{\vec{w}}$$

$$\frac{d}{d\vec{w}}\|X\vec{w} - \vec{y}\|^2 = 0$$

Аналитическое решение

$$\frac{1}{N}\|X\vec{w} - \vec{y}\|^2 \to \min_{\vec{w}}$$

$$\frac{d}{d\vec{w}}\|X\vec{w} - \vec{y}\|^2 = 0$$

$$\vec{w}^* = \arg\min_{\vec{w}}\|X\vec{w} - \vec{y}\|^2 = (X^T X)^{-1} X^T \vec{y}$$