3/7/18

Lecture 10

Math 241

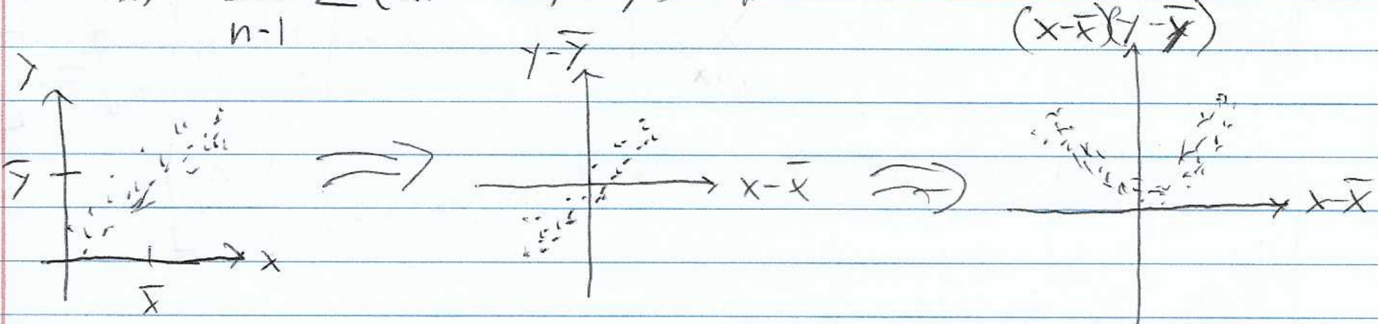
r.v.s X, Y were said to be "dependent" if knowing the value of one affects the distribution of the other:

$P(Y \mid X=x) \neq P(Y)$ or analogously with X

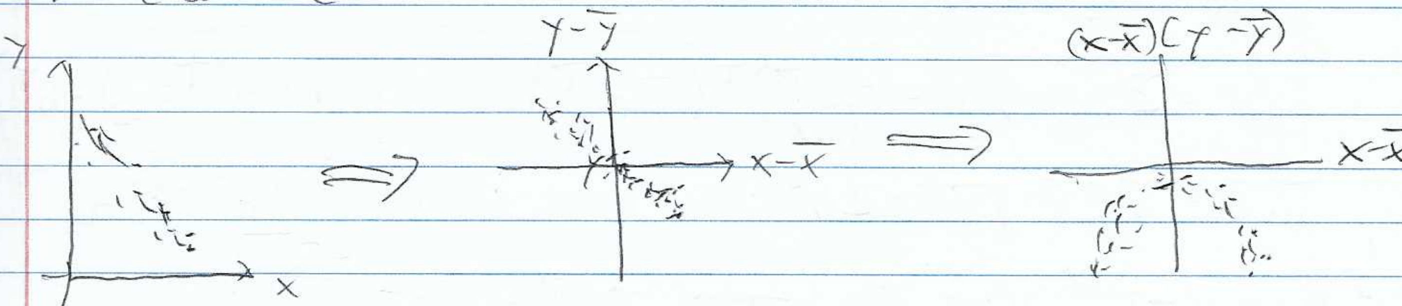
If knowing a prediction X's value allows to know "something" about y, then x, y are said to be "associated".

$\sigma_{xy} :=$ Covariance $Cov[X,Y] := E[(X-\mu_x)(Y-\mu_y)] \in \mathbb{R}$ estimated by

$S_{xy} := \frac{1}{n-1} \sum (x_i - \bar{x})(y_i - \bar{y}) \in \mathbb{R}$



Positive Covariance



Negative Covariance

(Proof Cauchy-Shwartz) interpretable scale

Correlation $\rho := Cor[X,Y] := \frac{Cov[X,Y]}{SE(X)SE(Y)} \in [-1,1]$ est. by $r = \frac{S_{xy}}{S_x S_y} \in [-1,1]$

We say X, Y are "pos. correlated" if $r > 0$ meaning $X\uparrow \Rightarrow Y\uparrow$

We say X, Y are "neg. correlated" if $r < 0$ meaning $X\uparrow \Rightarrow Y\downarrow$

if not correlated $r = 0$ meaning $X\uparrow \Rightarrow Y$

Correlation $\Rightarrow$ linear correlation

$\mathcal{Y} \subseteq \mathbb{R}$ regression, previously $p=1 \implies \mathcal{H} = \{w_0 + w_1 x : w_0 \in \mathbb{R}, w_1 \in \mathbb{R}\}$
and linear model

If $p=2 \implies \mathcal{H} = \{w_0 + w_1 x_1 + w_2 x_2 : \vec{w} \in \mathbb{R}^3\}$

$$SSE = \sum_{(\vec{x}_i, y_i) \in \mathbb{D}} (y_i - \hat{y}_i) = \sum \left(y_i - (w_0 + w_1 x_{i1} + w_2 x_{i2})\right)^2$$

Using $\mathcal{A} = L.S.$, I find... $\frac{\partial}{\partial w_0}[SSE] \overset{set}{=} 0$, $\frac{\partial}{\partial w_1}[SSE] \overset{set}{=} 0$, $\frac{\partial}{\partial w_2}[SSE] = 0$

$\mathbb{D} = \langle X, \vec{y} \rangle$

$$\begin{bmatrix} 1 & x_{11} & x_{12} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \end{bmatrix} \in \mathbb{R}^{n \times (\overset{2}{\overset{\parallel}{p}}+1)}$$

col of 1's

$$\underset{\in \mathbb{R}^{n \times 1}}{X\vec{w}} = \begin{bmatrix} w_0 + w_1 x_{11} + w_2 x_{12} \\ \vdots \\ w_0 + w_1 x_{n1} + w_2 x_{n2} \end{bmatrix} = \begin{bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_n \end{bmatrix} = \hat{\vec{y}} \implies \boxed{\hat{\vec{y}} = X\vec{w}}$$

$$SSE = \sum (y_i - \hat{y}_i)^2 = (\vec{y} - \hat{\vec{y}})^T (\vec{y} + \hat{\vec{y}})$$

(Note $(a+b)^T = a^T + b^T$)
(Note $a^T b = b^T a$)
(Note $(AB)^T = B^T A^T$)

$$\vec{v} \cdot \vec{v} = \vec{v}^T \vec{v} = \|\vec{v}\|^2 = \sum v_i^2$$

cont

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

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

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

$$\frac{\partial SSE}{\partial \vec{w}} := \begin{bmatrix} \frac{\partial}{\partial w_0}[SSE] \\ \frac{\partial}{\partial w_1}[SSE] \\ \vdots \\ \frac{\partial}{\partial w_p}[SSE] \end{bmatrix} = \vec{0}_{p+1}$$

e.g. $\frac{\partial}{\partial \vec{c}}[a] = \begin{bmatrix} \frac{\partial}{\partial c_1}[a] \\ \vdots \\ \frac{\partial}{\partial c_n}[a] \end{bmatrix} = \vec{0}_n$

$a \in \mathbb{R}$
$\vec{c} \in \mathbb{R}^n$

$$\text{eg}\quad \frac{\partial}{\partial \vec{c}}\left[\vec{c}^T\vec{a}\right] = \begin{bmatrix} \frac{\partial}{\partial c_1}\left[c_1a_1+\cdots+c_na_n\right] \\ \frac{\partial}{\partial c_2}\left[c_1a_1+c_2a_2+\cdots c_na_n\right] \\ \vdots \\ \frac{\partial}{\partial c_n}\left[c_1n_1+\cdots\cdots+c_na_n\right] \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

$\vec{a}\in\mathbb{R}^n$

quadratic form

$$\frac{\partial}{\partial \vec{c}}\overbrace{\left[\underbrace{\vec{c}^T A\vec{c}}_{\in\mathbb{R}}\right]}$$

$$\underset{\in\mathbb{R}^{n\times 1}}{A\vec{c}} = \begin{bmatrix} a_{11}c_1+a_{12}c_2+\cdots+a_{1n}c_n \\ a_{21}c_1+a_{22}c_2+\cdots+a_{2n}c_n \\ \vdots \\ a_{n1}c_1+a_{n2}c_2+\cdots+a_{nn}c_n \end{bmatrix}$$

$A\in\mathbb{R}^{n\times n}$ and symmetric

$$\vec{c}^T(A\vec{c}) = c_1(a_{11}c_1+a_{12}c_2+\cdots a_{1n}c_n) + c_2(a_{21}c_1+a_{22}c_2+\cdots+a_{2n}c_n) + \cdots + c_n(a_{n1}c_1+a_{n2}c_2+\cdots a_{nn}c_n)$$

$$\frac{\partial}{\partial \vec{c}}\left[\vec{c}^TA\vec{c}\right] = \begin{bmatrix} \frac{\partial}{\partial c_1}[\,] = 2c_1a_{11}+c_2a_{12}+\cdots+c_na_{1n} + c_2a_{21}+\cdots\cdots+c_na_{n1} = 2(c_1a_{11})+2c_2a_{12}+\cdots+2c_na_{n1} = 2(c_1a_{11}+\cdots\cdots+2_2a_{n1}) \\ \vdots \\ \frac{\partial}{\partial c_n}[\,] = c_1a_{1n}+c_2a_{2n}+\cdots+c_{n-1}a_{n-1,n}+2c_na_{nn} + c_1a_{n1}+c_2a_{n2}+\cdots+c_{n-1}a_{n,n-1} = 2c_1a_{1n}+2c_2a_{n2}+\cdots+2c_na_{nn} = 2(c_1a_{1n}+\cdots+c_na_{nn}) \end{bmatrix}$$

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

(Side note $\frac{\partial}{\partial x}\left[ax^2\right] = 2ax$)

Prove $A^T A$ symmetric $A \in \mathbb{R}$: $\Rightarrow A^T(A^T)^T \Rightarrow A^T A$ $\quad (A^T A)^T = A^T A$

$$\frac{\partial}{\partial \vec{w}}\left[\vec{y}^T\vec{y} - 2\vec{w}(X^T\vec{y}) + \vec{w}^T(X^TX)\vec{w}\right] = \vec{0}_{p+1} - 2X^T\vec{y} + 2X^TX\vec{w} \overset{set}{=} \vec{0}_{p+1}$$

$$\Rightarrow \underbrace{(X^TX)^{-1}X^TX}_{I_{p+1}}\vec{w} = (X^TX)^{-1}X^T\vec{y} \Rightarrow \boxed{\vec{b} = \underbrace{(X^TX)^{-1}}_{\text{only possible if } rank[X^TX] = p+1}X^T\vec{y}}$$

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad V = colsp(\vec{a}_{\cdot 1}, \vec{a}_{\cdot 2}) \quad Kernel(A) = 1 \quad rank(A) = 1$$

If $rank(x) < p+1$ (not full rank) $\Rightarrow$ theres a nullspace $\exists \vec{n} \neq 0 \in \mathbb{R}^{p+1}$
$X\vec{n} = \vec{0}_{p+1} \Rightarrow (X^TX)\vec{u} \Rightarrow X^T(X\vec{u}) = X^T\vec{0}_{p+1} = \vec{0}_{p+1} \Rightarrow X^TX$ not full rank

$$\hat{y}^* = g(\vec{x}^*) = \vec{x}^{*T}\vec{b} = b_0 + b_1x_1^* + \cdots + b_px^*_p \in Colsp[X]$$

$$\hat{\vec{y}} = X\vec{b} = \underbrace{X(X^TX)^{-1}X^T}_{H \text{ "hat matrix"}}\vec{y} = H\vec{y}$$