Linear regression

Ordinary least square regression

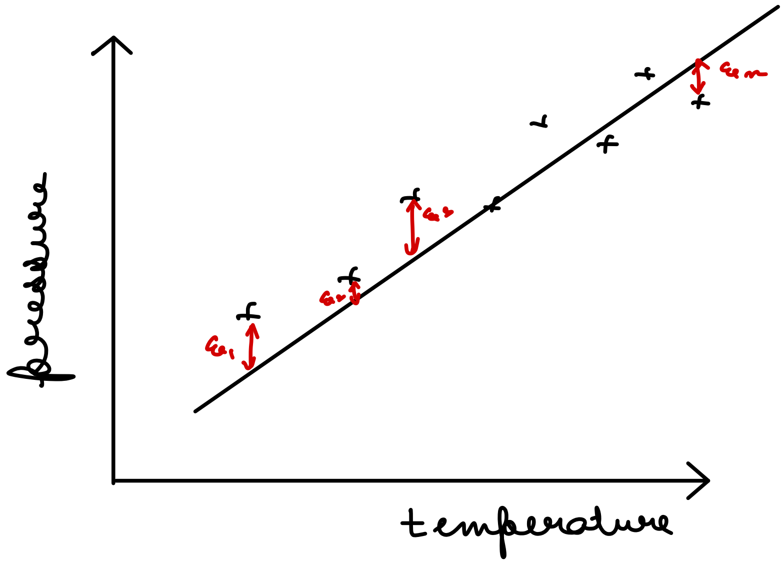
Case of one - variable

1. As we learnt in last video, one of the tasks in regression problem is to determine the parameters of the model.

2. We want to fit the generalized linear model, we can write it as:

$$y_i = \alpha x_i + \beta + \varepsilon_i$$

$$\varepsilon_i = y_i - \alpha x_i - \beta$$



For n samples:

$$\varepsilon = \sum_{i=1}^{n} \varepsilon_i = \sum_{i=1}^{n} (y_i - \alpha x_i - \beta)$$

$$J = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \alpha x_i - \beta)^2$$

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \ \varepsilon^2 \quad \text{——(a)}$$

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \ \varepsilon^2 \quad \text{——(b)}$$

$$\frac{\partial J}{\partial \alpha} = 0 \ ; \quad \frac{\partial J}{\partial \beta} = 0$$

$$J = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \alpha x_i - \beta)^2$$

$$\frac{\partial J}{\partial \alpha} = \sum_{i=1}^{n} (2)(y_i - \alpha x_i - \beta)(-x_i)$$

$$0 = \sum_{i=1}^{n} (y_i - \alpha x_i - \beta)(x_i)$$

$$0 = \sum_{i=1}^{n} (x_i y_i - \alpha x_i^2 - \beta x_i)$$

$$-\sum_{i=1}^{n} \beta x_i + \sum_{i=1}^{n} x_i y_i = \alpha \sum_{i=1}^{n} x_i^2$$

$$\frac{\partial J}{\partial \beta} = 0$$

$$\Rightarrow \sum_{i=1}^{n} 2(y_i - \alpha x_i - \beta)(-1) = 0$$

$$\Rightarrow \frac{\sum_{i=1}^{n} y_i}{n} - \frac{\alpha \sum_{i=1}^{n} x_i}{n} = \beta$$

$$\Rightarrow \boxed{\bar{y} - \alpha \bar{x} = \beta}$$

$$\Rightarrow -\sum_{i=1}^{n} (\bar{y} - \alpha \bar{x})(x_i) + \alpha \sum_{i=1}^{n} x_i^2 = +\sum_{i=1}^{n} x_i y_i$$

$$\sum_{i=1}^{n} -x_i \bar{y} + \alpha\left(\sum_{i=1}^{n} x_i \bar{x} + \sum_{i=1}^{n} x_i^2\right) = \sum_{i=1}^{n} x_i y_i$$

$$\alpha = \frac{\sum_{i=1}^{n} x_i y_i - N \bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - N \bar{x}^2}$$

$$\alpha = \frac{E(XY) - E(X)E(Y)}{E(X^2) - [E(X)]^2}$$

$$= \mathrm{cov}(X, Y) / \mathrm{var}(X)$$

Case with multiple variables

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \underset{n \times k}{\begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ \vdots \\ \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix}} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \underset{n \times 1}{\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}}$$

$$\vec{Y} = [X] \vec{\beta} + \vec{e}$$

$$\vec{e} = \vec{Y} - [X] \vec{\beta}$$

$$\vec{e}^T \vec{e} = [e_1 \ldots e_n] \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}$$

$$= e_1^2 + e_2^2 + \ldots + e_n^2$$

$$\Rightarrow \quad \vec{e}^T \vec{e} = (Y - X\beta)^T (Y - X\beta)$$

$$= (Y^T - \beta^T X^T)(Y - X\beta)$$

$$= Y^T Y - \beta^T X^T Y - \boxed{Y^T X \beta} + \beta^T X^T X \beta$$

scalar

$$\vec{e}^T \vec{e} = Y^T Y - 2\beta' X^T Y + \beta^T X^T X \beta$$

$$\frac{\partial \vec{e}^T \vec{e}}{\partial \beta} = 0$$

$$-2X^T Y + 2X^T X \beta = 0$$

$$X^T Y = X^T X \beta$$

$$(X^T X)^{-1} X^T Y = (X^T X)^{-1} (X^T X) \beta$$

$$\Rightarrow \beta = (X^T X)^{-1} X^T Y$$