Lecture - 08 02/20/2020

Review for lab.

$$g(x) = b_0 + b_1 x$$

$b_0 \uparrow \bar{y}_0$, $b_1 \uparrow \bar{y}_1 - \bar{y}_0$, $x \leftarrow [0,1]$

$$g(x) = b_0 \mathbb{1}_{x=0} + b_1 \mathbb{1}_{x=1}$$

$b_0 \uparrow \bar{y}_0$, $b_1 \uparrow \bar{y}_1$

$Y = \mathbb{R}$, $p = 2$

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

$$\mathbb{D} = \langle x, y \rangle \qquad X = \begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \end{bmatrix} \in \mathbb{R}^{n \times 3}$$

Algorithm will return one $\vec{w}$ and all $\hat{y}$ is can be!

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

dot / inner product

A: OLS

$$SSE := \sum e_i^2 = \vec{e}^{\,T} \vec{e} = A \cdot b = \underset{\vec{w} \in \mathbb{R}^3}{\text{argmin}} \left[\vec{e}^{\,T} \vec{e} \right] =$$

$\downarrow \mathrm{I}$ $\qquad$ $\mathrm{II} \downarrow$

I

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

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

$$\xrightarrow{\text{II}} \quad = \vec{y}^T \vec{y} - 2\vec{w}^T X^T + \vec{w}^T X^T X \vec{w} = SSE$$

$$\frac{\partial}{\partial \vec{w}}[SSE] = \begin{array}{l} b_0 : \\ b_1 : \\ b_2 : \end{array} \begin{bmatrix} \frac{\partial}{\partial w_0}[SSE] \overset{set}{=} 0 \\ \frac{\partial}{\partial w_1}[SSE] \overset{set}{=} 0 \\ \frac{\partial}{\partial w_2}[SSE] \overset{set}{=} 0 \end{bmatrix} \overset{set}{=} \vec{0}_3$$

let $\vec{x} \in \mathbb{R}^n$, $\vec{a}$ column vector (scalar) constant w.r.t all x_j's

$$\frac{\partial}{\partial \vec{x}}[a] = \vec{0}_n$$

$$\frac{\partial}{\partial \vec{x}}[\vec{a}^T \vec{x}] = \begin{bmatrix} \frac{\partial}{\partial x_1}\left[a_1 x_1 + a_2 x_2 + \dots + a_n x_n\right] \\ \\ \end{bmatrix}$$

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

f, g are both $\mathbb{R}^n \longrightarrow \mathbb{R}^n$ functions, a, b scalar constant

$$\frac{\partial}{\partial \vec{x}}\left[a f(\vec{x}) + b g(\vec{x})\right] = \begin{bmatrix} \frac{\partial}{\partial x_1}\left[a f_1(\vec{x}) + b g_1(\vec{x})\right] \\ \\ \end{bmatrix}$$

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

let $H \in \mathbb{R}^{n \times n}$ be a symmetric matrix of constants w.r.t x_j's

$$\frac{\partial}{\partial \vec{x}}\left[\vec{x}^{\,T} A \vec{x}\right] = \frac{\partial}{\partial \vec{x}}\left[\vec{x}^{\,T} \begin{bmatrix} \longleftarrow \vec{a}_1 \longrightarrow \\ \longleftarrow \vec{a}_2 \longrightarrow \\ \vdots \\ \longleftarrow \vec{a}_n \longrightarrow \end{bmatrix} \begin{bmatrix} \vec{x} \end{bmatrix}\right] =$$

FYI :- Quadratic form

$$= \frac{\partial}{\partial \vec{x}}\left[\vec{x}^{\,T} \begin{bmatrix} \vec{a}_1 \cdot \vec{x} \\ \vec{a}_2 \cdot \vec{x} \\ \vdots \\ \vec{a}_n \cdot \vec{x} \end{bmatrix}\right] = \frac{\partial}{\partial \vec{x}}\left[\begin{bmatrix} x_1, x_2, \ldots x_n \end{bmatrix} \begin{bmatrix} \vec{a}_1 \vec{x} \\ \vdots \\ \vec{a}_n \vec{x} \end{bmatrix}\right]$$

$$= \frac{\partial}{\partial \vec{x}}\left[x_1 \vec{a}_1 \cdot \vec{x} + x_2 \vec{a}_2 \cdot \vec{x} + \ldots + x_n \vec{a}_n \cdot \vec{x}\right]$$

$$= \frac{\partial}{\partial \vec{x}}\Big[x_1(a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n) + x_2(a_{21}x_1 + a_{22}x_2 + \ldots +$$

$$a_{2n}x_n) + \ldots + x_n(a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n)\Big]$$

$$= \frac{\partial}{\partial \vec{x}}\Big[a_{11}x_1^2 + a_{12}x_1x_2 + \ldots + a_{1n}x_1x_n + a_{21}x_2x_1 + a_{22}x_2^2 +$$

$$\ldots + a_{2n}x_2x_n + a_{n1}x_nx_1 + a_{n2}x_nx_2 + \ldots +$$

$$a_{nn}x_n^2\Big]$$

equal since $A = A^T$

$$\frac{\partial}{\partial x_1}\left[\ldots\right] = \left[a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + a_{21}x_2 +\right.$$

$$a_{31}x_3 + \ldots + a_{n1}x_n.$$

$$= 2a_{11}x_1 + 2a_{21}x_2 + \ldots + 2a_{n1}x_n = 2\vec{a}_1 \cdot \vec{x}$$

$$\frac{\partial}{\partial x_2}\left[\ldots\right] = 2\vec{a}_2 \vec{x}$$

$$\Longrightarrow \frac{\partial}{\partial x}\left[\vec{x} A \vec{x}\right] = \begin{bmatrix} 2\vec{a}_1 \vec{x} \\ 2\vec{a}_2 \vec{x} \\ \vdots \\ 2\vec{a}_n \vec{x} \end{bmatrix} = 2 \begin{bmatrix} \vec{a}_1 \\ \vec{a}_2 \\ \vdots \\ \vec{a}_n \end{bmatrix} \vec{x} = 2A\vec{x}$$

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

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

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

$\Longrightarrow (X^TX)^{-1}X^TX\vec{w} = (X^TX)^{-1}X^T\vec{y}$ assumed .. X^TX was invertible.

$\Longrightarrow \boxed{\vec{b} = (X^TX)^{-1}X^T\vec{y}}$ OLS estimate valid for all P

$\quad p\times p \quad p\times n \quad n\times 1$

Assume X is full rank $\Longrightarrow X^TX$ full rank

linearly independent

$$X = \begin{bmatrix} 1 & \uparrow & \uparrow \\ 1 & & \\ \vdots & \vec{x}_1 & \vec{x}_2 \\ \vdots & & \\ 1 & \downarrow & \downarrow \end{bmatrix}$$

In general X is full rank $= p+1$

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

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

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

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

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

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

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

Get back $\vec{\hat{y}}$ for all n ordinal subjects in $\mathbb{D}$

$$\hat{y} = g(X) = X\vec{b} = \underbrace{X}_{n\times(p+1)} \underbrace{(X^T X)^{-1}}_{(p+1)(p+1)} \underbrace{X^T}_{(p+1)\times n} \vec{y}$$

H "Hat matrix"

$$= \boxed{\vec{\hat{y}} = H\vec{y}} = T(\vec{y}) \text{ linear function}$$

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

$$\text{rank} + \text{nullity} = n$$

$$\dim[\text{Colspace}] + \dim[\text{null space}] = n$$

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

degrees of freedom

$$\vec{y} = \underset{\text{prediction}}{\vec{\hat{y}}} + \underset{\text{error}}{\vec{e}}$$

$T(\vec{y})$ $\quad U(\vec{y})$ (function of nullspace)

Geometric Interpretation

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

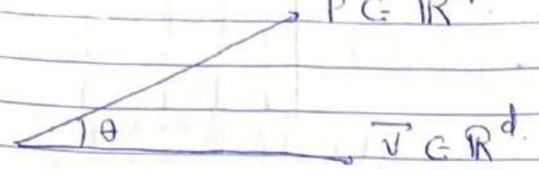
$$\vec{e} = \vec{y} - \vec{\hat{y}} = \vec{y} - H\vec{y} = (I-H)\vec{y}$$

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

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

$$RMSE = \sqrt{MSE}$$

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



$\vec{l} = \text{proj}_{\vec{v}}(\vec{a})$ "Orthonogal projection"

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

Recall the "law of cosines"

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

$$\Longrightarrow \|\vec{l}\| = \frac{\vec{a} \cdot \vec{v}}{\|\vec{v}\|}$$

$$\vec{l} = \|\vec{l}\| \frac{\vec{v}}{\|\vec{v}\|} = \frac{\vec{a} \cdot \vec{v} \, \vec{v}}{\|\vec{v}\|^2} = \overbrace{\left(\frac{\overbrace{\underbrace{\vec{v}}_{d \times 1} \underbrace{\vec{v}^T}_{1 \times d}}^{d \times d}}{\|\vec{v}\|^2} \right)}^{H} \vec{a}$$

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