$$f(x,\theta) = f\left(x, \begin{bmatrix} a \\ b \end{bmatrix}\right) = ax + b = \begin{bmatrix} x & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} x & 1 \end{bmatrix} \theta$$

linear in x? Not in general

linear in $\begin{bmatrix} a \\ b \end{bmatrix}$? YES

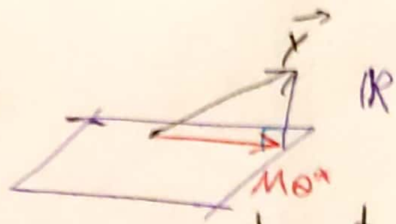
$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_N)\}$$

$$C(\theta) = C(a,b) = \sum_{i=1}^{N} d(f(x_i,\theta), y_i) = \sum_{i=1}^{N} (f(x_i,\theta) - y_i)^2$$

$$C(a,b) = \sum_{i=1}^{N} (ax_i + b - y_i)^2 = \sum_{i=1}^{N} \left(\begin{bmatrix} x_i & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} - y_i \right)^2$$

$$= \left\| \begin{bmatrix} ax_1 + b - y_1 \\ ax_2 + b - y_2 \\ \vdots \\ ax_N + b - y_n \end{bmatrix} \right\|^2 = \left\| \underbrace{\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ \vdots & \\ x_N & 1 \end{bmatrix}}_{N \times 2} \underbrace{\begin{bmatrix} a \\ b \end{bmatrix}}_{2 \times 1} - \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{N \times 1} \right\|^2$$

$$C(\theta) = \|M\theta - \vec{y}\|^2$$



$$\begin{bmatrix} a^* \\ b^* \end{bmatrix} = \theta^* = \operatorname{argmin} \|M\theta - \vec{y}\|$$

least squares

$$\theta^* = (M^T M)^{-1} M^T y$$

$$M^T M = \begin{bmatrix} x_1 & x_2 & \cdots & x_N \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} x_i^2 & \sum x_i \\ \sum x_i & N \end{bmatrix}$$

$$M^T y = \begin{bmatrix} x_1 & x_2 & \cdots & x_N \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum x_i y_i \\ \sum y_i \end{bmatrix}$$

Polynomial Regression

$$f(x, \theta) = a_p x^p + a_{p-1} x^{p-1} + \cdots + a_2 x^2 + a_1 x + a_0$$

$$\theta = [a_p, a_{p-1}, \cdots, a_2, a_1, a_0]^T \in \mathbb{R}^{p+1}$$

$$f : \mathbb{R} \times \mathbb{R}^{p+1} \longrightarrow \mathbb{R}$$

$$f(x, \theta) = \underbrace{\begin{bmatrix} x^p & x^{p-1} & \cdots & x^2 & x & 1 \end{bmatrix}}_{V^T} \underbrace{\begin{bmatrix} a_p \\ a_{p-1} \\ \vdots \\ a_2 \\ a_1 \\ a_0 \end{bmatrix}}_{\theta} = V^T \theta$$

$$C(\theta) = \sum_{i=1}^{N} \left(f(x_i, \theta) - y_i \right)^2$$

$$D = D_{test} \cup D_{train}$$

$$\text{Train Error} = \sum_{(x_i, y_i) \in D_{train}} \left(f(x_i, \theta) - y_i \right)^2$$

$$\text{Test Error} = \sum_{(x_i, y_i) \in D_{test}} \left(f(x_i, \theta) - y_i \right)^2$$