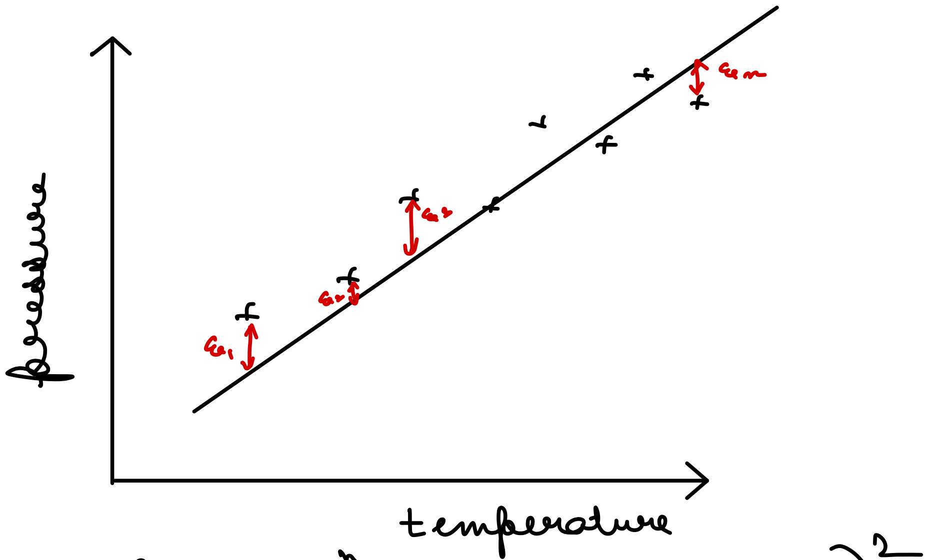
Linear Regression

Gradient Descent approach

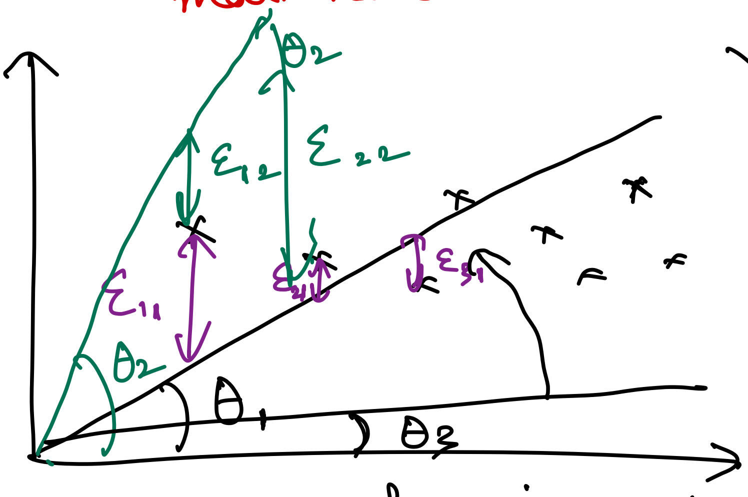


$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \alpha x_i - \beta)^2$$

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

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

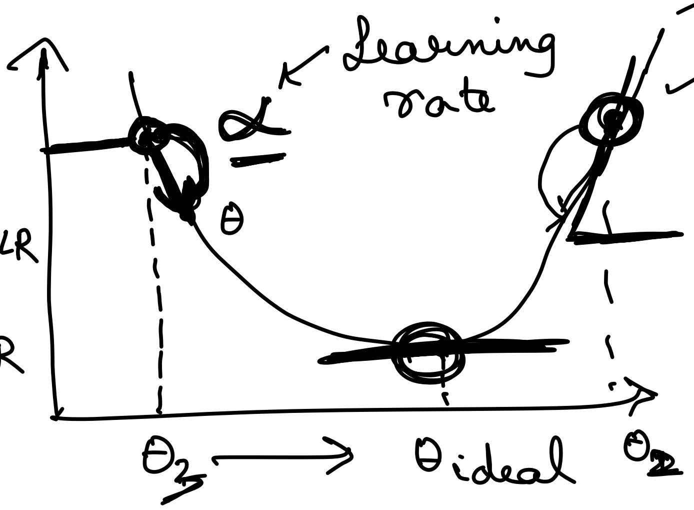
mathematical convenience



$Y = \theta x$

θ

$J(\theta)$



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

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

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

$$\vec{Y} = \theta_1 \vec{X} + \theta_2$$

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{\text{predictand}} \qquad \underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}}_{\text{predictor}}$$

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

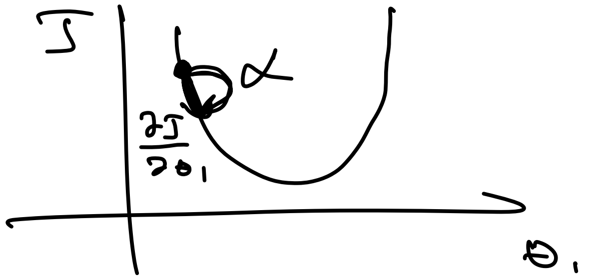
$$\theta_2 = ? \qquad y = f(x)$$

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

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

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

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



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

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

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

value of θs stop changing
or

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

3 Questions

1. Gradient descent to multivariate regression with k parameters?
2. What happens if α is too large or α is too small?
3.

$$\vec{X} = \begin{bmatrix} 1 \\ 3 \\ 5 \\ 7 \end{bmatrix} \rightarrow \vec{Y} = \begin{bmatrix} 2 \\ 4.5 \\ 6.25 \\ 6.75 \end{bmatrix}$$

$Y = \theta_1 X + \theta_2$

$\theta_1 = 0.1$

$\theta_2 = 1$ } $\alpha = 0.01$

$\theta_1 :=$

$\theta_2 :=$