

Аналитическое решение линейной регрессии методом наименьших квадратов

Рассмотрим 5 наблюдений:

$$x = [1, 2, 3, 4, 5]$$

$$y = [2, 4, 5, 4, 5]$$

Шаг 1: Постановка задачи

Ищем параметры линейной модели:

$$y = w \cdot x + b$$

Шаг 2: Матричная форма

Для удобства представим задачу в матричной форме:

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \\ 5 & 1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 2 \\ 4 \\ 5 \\ 4 \\ 5 \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} w \\ b \end{bmatrix}$$

Где первая колонка \mathbf{X} — значения x , вторая колонка — единицы для intercept.

Шаг 3: Аналитическое решение

Параметры оптимальной прямой находятся по формуле:

$$\mathbf{w} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

Шаг 4: Вычисление необходимых матриц

4.1. Вычисляем $\mathbf{X}^T \mathbf{X}$:

$$\mathbf{X}^T = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$
$$\mathbf{X}^T \mathbf{X} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \\ 5 & 1 \end{bmatrix}$$

Первый элемент (1,1):

$$1 \cdot 1 + 2 \cdot 2 + 3 \cdot 3 + 4 \cdot 4 + 5 \cdot 5 = 1 + 4 + 9 + 16 + 25 = 55$$

Первый элемент (1,2):

$$1 \cdot 1 + 2 \cdot 1 + 3 \cdot 1 + 4 \cdot 1 + 5 \cdot 1 = 1 + 2 + 3 + 4 + 5 = 15$$

Второй элемент (2,1) = первому элементу (1,2) = 15

Второй элемент (2,2):

$$1 \cdot 1 + 1 \cdot 1 + 1 \cdot 1 + 1 \cdot 1 + 1 \cdot 1 = 5$$

Итого:

$$\mathbf{X}^T \mathbf{X} = \begin{bmatrix} 55 & 15 \\ 15 & 5 \end{bmatrix}$$

Шаг 5: Вычисление обратной матрицы

5.1. Определитель матрицы:

$$\det(\mathbf{X}^T \mathbf{X}) = 55 \cdot 5 - 15 \cdot 15 = 275 - 225 = 50$$

5.2. Обратная матрица:

$$(\mathbf{X}^T \mathbf{X})^{-1} = \frac{1}{50} \begin{bmatrix} 5 & -15 \\ -15 & 55 \end{bmatrix} = \begin{bmatrix} 0.1 & -0.3 \\ -0.3 & 1.1 \end{bmatrix}$$

Шаг 6: Вычисление $\mathbf{X}^T \mathbf{y}$

$$\mathbf{X}^T \mathbf{y} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \\ 5 \\ 4 \\ 5 \end{bmatrix}$$

Первый элемент:

$$1 \cdot 2 + 2 \cdot 4 + 3 \cdot 5 + 4 \cdot 4 + 5 \cdot 5 = 2 + 8 + 15 + 16 + 25 = 66$$

Второй элемент:

$$1 \cdot 2 + 1 \cdot 4 + 1 \cdot 5 + 1 \cdot 4 + 1 \cdot 5 = 2 + 4 + 5 + 4 + 5 = 20$$

Итого:

$$\mathbf{X}^T \mathbf{y} = \begin{bmatrix} 66 \\ 20 \end{bmatrix}$$

Шаг 7: Вычисление весов

$$\mathbf{w} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} = \begin{bmatrix} 0.1 & -0.3 \\ -0.3 & 1.1 \end{bmatrix} \begin{bmatrix} 66 \\ 20 \end{bmatrix}$$

7.1. Вычисление w:

$$\mathbf{w} = 0.1 \cdot 66 + (-0.3) \cdot 20 = 6.6 - 6.0 = 0.6$$

7.2. Вычисление b:

$$\mathbf{b} = (-0.3) \cdot 66 + 1.1 \cdot 20 = -19.8 + 22.0 = 2.2$$

Шаг 8: Результат

$$\boxed{y = 0.6x + 2.2}$$

Шаг 9: Проверка предсказаний

Вычислим предсказания для всех x:

$$\hat{y}_1 = 0.6 \cdot 1 + 2.2 = 2.8$$

$$\hat{y}_2 = 0.6 \cdot 2 + 2.2 = 3.4$$

$$\hat{y}_3 = 0.6 \cdot 3 + 2.2 = 4.0$$

$$\hat{y}_4 = 0.6 \cdot 4 + 2.2 = 4.6$$

$$\hat{y}_5 = 0.6 \cdot 5 + 2.2 = 5.2$$

Шаг 10: Вычисление ошибок

Ошибки:

$$e_1 = 2 - 2.8 = -0.8$$

$$e_2 = 4 - 3.4 = 0.6$$

$$e_3 = 5 - 4.0 = 1.0$$

$$e_4 = 4 - 4.6 = -0.6$$

$$e_5 = 5 - 5.2 = -0.2$$

Шаг 11: Вычисление MSE

$$\begin{aligned} \text{MSE} &= \frac{1}{5} [(-0.8)^2 + (0.6)^2 + (1.0)^2 + (-0.6)^2 + (-0.2)^2] \\ &= \frac{1}{5} [0.64 + 0.36 + 1.00 + 0.36 + 0.04] \\ &= \frac{1}{5} \cdot 2.40 = 0.48 \end{aligned}$$

Численный пример градиентного спуска с $\eta = 0.01$

Исходные данные

$$X = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{bmatrix}, \quad y = \begin{bmatrix} 2 \\ 4 \\ 5 \\ 4 \\ 5 \end{bmatrix}$$

Ищем параметры: $y = w \cdot x + b$

Начальные веса: $w_0 = 0, b_0 = 0$

Learning rate: $\eta = 0.01$

Количество примеров: $n = 5$

Итерация 1

1. Вычисление предсказаний:

$$\begin{aligned}\hat{y}_1 &= 0 \cdot 1 + 0 = 0 \\ \hat{y}_2 &= 0 \cdot 2 + 0 = 0 \\ \hat{y}_3 &= 0 \cdot 3 + 0 = 0 \\ \hat{y}_4 &= 0 \cdot 4 + 0 = 0 \\ \hat{y}_5 &= 0 \cdot 5 + 0 = 0\end{aligned}$$

2. Вычисление градиентов:

Градиент по w :

$$\begin{aligned}\frac{\partial L}{\partial w} &= \frac{2}{n} \sum_{i=1}^n x_i(\hat{y}_i - y_i) \\ &= \frac{2}{5} [1(0 - 2) + 2(0 - 4) + 3(0 - 5) + 4(0 - 4) + 5(0 - 5)] \\ &= \frac{2}{5} [(-2) + (-8) + (-15) + (-16) + (-25)] \\ &= \frac{2}{5} \cdot (-66) = -26.4\end{aligned}$$

Градиент по b :

$$\begin{aligned}\frac{\partial L}{\partial b} &= \frac{2}{n} \sum_{i=1}^n (\hat{y}_i - y_i) \\ &= \frac{2}{5} [(0 - 2) + (0 - 4) + (0 - 5) + (0 - 4) + (0 - 5)] \\ &= \frac{2}{5} \cdot (-20) = -8\end{aligned}$$

3. Обновление параметров:

$$\begin{aligned}w_1 &= w_0 - \eta \cdot \frac{\partial L}{\partial w} = 0 - 0.01 \cdot (-26.4) = 0.264 \\ b_1 &= b_0 - \eta \cdot \frac{\partial L}{\partial b} = 0 - 0.01 \cdot (-8) = 0.08\end{aligned}$$

Итерация 2

1. Вычисление предсказаний:

$$\begin{aligned}\hat{y}_1 &= 0.264 \cdot 1 + 0.08 = 0.344 \\ \hat{y}_2 &= 0.264 \cdot 2 + 0.08 = 0.608 \\ \hat{y}_3 &= 0.264 \cdot 3 + 0.08 = 0.872 \\ \hat{y}_4 &= 0.264 \cdot 4 + 0.08 = 1.136 \\ \hat{y}_5 &= 0.264 \cdot 5 + 0.08 = 1.400\end{aligned}$$

2. Вычисление градиентов:

Градиент по w :

$$\begin{aligned}\frac{\partial L}{\partial w} &= \frac{2}{5} [1(0.344 - 2) + 2(0.608 - 4) + 3(0.872 - 5) + 4(1.136 - 4) + 5(1.400 - 5)] \\ &= \frac{2}{5} [(-1.656) + (-6.784) + (-12.384) + (-11.456) + (-18.0)] \\ &= \frac{2}{5} \cdot (-50.28) = -20.112\end{aligned}$$

Градиент по b :

$$\begin{aligned}\frac{\partial L}{\partial b} &= \frac{2}{5} [(0.344 - 2) + (0.608 - 4) + (0.872 - 5) + (1.136 - 4) + (1.400 - 5)] \\ &= \frac{2}{5} [(-1.656) + (-3.392) + (-4.128) + (-2.864) + (-3.6)] \\ &= \frac{2}{5} \cdot (-15.64) = -6.256\end{aligned}$$

3. Обновление параметров:

$$\begin{aligned}w_2 &= w_1 - \eta \cdot \frac{\partial L}{\partial w} = 0.264 - 0.01 \cdot (-20.112) = 0.46512 \\ b_2 &= b_1 - \eta \cdot \frac{\partial L}{\partial b} = 0.08 - 0.01 \cdot (-6.256) = 0.14256\end{aligned}$$

Итерация 3

1. Вычисление предсказаний:

$$\begin{aligned}\hat{y}_1 &= 0.46512 \cdot 1 + 0.14256 = 0.60768 \\ \hat{y}_2 &= 0.46512 \cdot 2 + 0.14256 = 1.07280 \\ \hat{y}_3 &= 0.46512 \cdot 3 + 0.14256 = 1.53792 \\ \hat{y}_4 &= 0.46512 \cdot 4 + 0.14256 = 2.00304 \\ \hat{y}_5 &= 0.46512 \cdot 5 + 0.14256 = 2.46816\end{aligned}$$

2. Вычисление градиентов:

Градиент по w :

$$\begin{aligned}\frac{\partial L}{\partial w} &= \frac{2}{5} [1(0.60768 - 2) + 2(1.07280 - 4) + 3(1.53792 - 5) + 4(2.00304 - 4) + 5(2.46816 - 5)] \\ &= \frac{2}{5} [(-1.39232) + (-5.8544) + (-10.38624) + (-7.98784) + (-12.6592)] \\ &= \frac{2}{5} \cdot (-38.28) = -15.312\end{aligned}$$

Градиент по b :

$$\begin{aligned}\frac{\partial L}{\partial b} &= \frac{2}{5} [(0.60768 - 2) + (1.07280 - 4) + (1.53792 - 5) + (2.00304 - 4) + (2.46816 - 5)] \\ &= \frac{2}{5} [(-1.39232) + (-2.9272) + (-3.46208) + (-1.99696) + (-2.53184)] \\ &= \frac{2}{5} \cdot (-12.3104) = -4.92416\end{aligned}$$

3. Обновление параметров:

$$\begin{aligned}w_3 &= w_2 - \eta \cdot \frac{\partial L}{\partial w} = 0.46512 - 0.01 \cdot (-15.312) = 0.61824 \\ b_3 &= b_2 - \eta \cdot \frac{\partial L}{\partial b} = 0.14256 - 0.01 \cdot (-4.92416) = 0.1918016\end{aligned}$$

Результаты первых 3 итераций

Итерация	w	b	$\frac{\partial L}{\partial w}$	$\frac{\partial L}{\partial b}$	MSE
0	0.000000	0.000000	—	—	17.200000
1	0.264000	0.080000	-26.400000	-8.000000	10.131904
2	0.465120	0.142560	-20.112000	-6.256000	6.469662
3	0.618240	0.191802	-15.312000	-4.924160	4.258099