

Линейная регрессия и градиентный спуск. Библиотека `sklearn`



Урок 1

Машинное обучение, его суть и виды



Урок 2

Линейная регрессия
и градиентный спуск.
Библиотека sklearn



Урок 3

Оценка качества и
улучшение модели.
Пайплайн обучения

Что будет на уроке?

1. Теория алгоритма линейной регрессии.
2. Что такое градиентный спуск и зачем он нужен.
3. Напишем линейную регрессию своими руками.
4. Познакомимся с библиотекой `sklearn`.
5. Метрики качества модели и функция потерь.



Какие задачи относятся к регрессии?

1. Определить пол пользователя.
2. Спрогнозировать количество проданных бананов и апельсинов.
3. Основываясь на кредитной истории, предсказать, отдаст ли человек кредит или нет.
4. Основываясь на кредитной истории, предсказать вероятность, отдаст ли человек кредит или нет.

Какие задачи относятся к регрессии?

1. Определить пол пользователя.
2. Спрогнозировать количество проданных бананов и апельсинов.
3. Основываясь на кредитной истории, предсказать, отдаст ли человек кредит или нет.
4. Основываясь на кредитной истории, предсказать вероятность, отдаст ли человек кредит или нет.

Метрики качества модели

Для проверки работы модели и её обучения нужно задать **метрику** и **функцию потерь** (loss function). Суть обучения модели состоит в минимизации (максимизации) такой функции.

Метрики регрессии:

1. Mean squared error (MSE).
2. Mean absolute error (MAE) etc.

Метрики классификации:

1. Precision, recall, accuracy.
2. ROC AUC score.

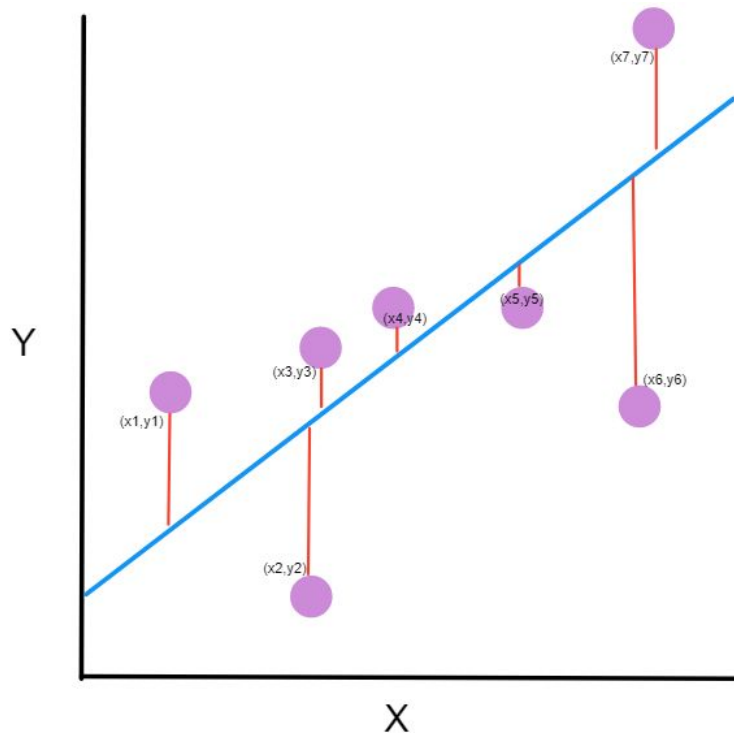
Важно!

**Метрика качества модели —
не всегда функция потерь, в
силу некоторых
математических
ограничений.**

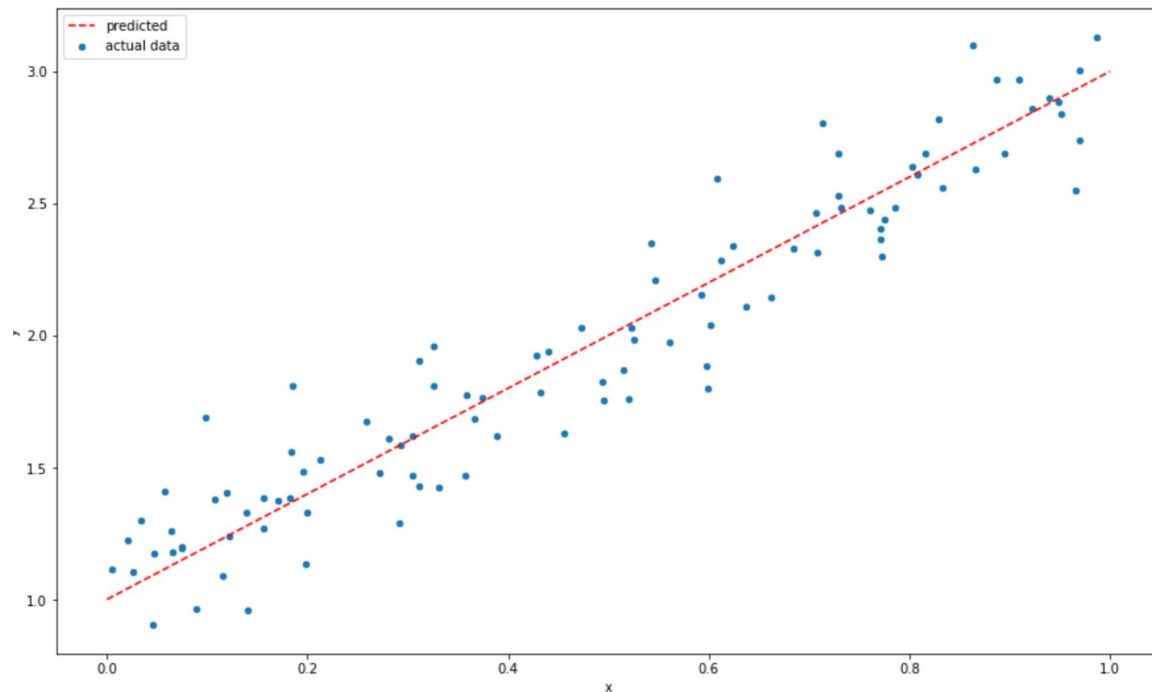
Среднеквадратичная ошибка

Среднеквадратичная ошибка (mean squared error) представляет собой среднее значение суммы **квадратов разности истинного и предсказанного значения**:

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



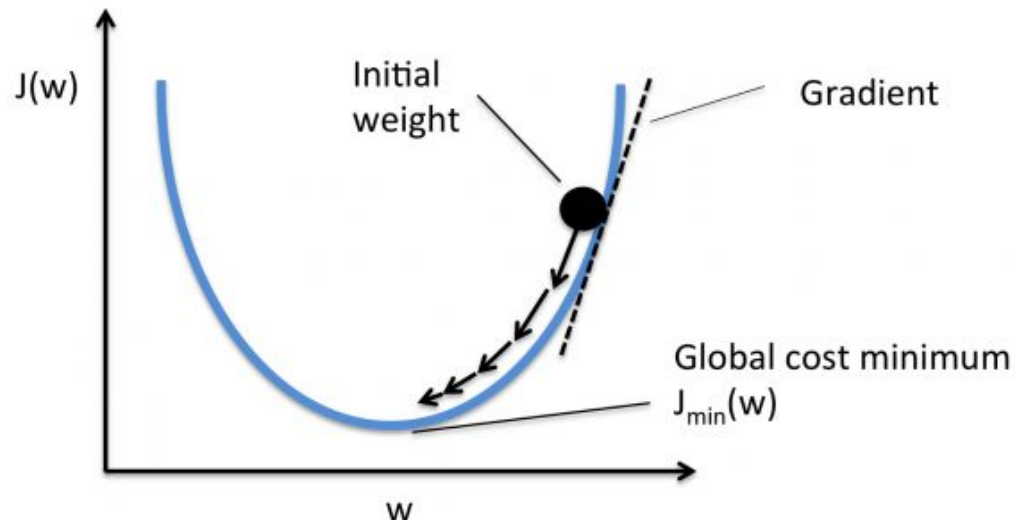
Линейная регрессия



Что такое градиент?

Градиент функции показывает, в каком направлении находится минимум функции.

Градиент векторной функции состоит из производных этой функции по каждому из её аргументов.



Градиентный спуск

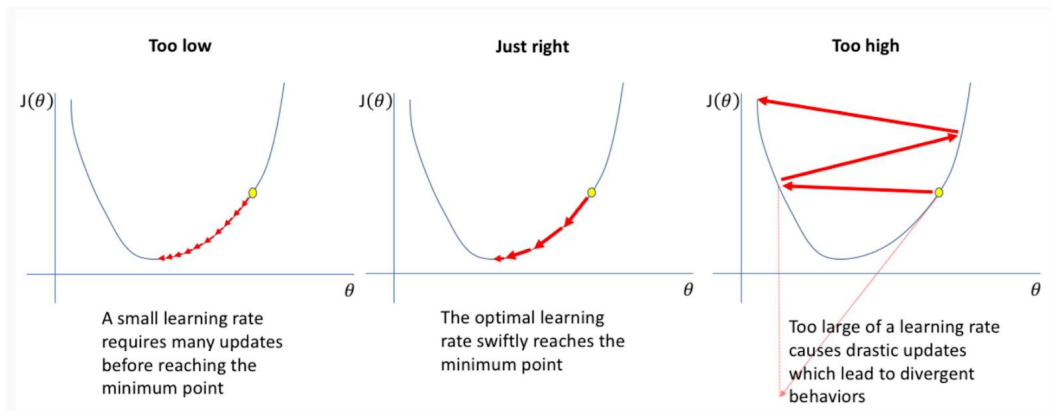
Используя градиентный спуск, можно **быстро** найти один из минимумов функции потерь.

$$w^{(t)} = w^{(t-1)} - \mu \nabla Q(w^{(t-1)})$$

Формула градиентного спуска:

$$Q(w) = \frac{1}{n} \sum_{i=1}^n q_i(w)$$

Ошибка вычисляется по всем объектам выборки.



Проблема выбора шага обучения

Практическое задание

1. Изучите методические материалы к занятию.
2. Пройдите тест с выбором варианта ответа.

