Training Model

Problem: theory: $Y = f(X) + \varepsilon$

Given a training set S

find $\hat{f}$, an estimate for f

Finding $\hat{f}$

$\Leftrightarrow$ finding $\hat{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$

p pieces of info

n example

$$S = \left\{ \begin{array}{ll} 1 & (x_{11}, x_{12}, x_{13}, \dots, x_{1p}, y_1), \\ 2 & (x_{21}, x_{22}, x_{23}, \dots, x_{2p}, y_2), \\ & \\ n & (x_{n1}, x_{n2}, \dots, x_{np}, y_n) \end{array} \right\}$$

$$\vec{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1p} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2p} \\ & & & & \\ x_{n1} & x_{n2} & & \dots & x_{np} \end{bmatrix} \qquad \begin{bmatrix} y_1 \\ y_2 \\ \\ y_n \end{bmatrix} = \vec{y}$$

$$\vec{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$$

$$a_1^2 + a_2^2 + a_3^2 = \vec{a}^T \vec{a}$$

$$= \begin{pmatrix} a_1 & a_2 & a_3 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$$

$$= a_1^2 + a_2^2 + a_3^2$$

$$\hat{Y} = \hat{f}(\vec{X})$$

$$= \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \cdots \hat{\beta}_p X_p$$

$$\hat{y}_1 = \hat{\beta}_0 + \hat{\beta}_1 X_{11} + \hat{\beta}_2 X_{12} + \cdots \hat{\beta}_p X_{1p}$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \cdots + \hat{\beta}_p X_{ip}$$

ith example

Linear Regression

Assume $\hat{Y} = \hat{f}(\vec{X}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p$

$$\vec{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}$$

Find $\hat{\beta}$ by minimize mean squared error using S.

$$MSE = \frac{(y_1-\hat{y}_1)^2+(y_2-\hat{y}_2)^2+\cdots+(y_n-\hat{y}_n)^2}{n}$$

$$= \frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 = \frac{1}{n}(\vec{y}-\hat{\vec{y}})^T(\vec{y}-\hat{\vec{y}})$$

$$= \frac{1}{n}(\underset{(n\times 1)}{\vec{y}} - \underbrace{\vec{X}\vec{\beta}}_{\substack{n\times \cancel{(p+1)}\cdot\cancel{(p+1)}\times 1 \\ n\times 1}})^T(\vec{y}-\vec{X}\vec{\beta})$$

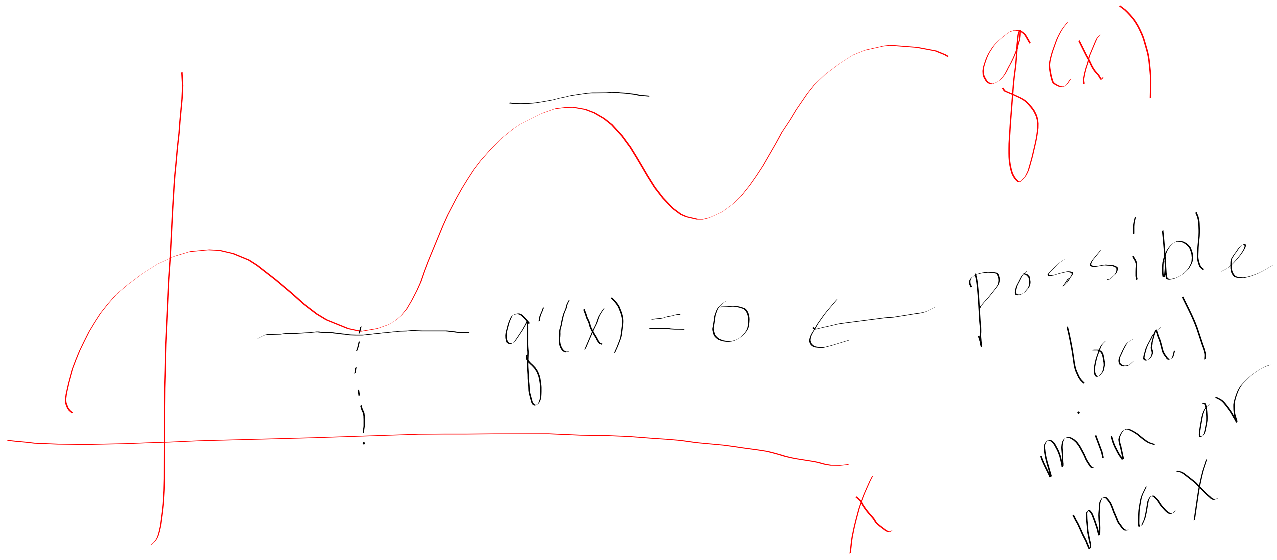
$$\underset{1\times \cancel{n}}{(n\times 1)^T}\;\underset{\cancel{n}\times 1}{(n\times 1)}$$

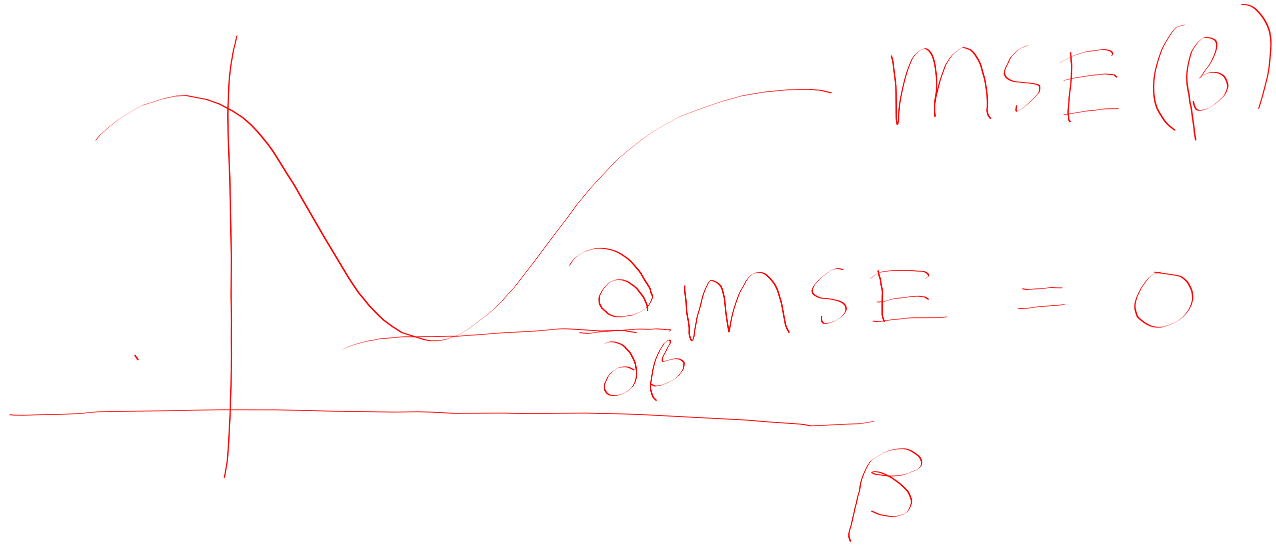
$$1\times 1$$

$$X \quad \vec{\beta}$$

$$\underbrace{\begin{bmatrix} 1 & x_{11} & x_{12} & x_{13} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & x_{23} & & x_{2p} \\ \vdots & \vdots & \vdots & & & \\ 1 & x_{n1} & x_{n2} & x_{n3} & \cdots & x_{np} \end{bmatrix}}_{\vec{X}} \underbrace{\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \vdots \\ \hat{\beta}_p \end{bmatrix}}_{\vec{\beta}}$$

$$\vec{X}\vec{\beta} = \hat{\vec{y}}$$


g(x)
g'(x) = 0 ← possible local min or max
x


MSE(β)
∂/∂β MSE = 0
β

$$\frac{\partial}{\partial \vec{\beta}} MSE = \frac{2}{n}(-\vec{X})^T(\vec{y} - \vec{X}\hat{\vec{\beta}})$$

$$\frac{\partial}{\partial \vec{\beta}} MSE = 0$$

$$0 = \cancel{\frac{2}{n}}(\cancel{-}\vec{X})^T(\vec{y} - \vec{X}\hat{\vec{\beta}})$$

solve for $\vec{\beta}$

$$0 = \vec{X}^T\vec{y} - \underbrace{\vec{X}^T\vec{X}\hat{\vec{\beta}}}$$

$$\cancel{(\vec{X}^T\vec{X})^{-1}\vec{X}^T\vec{X}}\,\hat{\vec{\beta}} = (\vec{X}^T\vec{X})^{-1}\vec{X}^T\vec{y}$$

$$\hat{\vec{\beta}} = (\vec{X}^T\vec{X})^{-1}\vec{X}^T\vec{y}$$