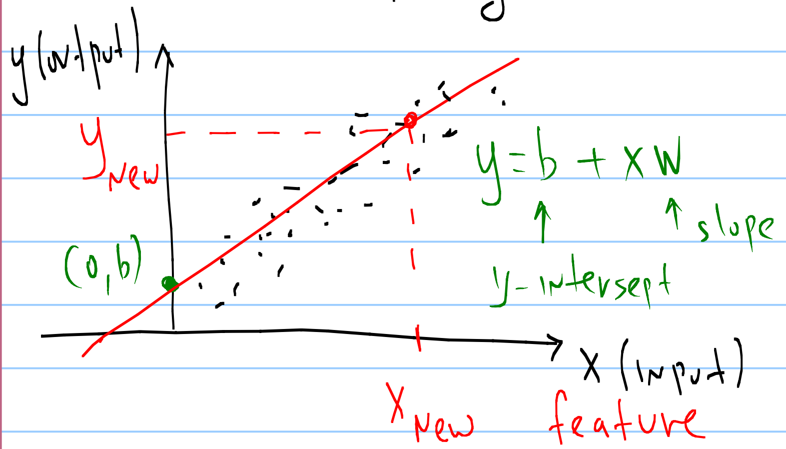
Note Title 4/11/18 1/6/2000

Regression (Linear)

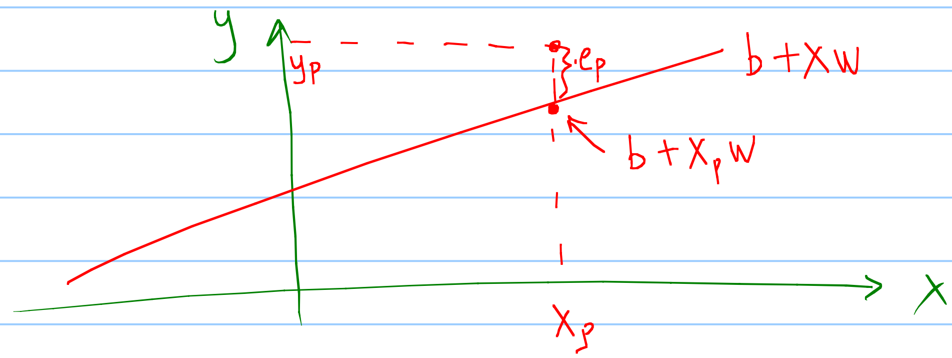
N=1



Basics of Regression

- Regression data : P input/output pairs

$$\{(\bar{x}_1, y_1), (\bar{x}_2, y_2), \ldots, (\bar{x}_p, y_p)\}, \quad \bar{x} \in \mathbb{R}^N$$



$$e_p = (b + x_p W - y_p)^2$$

for $\bar{x} \in \mathbb{R}^N$: $\underset{1\times 1}{b} + \underset{(1\times N)}{\bar{x}^T} \underset{(N\times 1)}{\bar{w}}$

$$\boxed{g(b, \bar{w}) = \sum_{p=1}^{P} (b + \bar{x}^T \bar{w} - y_p)^2}$$

$$\tilde{x}_p = \begin{bmatrix} 1 \\ \bar{x}_p \end{bmatrix}_{(N+1)\times 1} \qquad \tilde{w} = \begin{bmatrix} b \\ \bar{w} \end{bmatrix}_{(N+1)\times 1}$$

$$\tilde{x}_p^T \tilde{w} = \begin{bmatrix} 1 & \bar{x}_1^T \end{bmatrix} \begin{bmatrix} b \\ \bar{w} \end{bmatrix} = b + \bar{x}_p^T \bar{w}$$

$$g(\tilde{w}) = \sum_{p=1}^{P} \left(\tilde{x}_p^T \tilde{w} - y_p\right)^2$$

$$\min_{\tilde{w}} g(\tilde{w}) \longrightarrow \tilde{w}^* = \arg\min_{\tilde{w}} g(\tilde{w})$$

Stationary point

$$\nabla_{\tilde{w}} g(\tilde{w}) = 0$$

$$\rightarrow \nabla \sum_{p=1}^{P} \left(\tilde{x}_p^T \tilde{w} - y_p\right)^2 = 0$$

$$\rightarrow \sum_{p=1}^{P} \nabla \left(\tilde{x}_p^T \tilde{w} - y_p\right)^2 = 0$$

$$\rightarrow \sum_{p} 2\left(\tilde{x}_p^T \tilde{w} - y_p\right) \cdot \nabla\left(\tilde{x}_p^T \tilde{w} - y_p\right) = 0$$

$$\rightarrow \sum_{p} 2\left(\tilde{x}_p^T \tilde{w} - y_p\right)\left(\nabla_{\tilde{w}}\left(\tilde{x}_p^T \tilde{w}\right) - \cancel{\nabla y_p}^{\,0}\right) = 0$$

$$\nabla\left(\tilde{x}_p^T \tilde{w}\right) = \nabla\left(\tilde{w}^T \tilde{x}_p\right) = \tilde{x}_p$$

$$\rightarrow 2 \sum_{p=1}^{P} \left(\tilde{x}_p^T \tilde{w} - y_p\right)\tilde{x}_p = 0$$

$$\rightarrow \sum_{p=1}^{P} \tilde{x}_p \left(\tilde{x}_p^T \tilde{w} - y_p\right) = 0$$

$$\rightarrow \sum_{p=1}^{P} \tilde{x}_p \tilde{x}_p^T \tilde{w} - \sum_{p=1}^{P} \tilde{x}_p y_p = 0$$

$$\rightsquigarrow \sum_{p=1}^{P} \underbrace{\tilde{x}_p \tilde{x}_p^T}_{[(N+1)\times 1][1\times(N+1)\times 1]} \tilde{w} = \sum_{p=1}^{P} y_p \cdot \tilde{x}_p$$

$$\rightsquigarrow \underbrace{\left(\sum_{p=1}^{P} \tilde{x}_p \tilde{x}_p^T \right)}_{A} \tilde{w} = \underbrace{\sum_{p=1}^{P} y_p \cdot \tilde{x}_p}_{b}$$

$$A \tilde{w} = b$$

- if A is invertible $\longrightarrow \tilde{w}^* = A^{-1} b$
- if A is non-invertible $\longrightarrow \tilde{w}^* = \underbrace{A^{+}}_{\text{pseudo-inverse}} b$

Feature transformation

$$x \longrightarrow f(x)$$

example $f(x) = \sin(2\pi x)$

$y_p \approx b + f(x_p) \cdot w$, for all p (linear regression)

General Case

$$\tilde{f}_p = \begin{bmatrix} 1 \\ \hline f_p \end{bmatrix}, \quad \tilde{w} = \begin{bmatrix} b \\ \hline w \end{bmatrix}$$

cost function

$$g(\tilde{w}) = \sum_{p=1}^{P} \left(\tilde{f}_p^T \tilde{w} - y_p \right)^2$$

$$\nabla g(\tilde{w}) = 0$$

$$\Rightarrow \left(\sum_{p=1}^{P} \tilde{f}_p \tilde{f}_p^T \right) \tilde{w} = \sum_{p=1}^{P} \tilde{f}_p y_p$$

$$A x = b$$