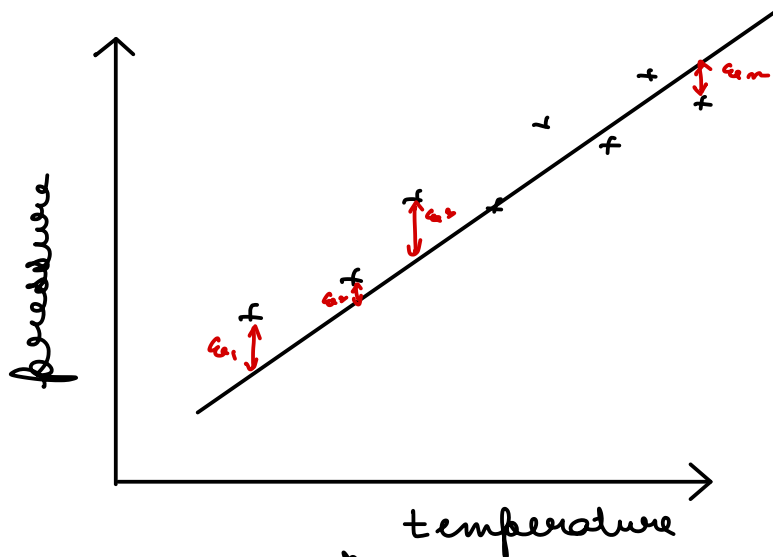


Linear Regression

Gradient Descent approach

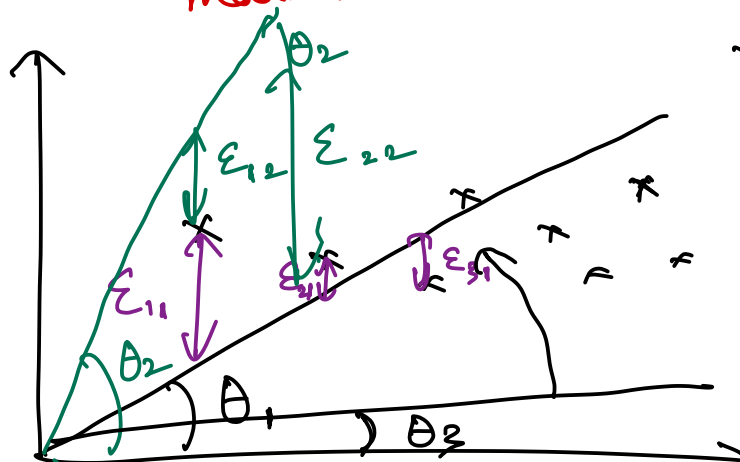


$$\sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - \alpha x_i - \beta)^2$$

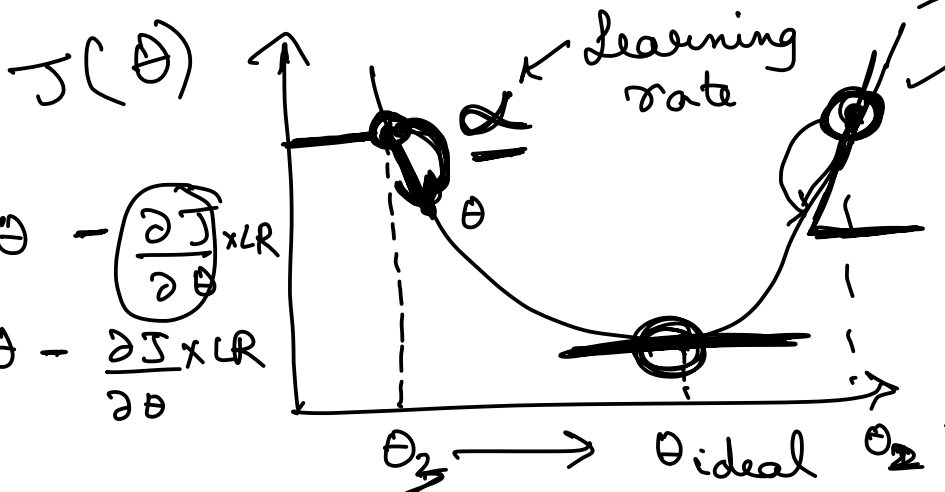
$$J(\alpha, \beta) = \frac{\epsilon_i^2}{n}$$

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

↑
mathematical convenience



$$y = \theta x$$



$$\theta := \theta - \left(\frac{\partial J}{\partial \theta} \right) \times LR$$

$$\theta := \theta - \frac{\partial J}{\partial \theta} \times LR$$

$$\propto \frac{\partial J}{\partial \theta}$$

$$\vec{y} = \theta_1 \vec{x} + \theta_2$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

predictand

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

predictor

$$\theta_1 = ?$$

$$\theta_2 = ?$$

$$y = \theta_1 x + \theta_2$$

$$y = f(x)$$

$$J = \frac{1}{2n} \sum_{i=1}^n (y_i - \theta_1 x_i - \theta_2)^2$$

$$\frac{\partial J}{\partial \theta_1} = \frac{1}{n} \sum_{i=1}^n ((y_i - \theta_1 x_i - \theta_2)(-x_i))$$

$$\frac{\partial J}{\partial \theta_1} = \frac{1}{n} \sum_{i=1}^n ((f(x_i) - y_i)(x_i))$$

$$\frac{\partial J}{\partial \theta_2} = \frac{1}{n} \sum_{i=1}^n (f(x_i) - y_i)$$



$$\theta_1 := \theta_1 - \frac{\partial J}{\partial \theta_1} (\text{learning rate})$$

$$\theta_2 := \theta_2 - \alpha \frac{\partial J}{\partial \theta_2}$$

$$\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} := \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} - \alpha \begin{bmatrix} \frac{\partial J}{\partial \theta_1} \\ \frac{\partial J}{\partial \theta_2} \end{bmatrix}$$

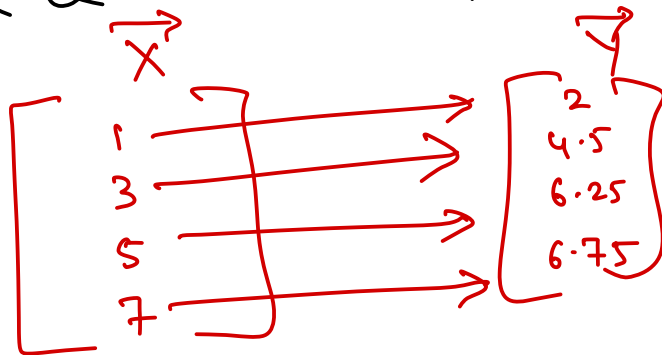
value of θ_1 stop changing
or

$$(\theta_{\text{new}} - \theta_{\text{prev}}) \leq \text{Tolerable limits}$$

3 Questions

- ① Gradient descent to multivariate regression with 2 parameters?
- ② What happens if α is too large or α is too small?

③



$$Y = \theta_1 X + \theta_2$$

$$\theta_1 = 0.1$$

$$\theta_2 = 1$$

$$\alpha = 0.01$$

$$\theta_1 :=$$

$$\theta_2 :=$$