Linear Regression

(x_i, y_i), $\quad i = 1, 2, \dots, N$

$x_i \in \mathbb{R}^d$, $\quad y_i \in \mathbb{R}$. (used to be ∓ 1)

In linear regression y_i's can take any value.

Our goal: find a function $y = h(x)$ s.t.

$y_i \approx h(x_i)$, $\quad i = 1, 2, \dots, N$.

$$h(x_i) = \sum_{j=1}^{d} w_j x_{ij} = w^T x_i$$

Determine $\quad w = [w_1 \; w_2 \dots w_d]^T$

(You may use $\quad w = [1 \; w_1 \; w_2 \dots w_d]^T$

$$h(x_i) = \sum_{j=1}^{d} w_j x_{ij} + b = \sum_{j=0}^{d} w_j x_{ij}$$

where $\quad x_i = [b \; x_{i1} \; x_{i2} \dots x_{id}]^T$)

Use the method of least squares to find w vector:

Cost function:

$$J(w) = \frac{1}{2} \sum_{i=1}^{N} \left(h(x_i) - y_i\right)^2$$

To minimize the cost function

$$\nabla_w (J) = \begin{bmatrix} \frac{\partial J(w)}{\partial w_1} \\ \vdots \\ \frac{\partial J(w)}{\partial w_d} \end{bmatrix} = 0$$

e.g. $$\frac{\partial(J(w))}{\partial w_1} = \frac{1}{2}\sum_{i=1}^{N}\frac{\partial}{\partial w_1}\left(\left(\sum_{j=0}^{d} w_j x_{ij}\right)-y_i\right)^2$$

$$= \sum_{i=1}^{N} x_{1i}\left(\sum_{j=0}^{d}(w_j x_{ij}) - y_i\right) = 0$$

$$\vdots \qquad\qquad \vdots$$

$$\frac{\partial J(w)}{\partial w_d} = \sum_{i=1}^{N} x_{di}\left(\sum_{j=0}^{d}(w_j x_{ij}) - y_i\right) = 0$$

We find the w from these equations!

$$w = [w_0\ w_1\ w_2 \dots w_d]^T$$

Heuristic solution:

$$\underbrace{Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}}_{N \times 1} \approx \underbrace{\begin{bmatrix} 1 & x_{11} & \cdots & x_{d1} \\ 1 & x_{12} & \cdots & x_{d2} \\ \vdots & & & \\ 1 & x_{1N} & \cdots & x_{dN} \end{bmatrix}}_{\substack{N \times d+1 \\ \phi}} \underbrace{\begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix}}_{\substack{d+1 \times 1 \\ w}}$$

, w_o =b

$$\underbrace{\phi^T\phi}_{\substack{d+1 \text{ by } d+1 \\ \text{matrix}}}\, w = \underbrace{\phi^T Y}_{\substack{d+1 \times 1 \\ \text{vector}}} \quad \left(\text{Multiply both sides by } \phi^T\right)$$

$$w = (\phi^T\phi)^{-1}\phi^T Y$$

Least squares solution

Assumption $\phi^T\phi$ is invertible!