Math 342W Lecture 8

$\mathcal{Y} = \mathbb{R}$, $\mathcal{H} = \{ \vec{w} \cdot \vec{x} : \vec{w} \in \mathbb{R}^{p+1} \}$ (linear models)

$\mathcal{A}_{OLS} : \underbrace{\frac{\partial}{\partial \vec{w}} [SSE]}_{} := \vec{0}_{p+1}$

$$\frac{\partial}{\partial \vec{w}} [\vec{e}^T \vec{e}] = \frac{\partial}{\partial \vec{w}} [(\vec{y} - \hat{\vec{y}})^T (\vec{y} - \hat{\vec{y}})]$$

$$= \frac{\partial}{\partial \vec{w}} [(\vec{y} - X\vec{w})^T (\vec{y} - X\vec{w})] \Longrightarrow$$ check previous notes for more...

$$= \frac{\partial}{\partial \vec{w}} [\vec{y}^T \vec{y} - 2\vec{w}^T X^T \vec{y} + \vec{w}^T X^T X \vec{w}] \Rightarrow$$ Here from previous notes

$$= \frac{\partial}{\partial \vec{w}} [\vec{y}^T \vec{y}] - 2 \cdot \frac{\partial}{\partial \vec{w}} [\vec{w}^T X^T \vec{y}] + \frac{\partial}{\partial \vec{w}} [\vec{w}^T X^T X \vec{w}]$$

$$= \vec{0}_{p+1} - 2 \frac{\partial}{\partial \vec{w}} [\vec{w}^T X^T \vec{y}] + \frac{\partial}{\partial \vec{w}} [\vec{w}^T X^T X \vec{w}]$$

$$= -2 X^T \vec{y} + \frac{\partial}{\partial \vec{w}} [\vec{w}^T X^T \vec{w}]$$

$$\frac{\partial}{\partial \vec{x}} [\vec{x}^T A \vec{x}] =$$

$$A\vec{x} = \begin{bmatrix} a_{11} x_1 + a_{12} x_2 + \ldots + a_{1d} x_d \\ \vdots \\ a_{d1} x_1 + a_{d2} x_2 + \ldots + a_{dd} x_d \end{bmatrix} = \begin{bmatrix} \leftarrow \vec{a}_1 \rightarrow \\ \vdots \\ \leftarrow \vec{a}_d \rightarrow \end{bmatrix} \vec{x} = \begin{bmatrix} \vec{a}_1 \vec{x} \\ \vdots \\ \vec{a}_d \vec{x}_d \end{bmatrix}$$

$$\vec{x}^T(A\vec{x}) = [x_1, x_2, \ldots, x_d](A\vec{x})$$

$\downarrow$ * Symmetric Property *

$$\frac{\partial}{\partial x_1}\left[\vec{x}^T A \vec{x}\right] = 2(a_{11}x_1 + \ldots + a_{1d}x_d) = 2\vec{a}_{1\cdot}\vec{x}$$

$$\frac{\partial}{\partial x_2}\left[\vec{x}^T A \vec{x}\right] = 2(a_{21}x_1 + \ldots + a_{2d}x_d) = 2\vec{a}_{2\cdot}\vec{x}$$

$$\Rightarrow \frac{\partial}{\partial \vec{x}}\left[\vec{x}^T A \vec{x}\right] = \begin{bmatrix} 2\vec{a}_{1\cdot}\vec{x} \\ 2\vec{a}_{2\cdot}\vec{x} \\ \vdots \\ 2\vec{a}_{d\cdot}\vec{x} \end{bmatrix} = 2A\vec{x}$$

$$= -2X^T\vec{y} + 2X^TX\vec{\omega} := \vec{0}_{p+1}$$

$$\hookrightarrow X^TX\vec{\omega} = X^T\vec{y}$$

$$\Rightarrow (X^TX)^{-1}(X^TX)\vec{\omega} = X^T\vec{y}(X^TX)^{-1}$$

$$\Rightarrow I_{p+1}\vec{\omega} = (X^TX)^{-1}X^T\vec{y}$$

$$\boxed{\Rightarrow \vec{b}_{OLS} = (X^TX)^{-1}X^T\vec{y}}$$

$\vec{x}^*$ is a new record

$$\hat{y}^* = g(\vec{x}^*) = b_0 + b_1x_1^* + \ldots + b_px_p^*$$

$$= \vec{x}^* \cdot \vec{b} \in \mathbb{R}$$

$$\hat{y}_1 = g(\vec{x}_{1\cdot})$$

$$= \vec{x}_{1\cdot}\vec{b}$$

In general, we say that

$$\begin{aligned} \hat{y}_1 &= \vec{x}_{1\cdot}\vec{b} \\ \hat{y}_2 &= \vec{x}_{2\cdot}\vec{b} \quad \Rightarrow \boxed{\vec{\hat{y}} = X\vec{b}} \\ &\vdots \\ \hat{y}_n &= \vec{x}_{n\cdot}\vec{b} \end{aligned}$$

$g_0 = \bar{y}$ null model, no features, average

$\Rightarrow X = [\vec{1}_n]$, $\vec{b} = b_0 = (\vec{1}_n^T \vec{1}_n)^{-1}(\vec{1}_n^T \vec{y})$ $\curvearrowright$ solve this's

$$\Rightarrow \vec{\hat{y}} = X\vec{b} = [\vec{1}_n]\bar{y} = \begin{bmatrix} \bar{y} \\ \vdots \\ \bar{y} \end{bmatrix}$$

If $X^T X$ is invertible, then $\text{rank}[X^T X] = p+1$,

$\forall \vec{v} \neq \vec{0}_{p+1}$, $X^T X \vec{v} \neq \vec{0}_{p+1} \Rightarrow \vec{v}^T X^T X \vec{v} \neq 0$ $\quad$ $\text{rank}[X] = p+1$

$\Rightarrow (X\vec{v})^T (X\vec{v}) \neq 0 \Rightarrow (X\vec{v}) \neq \vec{0} \Rightarrow$ Implies X full rank

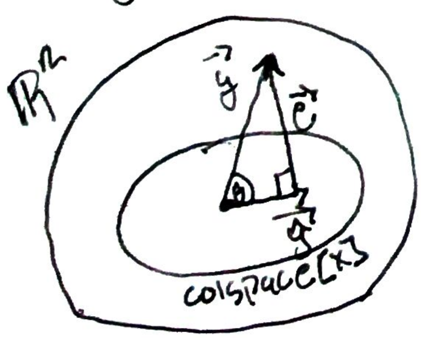
$$X = \begin{array}{c} \begin{bmatrix} \vec{1} & \vec{x}_{\cdot 1} & \cdots & \vec{x}_{\cdot p} \end{bmatrix} \\ \quad b_0 \quad\; b_1 \quad \cdots \quad b_p \end{array}$$

$\Rightarrow$ In order to compute $\vec{b}_{OLS}$ each feature in X must be (all) linearly independent

$$\text{Colspace}[X] = \text{span}\{\vec{1}, \vec{x}_{\cdot 1}, \ldots, \vec{x}_{\cdot p}\}$$

$$= \{\omega_0 \vec{1} + \omega_1 \vec{x}_{\cdot 1} + \ldots + \omega_p \vec{x}_{\cdot p} : \vec{\omega} \in \mathbb{R}^{p+1}\}$$

$$\Rightarrow \vec{\hat{y}} = b_0 \vec{1} + b_1 \vec{x}_{\cdot 1} + \ldots + b_p \vec{x}_{\cdot p} \in \text{colspace}[X]$$



$\vec{\hat{y}}$ is a projection of $\vec{y}$ onto the subspace of $\mathbb{R}^n$ which is colspace[X]

$\vec{\hat{y}} = X(X^T X)^{-1} X^T \vec{y} = H$ "Hat Matrix"