Math 342W Lecture 7

$y = \mathbb{R}$, $x = \mathbb{R}$ $\Rightarrow$ Ordinary Least Squares $\mathcal{A}$

$x_{raw} = \{\text{red}, \text{green}\}$

$x = \{0, 1\}$

$$g(x) = \begin{cases} \bar{y}_0 & \text{if } x=0 \\ \bar{y}_1 & \text{if } x=1 \end{cases}$$

$$g(x) = \underbrace{\bar{y}_0}_{b_0} + \underbrace{(\bar{y}_1 - \bar{y}_0)}_{b_1} x$$

<u>Formulas From Proof</u>

$$\bar{y} = (1-p)\bar{y}_0 + p\bar{y}_1$$

$$b_0 = \bar{y} - b_1$$

$$b_1 = \bar{y}_1 - \bar{y}_0$$

Proving "betas" b_0 and b_1

$$\bar{x} = \frac{\sum x_i}{n} = \frac{n_1}{n} = p \quad \text{x bar}$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{\sum_{i:x_i=0} y_i + \sum_{i:x_i=1} y_i}{n}$$

$$= \frac{\sum_{i:x_i=0} y_i}{n} \cdot \frac{n_0}{n_0} + \frac{\sum_{i:x_i=1} y_i}{n} \cdot \frac{n_1}{n_1}$$

$$= (1-p)\bar{y}_0 + p\bar{y}_1 \quad \text{y bar}$$

$$b_1 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \frac{\sum_{i:y_i=1} y_i - np\bar{y}}{n_1 - np^2} \cdot \frac{\frac{1}{n}}{\frac{1}{n}}$$

$$= \frac{p\bar{y}_1 - p\bar{y}}{p - p^2} = \frac{\bar{y}_1 - \bar{y}}{1-p} = \frac{\bar{y}_1 - (p\bar{y}_1 + (1-p)\bar{y}_0)}{1-p} = \frac{(1-p)\bar{y}_1 - (1-p)\bar{y}_0}{1-p}$$

$$= \bar{y}_1 - \bar{y}_0 \quad \text{beta 1}$$

$\mathcal{H} = w_0 + w_1 x_1 : w_0, w_1 \in \mathbb{R}$ $\Rightarrow$ Notice from 2 features we only "use" 1.

↳ $\mathbb{1}$ $(n-1)$ features $= 0$ imply $\mathbb{1}(n^{th}) = 1$

↳ Reference!!

How well does $g(x)$ predict? $\Rightarrow$ OLS minimizes SSE

$$\left.\begin{aligned} SSE &= \sum e_i^2 = \sum (y_i - \hat{y}_i)^2 \\ MSE &= \frac{1}{n-2} SSE \end{aligned}\right] \text{Not interpretable}$$

$RMSE = \sqrt{MSE} \rightarrow$ Interpretable

A 95% prediction interval $\approx \hat{y} \pm 2\,RMSE$

Another way to gauge a regression model's performance is R^2 which means "proportion of variance explained"

<u>Recall:</u>

$g_0 = \bar{y}$

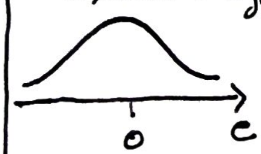
$SSE_0 = \sum (y_i - \bar{y})^2 = (n-1)s_y^2$

$\downarrow$

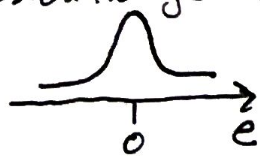
$SST \equiv$ Sum Squares Total

If your model can't beat SST $\Rightarrow$ Trash

Residuals (g_0)



Residuals ($g(OLS)$)



$$R^2 = \frac{\Delta s^2}{s_y^2} = \frac{s_y^2 - s_e^2}{s_y^2} = \frac{SST - SSE}{SST}$$

$$= 1 - \frac{SSE}{SST} \leq 1$$

R^2 and RMSE are inversely related, s.t

$R^2 \uparrow$, $RMSE \downarrow$

$$\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{R}^n, \quad \vec{x}_{\cdot 1} = \begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1n} \end{bmatrix} \in \mathbb{R}^n, \quad \{b_0, b_1\} \in \mathbb{R}$$

$$\begin{matrix} \hat{y}_1 = b_0 + b_1 x_{11} \\ \hat{y}_2 = b_0 + b_1 x_{12} \\ \vdots \\ \hat{y}_n = b_0 + b_1 x_{1n} \end{matrix} \equiv \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} = b_0 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + b_1 \begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1n} \end{bmatrix} \Rightarrow \vec{\hat{y}} = b_0 \vec{1} + b_1 \vec{x}_{\cdot 1}$$

s.t. $\vec{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$

$$X = [\vec{1} : \vec{x}_{\cdot 1}] \Rightarrow \vec{\hat{y}} = X\vec{b}$$

$\equiv \vec{\hat{y}} \in \text{Span}\{\vec{1}, \vec{x}_{\cdot 1}\}$

$\equiv \vec{\hat{y}} \in \text{Colspace}[X]$

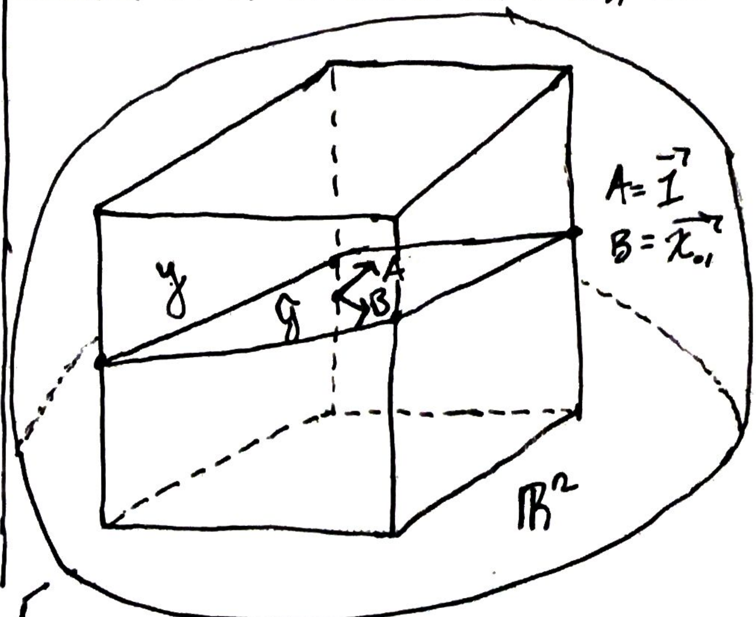
In general,

$$X = [\vec{1} \mid \vec{x}_{\cdot 1} \mid \dots \mid \vec{x}_{\cdot p}]$$

$$\mathcal{H} = \{w_0 + w_1 x_1 + \dots + w_p x_p\}$$

↳ $w_0, \dots, w_p \in \mathbb{R}$

There exists an algebraic abstract structure for this n-dimensional space and $(n-1)$ dimensional subspace



$\vec{\hat{y}}$ is casted down from y ↳ Currently there will always be loss

$$A_{OLS}: \vec{b} = \underset{\vec{w} \in \mathbb{R}^{p+1}}{\text{argmin}} \{SSE\} = \hat{\vec{y}} = \vec{w}_0 \vec{1}_n + \vec{w}_1 \vec{x}_{\cdot 1} + \dots + \vec{w}_p \vec{x}_{\cdot p}$$

$$\Rightarrow X\vec{w}, \quad \Rightarrow \hat{\vec{y}} = X\vec{w} \text{ and } \vec{e} = \vec{y} - X\vec{w}$$

$$SSE = \sum e_i^2 = \vec{e}^T \cdot \vec{e} = (\vec{y} - X\vec{w})^T(\vec{y} - X\vec{w})$$

$$= (\vec{y}^T (X\vec{w})^T)(\vec{y} - X\vec{w})$$

$$= \vec{y}^T\vec{y} - \vec{w}^T X^T \vec{y} - \vec{y}^T + X\vec{w}^T + \vec{w}^T X^T X \vec{w} = SSE$$

$$\frac{\partial SSE}{\partial w_0} := 0 \quad \vdots \quad \frac{\partial SSE}{\partial w_p} := 0 \quad \Rightarrow \frac{\partial SSE}{\partial \vec{w}} = \vec{0}_{p+1}$$

Work on $\frac{\partial}{\partial \vec{x}} [\underbrace{\vec{x}^T A \vec{x}}_{\text{Quadratic Form}}]$

Linear Algebra Tangent

$a \in \mathbb{R}$, costant WRT $x_1, \dots, x_n$

$$\frac{\partial}{\partial \vec{x}}[a] = \begin{bmatrix} \frac{\partial}{\partial x_1}[a] \\ \vdots \\ \frac{\partial}{\partial x_n}[a] \end{bmatrix} = \vec{0}_n$$

$\vec{a} \in \mathbb{R}^n$, costant WRT x_i's

$$\frac{\partial}{\partial \vec{x}}[\vec{a}^T \vec{x}] = \begin{bmatrix} \frac{\partial}{\partial x_1}[a_1x_1 + \dots + a_nx_n] \\ \vdots \\ \frac{\partial}{\partial x_n}[a_1x_1 + \dots + a_nx_n] \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \vec{a} \;\; \# \;\; \vec{a}^T$$

$a, b \in \mathbb{R}$, Constant WRT x_i's

$$\frac{\partial}{\partial \vec{x}}[af(\vec{x}) + bg(\vec{x})] = \begin{bmatrix} \frac{\partial}{\partial x_1}[af(\vec{x}) + bg(\vec{x})] \\ \vdots \\ \frac{\partial}{\partial x_n}[af(\vec{x}) + bg(\vec{x})] \end{bmatrix} = a\frac{\partial}{\partial \vec{x}} f(\vec{x}) + b\frac{\partial}{\partial \vec{x}} g(\vec{x})$$