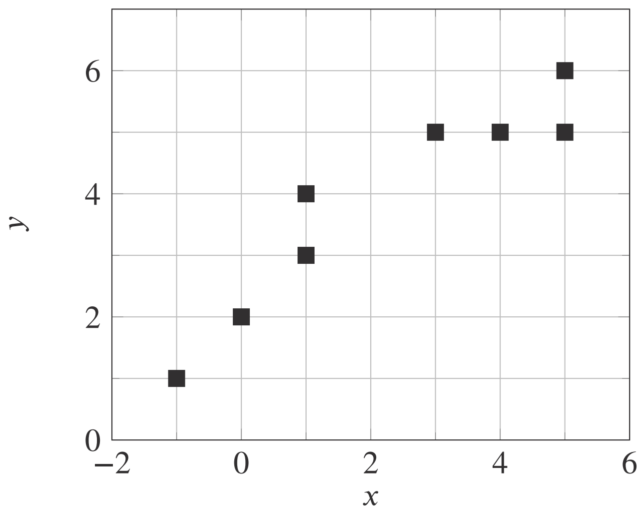
Подробный разбор решения некоторых задач машинного обучения

Задача регрессии. Метод линейной регрессии, оптимизация градиентным спуском

Задача регрессии

Обучающая выборка

x	y
-1	1
1	3
1	4
0	2
3	5
5	6
5	5
4	5



$x \in R$ — независимый признак;
$y \in R$ — целевой признак.

Требуется составить уравнение $\hat{y} = F(x)$, описывающее зависимость данных.

Используем метод линейной регрессии:

$$\hat{y} = b_0 + \sum_{i=1}^{m} b_i x_i$$

m — размерность пространства признаков (количество признаков).

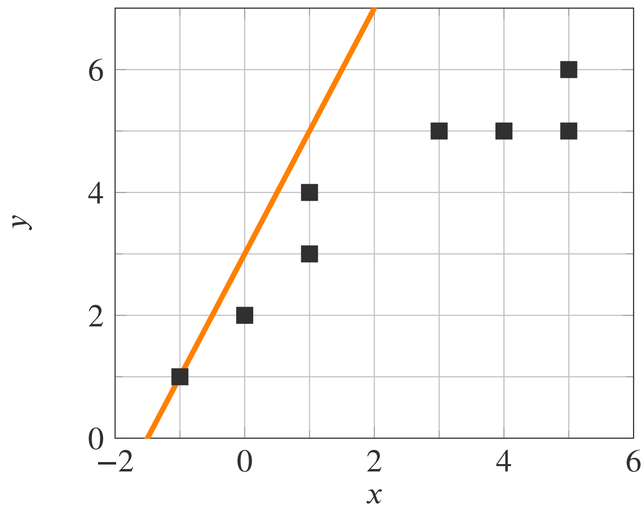
Для случая одной переменной:

$$\hat{y} = b_0 + b_1 x \tag{1}$$

Инициализируем *параметры* модели b_0, b_1 случайными числами:

пока эта зависимость достаточно плохо подходит под данные

$$\hat{y} = 3 + 2x \tag{2}$$



Используем один из показателей (метрик) качества модели регрессии — MSE чтобы оценить существующее уравнение

MSE одновременно является и функцией потерь и метрикой качества для линейной регрессии; см. также R2, MAE, ….

$$L_{MSE} = \sum_{i=1}^{n} \left(\hat{y}_i - y_i\right)^2$$

где n — количество объектов в обучающей выборке.

Подставим уравнение регрессии (1) в функцию ошибок:

$$L_{MSE} = \sum_{i=1}^{n} \left(b_0 + b_1 x - y_i\right)^2 \tag{3}$$

Подставим выбранные значения параметров из (2) и все значения (x, y) обучающей выборки:

$$L_{MSE} = \sum_{i=1}^{n} \left(b_0 + b_1 x_i - y_i\right)^2 = 21.375$$

Оптимальное значение функции — 0.

Задача обучения модели — это задача минимизации функции потерь в пространстве параметров модели, при фиксированном наборе данных. В данном случае, в пространстве b_0, b_1 (рис 1).

$$L_{MSE} = \sum_{i=1}^{n} \left(b_0 + b_1 x - y_i\right)^2 \to \min_{b} \qquad (4)$$

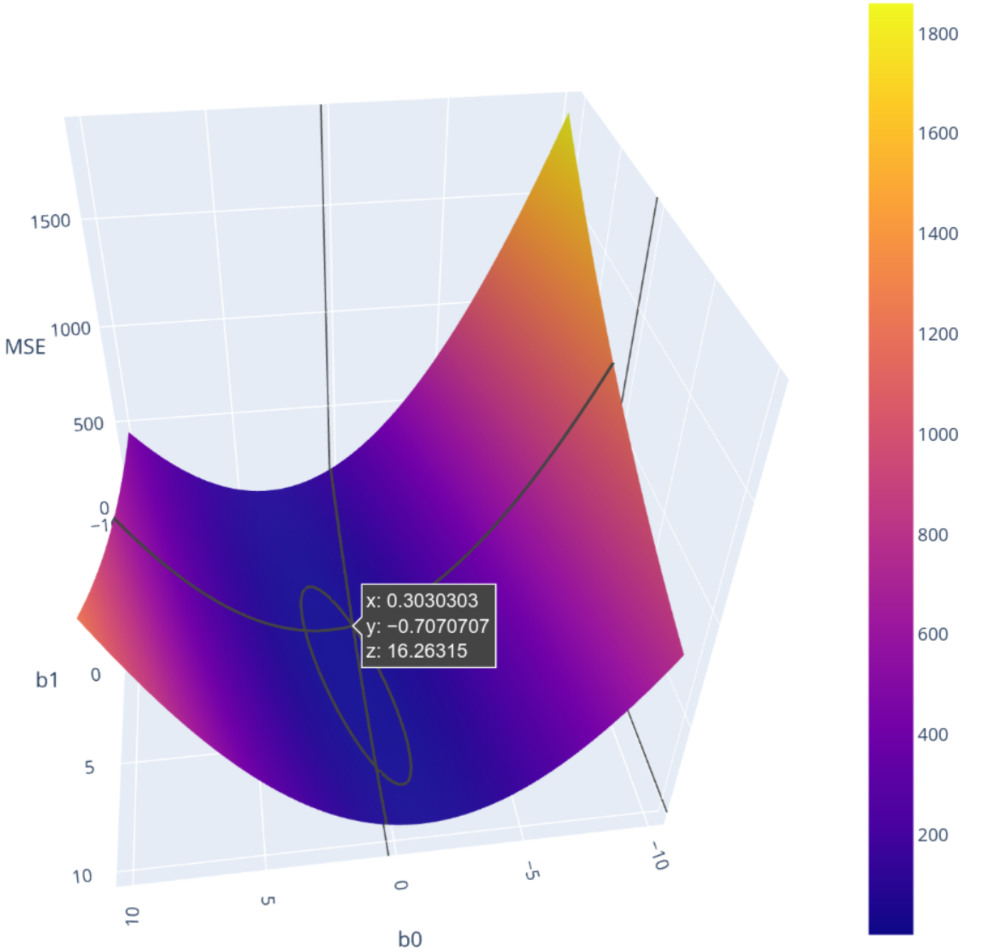


Рис. 1. Поверхность рассматриваемой функции потерь — вытянутый овраг. Для обучения модели требуется найти минимум этой функции

Оптимизация параметров методом градиентного спуска

Далее для простоты рассмотрим только две точки данных.

определить значения параметров уравнения линейной регрессии можно и аналитическим методом, но для примера используем наиболее часто применяемый метод оптимизации в МО — градиентный спуск

x	y
-1	1
4	5

Зафиксируем в формуле (4) все значения x_i и y_i, будем подбирать значения параметров b_0, b_1 методом градиентного спуска.

Выведем формулу для градиента из функции ошибок

$$grad\, L = \left(\frac{\partial L}{\partial b_0}, \frac{\partial L}{\partial b_1}\right)$$

где частные производные:
$\frac{\partial L}{\partial b_0} = \sum_{i=1}^{n} 2(b_0 + b_1 x_i - y_i)$

$\frac{\partial L}{\partial b_1} = \sum_{i=1}^{n} x_i 2(b_0 + b_1 x_i - y_i)$

Зададим коэффициент шага вдоль направления антиградиента $\alpha = 0.01$

Используем начальные значения
$textcolorblue b_0^0 = 3, b_1^0 = 2$

Значение функции ошибок до оптимизации: $L_{MSE}(b_0^0, b_1^0) = 241$

1. Итерация 1

   (a) Вычислим градиент в точке (b_0^0, b_1^0):
   $\frac{\partial L}{\partial b_0} = 2(3 - 2 - 1) + 2(3 + 2 \cdot 4 - 5) = 18$
   $\frac{\partial L}{\partial b_1} = -1 \cdot 2(3 - 2 - 1) + 4 \cdot 2(3 + 2 \cdot 4 - 5) = 80$

   (b) Сделаем шаг вдоль антиградиента, изменив значения параметров
   $b_0^1 = b_0^0 - \frac{\partial L}{\partial b_0}\alpha = 3 - 18 \cdot 0.01 = 2.82$
   $b_1^1 = b_1^1 - \frac{\partial L}{\partial b_1}\alpha = 2 - 80 \cdot 0.01 = 1.2$

   (c) Вычислим значение функции ошибок: $L^0(b_0^1, b_1^1) \approx 7.25$

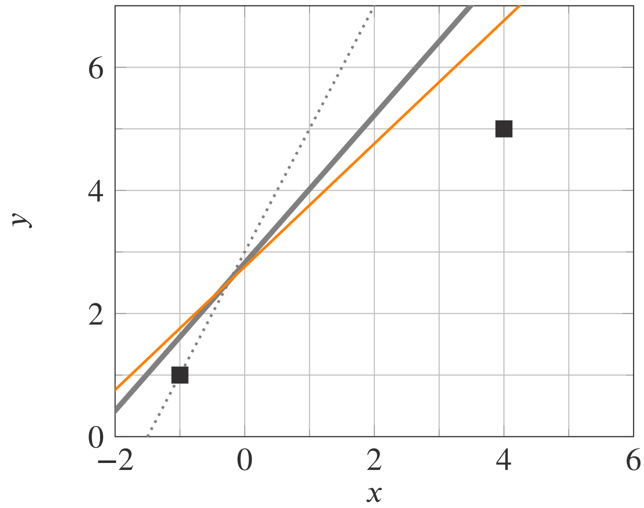
2. Итерация 2

   (a) Вычислим градиент в точке (b_0^0, b_1^0):
   $\frac{\partial L}{\partial b_0} = 2(2.82 + 1.2(-1) - 1) + 2(2.82 + 1.2 \cdot 4 - 5) = 6.48$
   $\frac{\partial L}{\partial b_1} = -1 \cdot 2(2.82 + 1.2(-1) - 1) + 4 \cdot 2(2.82 + 1.2 \cdot 4 - 5) = 19.72$

   (b) Сделаем шаг вдоль антиградиента
   $b_0^1 = b_0^0 - \frac{\partial L}{\partial b_0}\alpha = 2.82 - 6.48 \cdot 0.01 = 2.76$
   $b_1^1 = b_1^1 - \frac{\partial L}{\partial b_1}\alpha = 1.2 - 19.72 \cdot 0.01 = 1$

   (c) Вычислим значение функции ошибок: $L^0(b_0^1, b_1^1) \approx 7.25$

Новая прямая $\hat{y} = 1x + 2.76$ (оранжевая линия) проходит (в среднем) ближе к точкам чем старые (серая).



Повторим шаги 1-3 несколько раз, пока функция ошибок не станет существенно изменяться.