Lecture 8 02-20-2020

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

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

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

Algorithm return one $\vec{w}$ and all $\hat{y}$'s can be compute via:

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

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

$\mathcal{A}$: OLS

$$\begin{aligned} SSE &:= \sum e_i^2 \\ &= \vec{e}^T \vec{e} \\ &= (\vec{y} - \vec{\hat{y}})^T (\vec{y} - \vec{\hat{y}}) \\ &= (\vec{y}^T - \vec{\hat{y}}^T)(\vec{y} - \vec{\hat{y}}) \\ &= \vec{y}^T \vec{y} - \vec{\hat{y}}^T \vec{y} - \vec{y}^T \vec{\hat{y}} + \vec{\hat{y}}^T \vec{\hat{y}} \\ &= \vec{y}^T \vec{y} - 2\vec{\hat{y}}^T \vec{y} + \vec{\hat{y}}^T \vec{\hat{y}} \end{aligned}$$

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

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

$$= SSE$$

$$\frac{\partial}{\partial \vec{w}}[SSE] := \begin{bmatrix} \frac{\partial}{\partial w_0}[SSE] \\ \frac{\partial}{\partial w_1}[SSE] \\ \frac{\partial}{\partial w_2}[SSE] \end{bmatrix} \stackrel{set}{=} \vec{0}_3$$

Back to Math 231 ...

Let $\vec{x} \in \mathbb{R}^n$, $\vec{a}$ column vector constant with respect to all x_j's.

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

$$\frac{\partial}{\partial \vec{x}}[\vec{a}^T \vec{x}] = \begin{bmatrix} \frac{\partial}{\partial x_1}[a_1x_1 + a_2x_2 + \dots + a_nx_n] \\ \\ \\ \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \vec{a}$$

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

$$\frac{\partial}{\partial \vec{x}}[af(\vec{x}) + bg(\vec{x})] = \begin{bmatrix} \frac{\partial}{\partial x_1}[af_1(\vec{x}) + bg_1(\vec{x}) \\ \\ \end{bmatrix}$$

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

Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix of constant with respect to x_j's.

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

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

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

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

$$= \frac{\partial}{\partial \vec{x}} [x_1 (a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n) + x_2 (a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n) + \cdots + x_n (a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n)]$$

$$= \frac{\partial}{\partial \vec{x}} \Big[a_{11}x_1^2 + a_{12}x_1x_2 + \dots + a_{1n}x_1x_n + a_{21}x_2x_1 + a_{22}x_2^2 + \dots + a_{2n}x_2x_n + a_{n1}x_nx_1 + a_{n2}x_nx_2 + \dots + a_{nn}x_n^2 \Big]$$

equal since $A = A^T$

$$\frac{\partial}{\partial x_1}\Big[\ \Big] = 2a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n + a_{21}x_2 + a_{31}x_3 + \dots + a_{n1}x_n$$

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

$$\frac{\partial}{\partial x_2}\Big[- - - \Big] = 2\vec{a}_2\vec{x}$$

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

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

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

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

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

$$\Rightarrow \vec{b} = (X^TX)^{-1}X^T\vec{y}$$ (OLS estimate valid for all p.)

Assum X is full rank $\Rightarrow X^TX$ full rank linearly independent

$$X = \begin{bmatrix} 1 & \uparrow & \uparrow & \uparrow \\ 1 & \vec{x}_1 & \vec{x}_2 & \vec{x}_3 \\ 1 & \downarrow & \downarrow & \downarrow \\ \vdots & & & \\ 1 & & & \end{bmatrix}$$

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

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

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

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

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

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

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

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

$$\vec{\hat{y}} = g(X) = X\vec{b} = \underbrace{\underset{n \times (p+1)}{X} \underset{(p+1)\times(p+1)}{(X^T X)^{-1}} \underset{(p+1)\times n}{X^T}}_{H \in \mathbb{R}^{n \times n} \text{ "Hat Matrix"}} \vec{y} = \boxed{\vec{\hat{y}} = H\vec{y}} = \underset{\text{linear function}}{T(\vec{y})}$$

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

$$\text{rank} = \dim[\text{colspace}], \qquad \dim[\text{colspace}] + \dim[\text{nullspace}] = n$$

$$\dim[\text{colspace}] = (p+1), \qquad (p+1) + (n-(p+1)) = n$$

$(p+1)$: "degree of freedom"

$$\vec{y} = \hat{\vec{y}} + \vec{e} \qquad \left[\vec{e} = \vec{y} - \hat{\vec{y}} = I\vec{y} - H\vec{y} = (I-H)\vec{y}\right]$$

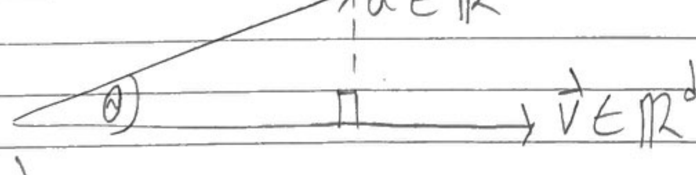
$$I\vec{y} = \underbrace{H\vec{y}}_{T(\vec{y})} + \underbrace{(I-H)\vec{y}}_{U(\vec{y})}$$

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

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

$$RMSE = \sqrt{MSE}$$

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



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

"orthogonal projection"

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

Recall the "law of cosines"

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

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

$$\vec{d} = \|\vec{d}\| \frac{\vec{v}}{\|\vec{v}\|} = \frac{\vec{a} \cdot \vec{v}\, \vec{v}}{\|\vec{v}\|^2} = \overbrace{\underbrace{\frac{\vec{v}\,\vec{v}^T}{\|\vec{v}\|^2}}_{H}}^{\substack{d \times d \\ d \times 1 \quad 1 \times d}} \vec{a}$$

$$= H \vec{a}$$

$$= \mathrm{Proj}_{\vec{v}}(\vec{a})$$