Lecture 8:

$y = R, \; P = 2$

linear models: $H = \{ w_0 + w_1 x_1 + w_2 x_2 : \vec{w} : \mathbb{R}^3 \}$

$\mathbb{D} := (X, Y)$

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \end{bmatrix} \in \mathbb{R}^{n \times 3}$$

Algorithm will return one $\vec{w}$

and all $\hat{y}$'s can be via:

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

$\vec{e} = \vec{y} - \vec{\hat{y}}$

$\mathcal{A}$: OLS

$SSE := \sum e_i^2$

$= \vec{e}^T \vec{e}$ [dot / inner Product]

$\mathcal{A} : \vec{b} = \underset{\vec{w} \in \mathbb{R}^3}{\operatorname{argmin}} \{ \vec{e}^T \vec{e} \}$

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

$([\;] - [\;])^T ([\;\;] - [\;\;]) = (\vec{y}^T - \vec{\hat{y}}^T)(\vec{y} - \vec{\hat{y}})$

$$= \vec{y}^T \vec{y} - \vec{\hat{y}}^T \vec{y} - \vec{y}^T \vec{\hat{y}} + \vec{\hat{y}}^T \vec{\hat{y}}$$

$$= \vec{y}^T \vec{y} - 2 \vec{\hat{y}}^T \vec{y} + \vec{\hat{y}}^T \vec{\hat{y}}$$

$$A. \; \vec{b} = \underset{\vec{w} \in R^3}{\operatorname{argmin}} \{ \vec{e}^T \vec{e} \}$$

$$= \vec{y}^T \vec{y} - 2 \vec{w}^T X^T + \vec{w}^T X^T X \vec{w}$$

$$= SSE$$

$$b_0: \; \frac{\partial}{\partial w_0} [SSE] \stackrel{set}{=} 0$$

$$b_1: \; \frac{\partial}{\partial w_1} [SSE] \stackrel{set}{=} 0$$

$$b_2: \; \frac{\partial}{\partial w_2} [SSE] \stackrel{set}{=} 0$$

$$\frac{\delta}{\delta \vec{w}}[SSE] := \begin{bmatrix} \frac{\delta}{\delta w_0}[SSE] \\ \frac{\delta}{\delta w_1}(SSE) \\ \frac{\delta}{\delta w_2}(SSE) \end{bmatrix} \overset{set}{=} \vec{0}_3$$

Back to math 231:-

Let $\vec{X} \in \mathbb{R}^n$, a scalar constant w.r.t all X_J's.

with respect to

$$\frac{\delta}{\delta \vec{X}}[a] = \vec{0}_n$$

Let $\vec{X} \in \mathbb{R}^n$, $\vec{a}$, column vectore constant w.r.t all X_J's

$$\frac{\delta}{\delta \vec{X}}[\vec{a}^T \vec{X}] = \begin{bmatrix} \frac{\delta}{\delta x_1}[a_1x_2 + a_2x_2 + \dots + a_nx_n] \\ \\ \\ \end{bmatrix}$$

$$= \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \vec{a}$$

f, g are both $R^n \rightarrow R^n$ func, a, b constants

$$\frac{\partial}{\partial \vec{x}} \left[a f(\vec{x}) + b g(\vec{x}) \right]$$

$$\stackrel{?}{=} a \frac{\partial}{\partial \vec{x}} \left[f(\vec{x}) \right] + b \frac{\partial}{\partial \vec{x}} \left[g(\vec{x}) \right]$$

$$\stackrel{?}{=} \begin{bmatrix} \frac{\partial}{\partial x_1} \left[a f_1(\vec{x}) + b g(\vec{x}) \right] \\ \vdots \\ \vdots \end{bmatrix}$$

Next page

ARNA

Let $A \in \mathbb{R}^{n \times n}$ be a ~~non~~ symmetric matrix of constants w.r.t. x_j's,

$$\frac{\partial}{\partial \vec{x}} \left[\overset{1\times n}{\vec{x}^T} \overset{n\times n}{A} \overset{n\times 1}{\vec{x}} \right] \quad (\text{dimen } (1,1))$$

⇓

FYI: Quadratic form

$$= \frac{\partial}{\partial \vec{x}} \left[\vec{x}^T \begin{bmatrix} \leftarrow \vec{a}_1 \rightarrow \\ \leftarrow \vec{a}_2 \rightarrow \\ \leftarrow \vec{a}_n \rightarrow \end{bmatrix} \begin{bmatrix} \uparrow \\ \vec{x} \\ \downarrow \end{bmatrix} \right]$$

$$= \frac{\partial}{\partial \vec{x}} \left[\vec{x}^T \begin{bmatrix} \vec{a}_1 \vec{x} \\ \vec{a}_2 \vec{x} \\ a_n \vec{x} \end{bmatrix} \right]$$

$$= \frac{\partial}{\partial \vec{x}} \left[\begin{bmatrix} x_1 \, x_2 \, \cdots \, x_n \end{bmatrix} \begin{bmatrix} \vec{a}_1 \vec{x} \\ \vdots \\ \vec{a}_n \vec{x} \end{bmatrix} \right]$$

$$= \frac{\partial}{\partial \vec{x}} \left[x_1 \vec{a}_1 \vec{x} + x_2 \vec{a}_2 \vec{x} + \dots x_n a_{n,} \vec{x} \right]$$

$$= \frac{\partial}{\partial \vec{x}} \left[x_1 \left(a_{11} x_1 + a_{12} x_2 + \dots + a_{1n} x_n \right) \right.$$

$$+ x_2 \left(a_{21} x_1 + a_{22} x_2 + \dots a_{2n} x_n \right)$$

$$+ \dots + x_n \left(a_{n1} x_{\cancel{n}\,2} + a_{n2} x_2 + \dots \right.$$

$$\left. \left. + a_{nn} x_n \right) \right]$$

$$= \frac{\partial}{\partial \vec{x}} \left[a_{11} x_1^2 + a_{12} x_1 x_2 + \dots + a_{1n} x_1 x_n + a_{21} x_2 x_1 \right.$$

$$+ a_{22} x_2^2 + \dots + a_{2n} x_2 x_n + a_{n1} x_n x_1$$

$$\left. + a_{n2} x_n x_2 + \dots + a_{nn} x_n^2 \right]$$

before

$$\frac{d}{\delta x_1}[\;] = 2a_{11}x_1 + a_{12}x_2 + a_{1n}x_n$$

$$+ a_{21}x_2 + a_{31}x_3 + \cdots + a_{n1}x_n$$

equals in $A = A^T$

$$= 2a_{11}x_1 + 2a_{21}x_2 + \cdots + 2a_{n1}x_n$$

$$= 2\vec{a}_1 \cdot x_1$$

$$\frac{\delta}{\delta x_2}(\cdot) = 2\vec{a}_2\vec{x}$$

$$\frac{\delta}{\delta\vec{x}}\left[\vec{x}^T A\vec{x}\right] =$$

$$= \begin{bmatrix} 2\vec{a}_1\vec{x} \\ 2\vec{a}_2 x_2 \\ \vdots \\ 2 a_n\vec{x} \end{bmatrix} = 2\begin{bmatrix} \vec{a}_1\vec{x} \\ \vec{a}_2\vec{x} \\ \vdots \\ \vec{a}_n\vec{x} \end{bmatrix}$$

$$= 2\begin{bmatrix} \vec{a}_1 \\ \vec{a}_2 \\ \vdots \\ \vec{a}_n \end{bmatrix}\vec{x} = 2A\vec{x}$$

$X \in \mathbb{R}^{n\times 3}$

$$\frac{\partial}{\partial \vec{w}}\left[SSE\right] = \frac{\partial}{\partial \vec{w}}\left[\vec{y}^T\vec{y} - 2\underbrace{\vec{w}^T X^T\vec{y}}_{(X^T\vec{y})^T\vec{w}} + \vec{w}^T X^T X\,\vec{w}\right]$$

$$= \vec{0}_3 - 2X^T\vec{y} + 2X^TX\vec{w}$$

$(X^Tx)^T$

$= X^T(X^T)^T$

$= X^TX$

$$\vec{0}_3 - \cancel{2}X^T\vec{y} + \cancel{2}X^TX\vec{w} \stackrel{set}{=} \vec{0}_3$$

$$\Rightarrow X^TX\vec{w} = X^T\vec{y}$$

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

assumed X^TX has inverse

[illegible]

$$\boxed{\vec{b} = \underset{\substack{p*p \\ (p+1)*(p+1)}}{(X^T X)^{-1}} \; \underset{\substack{p*n \\ (p+1)*n}}{X^T} \; \underset{\substack{n*1 \\ n*1}}{\vec{y}}}$$

OLS estimate valid

for all p.

$\Rightarrow X^T X$ full rank

Assume X is full rank

linearly Independent.

$$X = \begin{bmatrix} 1 & \uparrow & \uparrow \\ 1 & \vec{x}_1 & x_2 \\ 1 & \downarrow & \downarrow \\ 1 & & \end{bmatrix}$$

In general, X is full rank

$= p+1$

$$X = \begin{bmatrix} 1 & \uparrow & \uparrow & & \uparrow \\ 1 & \vec{x}_1 & \vec{x}_2 & \cdots & \vec{x}_p \\ 1 & \downarrow & \downarrow & & \downarrow \end{bmatrix} \; n$$

p by 1 row vector

$\vec{x}_*$ is a new observation $\in 1 \times \mathbb{R}^{p+1}$

$$g(\vec{x}_*) = \vec{x}_* \vec{b} = \hat{y}_*$$

$$\in \mathbb{R}^{n_* \times (p+1)}$$

$$g(X_*) = X_* \vec{b} = \vec{\hat{y}}_*$$

get back $\vec{\hat{y}}$ for all n original subjects in D

$$\vec{\hat{y}} = g(X) = X \cdot \vec{b}$$

$$= \underbrace{\overset{n\times(p+1)}{X} \overset{(p+1)\times(p+1)}{(X^T X)^{-1}} \overset{(p+1)\times n}{X^T}}_{H \in \mathbb{R}^{n\times n}} \vec{y} = \vec{\hat{y}}$$

"Hat matrix"

$$= H\vec{y}$$

$$= T(\vec{y})$$

linear function

$$[\quad] = [\quad][\quad]$$

rank + nullity = n

dim(colspace) + dim(nullspace) = n

$(p+1) + (n-(p+1)) = n$

degrees of freedom

Feel it

$$\vec{e} = \vec{y} - \vec{\hat{y}} = \vec{y} - H\vec{y} = (I - H)\vec{y}$$

$$\vec{y} = \vec{\hat{y}} + \vec{e}$$

$$I\vec{y} = \underbrace{H\vec{y}}_{T(\vec{y})} + \underbrace{(I-H)\vec{y}}_{U(\vec{y})}$$

ARNA

$$SSE = \vec{e}^T \vec{e}$$

$$MSE = \frac{1}{n-(p+1)} SSE$$

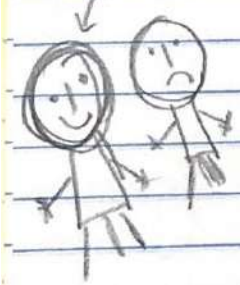
$$RMSE = \sqrt{MSE}$$

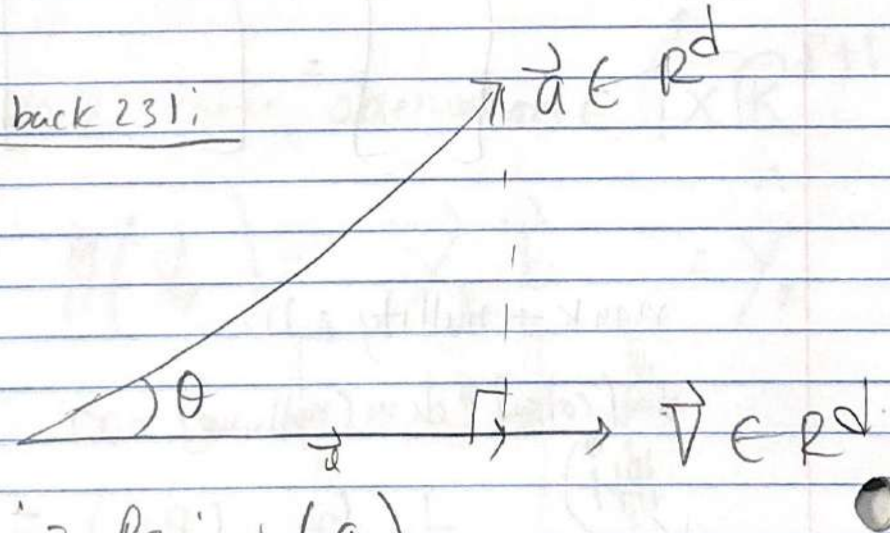
$$R^2 = 1 - \frac{SSE}{SST}$$

না মিসা হক অনা

বাংলাদেশ

going back 231;





$$l := \text{Proj}_{\vec{v}}(a)$$

orthogonal Projection

We want a formula for ℓ as a function of inputs $\vec{a}, \vec{v}$

Recall the "law of cosines"

$$\cos(\theta) = \frac{\vec{a}^T \vec{v}}{\|\vec{a}\| \, \|\vec{v}\|} = \frac{\|\vec{\ell}\|}{\|\vec{a}\|}$$

$$\Rightarrow \|\vec{\ell}\| = \frac{\vec{a} \cdot \vec{v}}{\|\vec{v}\|}$$

$$\vec{\ell} = \|\vec{\ell}\| \frac{\vec{v}}{\|\vec{v}\|} = \frac{\vec{a} \cdot \vec{v} \, \vec{v}}{\|\vec{v}\|^2}$$

$$= \frac{\overbrace{\underbrace{\vec{v}}_{d \times 1} \underbrace{\vec{v}^T}_{1 \times d}}^{d \times d}}{\|\vec{v}\|^2} \vec{a}$$

$$= \boxed{H \vec{a} = \text{Proj}_{\vec{v}}(\vec{a})}$$

Formula