Линейная регрессия

Метод наименьших квадратов как решение матричного уравнения. Часть 1



$$X = \begin{pmatrix} x_{00} & x_{01} & \dots & x_{0k} \\ x_{10} & x_{11} & \dots & x_{1k} \\ \dots & \dots & \dots & \dots \\ x_{n0} & x_{n1} & \dots & x_{nk} \end{pmatrix}$$

$$\vec{y} = \begin{pmatrix} y_0 \\ y_1 \\ \dots \\ y_n \end{pmatrix}$$

$$\hat{y} = Xw$$

Loss =
$$(Xw - y)^2 \rightarrow min$$

$$Xw = \hat{y}$$



 $X^T X w = X^T \hat{y},$ где X^T — матрица, транспонированная к матрице X



 $w = (X^T X)^{-1} X^T \hat{y},$ где $(X^T X)^{-1}$ — матрица, обратная к $X^T X$

Пример

Пусть у нас имеется выборка из 5 объектов, описанных всего одним признаком. То есть $n=5,\,k=1$

X_1	
1	
3	
4	
5	
7	

Y	
6	
16	
21	
26	
36	

Перед нами задача одномерной регрессии:

$$\hat{y} = w_0 + w_1 \times x_1$$

Наша задача — определить значения вектора весов: w₀ и w₁

Очевидно, что искомая функция:

$$\hat{y} = 1 + 5 \times x_1$$

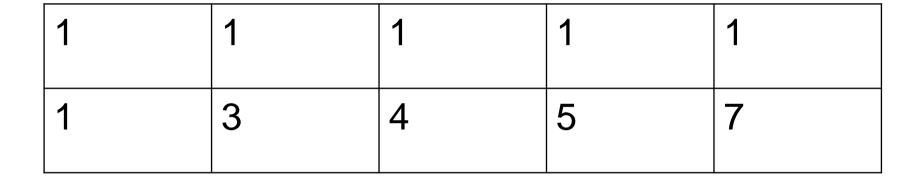
X_1	
1	
3	
4	
5	
7	

X_0	X ₁
1	1
1	3
1	4
1	5
1	7

Это и есть матрица Х

X ₀	X ₁
1	1
1	3
1	4
1	5
1	7

Тогда транспонированная матрица X^T



Тогда матрица Х^ТХ

5	20
20	100

Матрица, обратная к X^TX

1	-0,2
-0,2	0,05

Значение X^T ŷ

105	
520	

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

 1

 5

Таким образом, уравнение регрессии действительно имеет вид: $\hat{y} = 1 \times 1 + 5 \times x = 1 + 5x$