LECT 10 3/7/18

COVARIANCE:

X, Y were srd to be "depdt" if knowing the value of f.

$$p(Y \mid X = x) \neq p(Y)$$

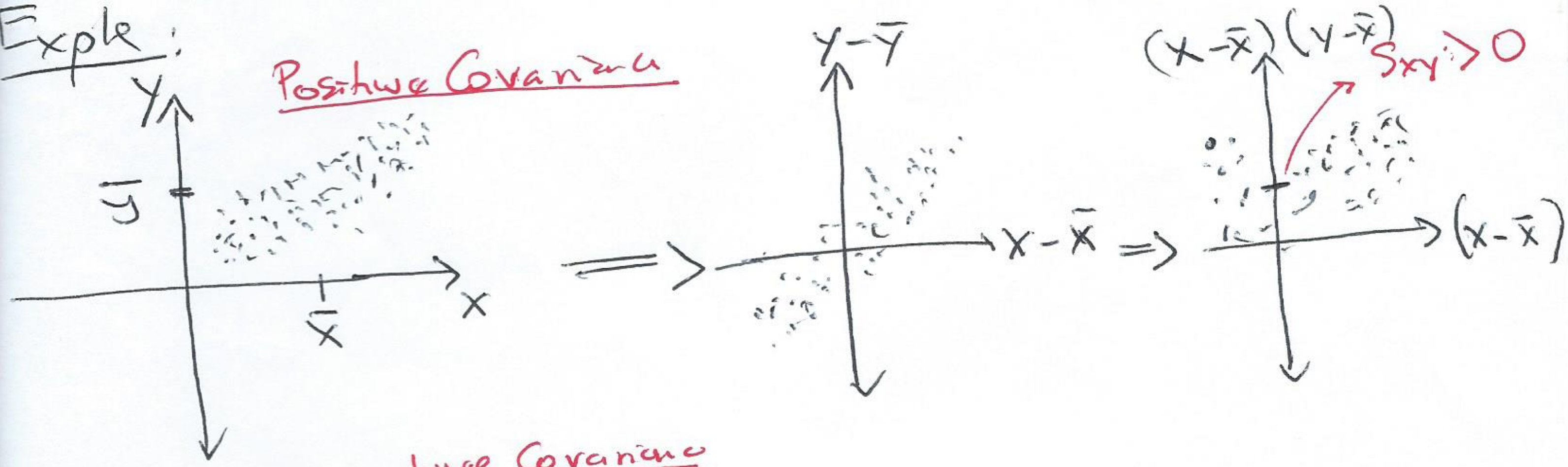
If knowing" pediction X"s value allows to know "something above Y, then X, y are said to be associated.

Covariance $Cor(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] \in \mathbb{R}$

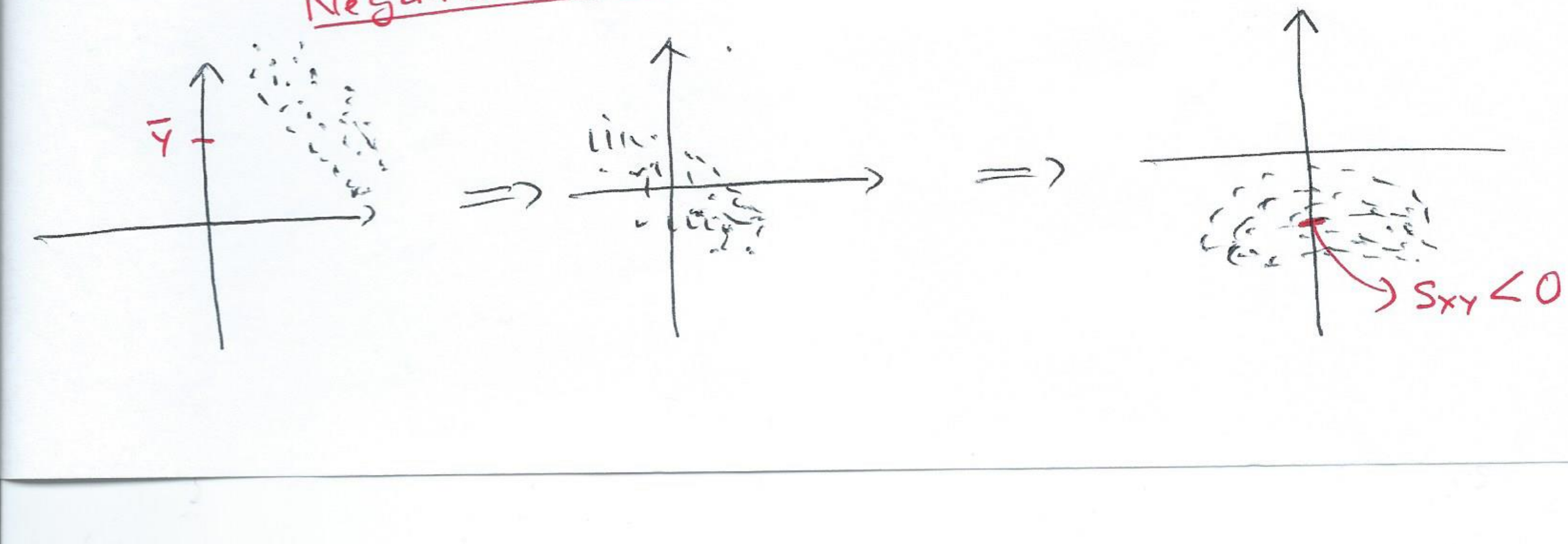
estimate by $S_{XY} = \frac{1}{n-1} \sum (x_i - \bar{x})(y_i - \bar{y}) \in \mathbb{R}$

Exple:

Positive Covariance



Negative Covariance



CORRELATION: $\rho = Corr[X,Y] = \dfrac{Corr[X,Y]}{SE(X)SE(Y)}$

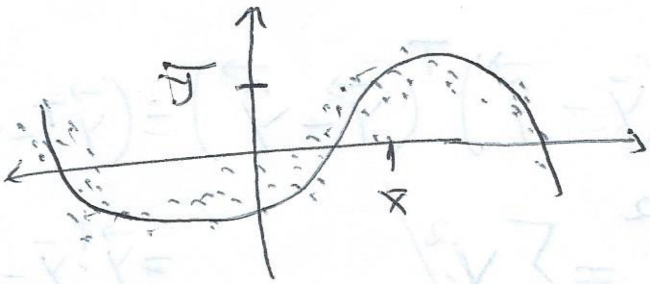
$\dfrac{Corr[X,Y]}{SE(X)SE(Y)} \in [-1,1] \leftarrow$ proof. $\qquad$ e .st by $r = \dfrac{S_{XY}}{S_X S_Y} \in [-1,1]$

We say X, Y are "pass combine" if $r > 0 \Rightarrow X\nearrow \Rightarrow Y\nearrow$

neg combine if $r < 0 \Rightarrow X\nearrow \Rightarrow Y\searrow$

not combine if $r = 0 \Rightarrow X\nearrow \Rightarrow Y$

Hence ~~Corr~~ $\Rightarrow$ "linear Correlation"



NEW UNIT

$y \subseteq \mathbb{R}$ regress, pre, $p = 1 \Rightarrow \mathcal{H}$

$\mathcal{H} = \{w_0 + w_1 x : w_0 \in \mathbb{R}, w_1 \in \mathbb{R}\}$ and linear model.

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

$SSE = \sum (y_i - \hat{Y}_i) = \sum \left(y_i - (w_0 + w_1 x_{i1} + w_2 x_{i2})\right)^2$

using $\mathcal{A} = L.S$, I find ....

$\dfrac{\partial}{\partial w_0}[SSE] = 0 \; ; \; \dfrac{\partial}{\partial w_1}[SSE] = 0 \; ; \; \dfrac{\partial}{\partial w_2}[SSE] = 0$

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

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \end{bmatrix} \in \mathbb{R}^{b \times (p+1)}$$

Col of 1's

$X_{\vec{w}} =$

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

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

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

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

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

Note: $(a+b)^T = a^T + b^T$

$a^T \cdot b = b^T \cdot a$

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

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

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

$$\begin{bmatrix} \frac{\partial}{\partial w_0}[SSE] \\ \frac{\partial}{\partial w_1}[SSE] \\ \frac{\partial}{\partial w_2}[SSE] \\ \vdots \\ \frac{\partial}{\partial w_p}[SSE] \end{bmatrix} \overset{set}{=} \vec{0}_{p+1}$$

Called $\frac{\partial SSE}{\partial \vec{w}}$

Exple 1

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

$a \in \mathbb{R}$

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

Exple 2

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

$\vec{a} \in \mathbb{R}^2$

Exple 3

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

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

Exple 4

$$\frac{\partial}{\partial \vec{c}}\left[\underbrace{\vec{c}^T A \vec{c}}_{\text{quadratic formula}}\right]$$

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

$$A\vec{C} = \begin{bmatrix} a_{11}c_1 + a_{12}c_2 + \cdots + a_{1n}c_y \\ a_{21}c_1 + a_{22}c_2 + \cdots + a_{2n}c_y \\ \vdots \\ a_{n1}c_n + \cdots + a_{nn}c_y \end{bmatrix}, \quad A\vec{C} \in \mathbb{R}^{n\times y}$$

$$\vec{c}\,(A\vec{C}) = \begin{matrix} +c_1(a_{11}c_1 + a_{12}c_2 + \cdots + a_{1n}c_i) \\ +c_2(\qquad\qquad\qquad\qquad) \\ \vdots \\ +c_n(a_{n1}c_1 + a_{n2}c_2 + \cdots + a_{nn}c_y) \end{matrix}$$

$$\frac{\partial}{\partial \vec{c}}\left[\vec{c}^{\,T} A \vec{c}\right] = \begin{bmatrix} \frac{\partial}{\partial c_1}[\;] = 2c_1 a_{11} + c_2 a_{12} + \cdots + c_n a_{1n} + c_2 a_{22} + \cdots + c_2 a_{n\ldots} \\ = 2c_1 a_{11} + 2c_2 a_{12} + \cdots + 2c_c \\ \vdots \\ \frac{\partial}{\partial c_n}[\;] = c_1 a_{1n} + c_2 a_{2n} + \cdots + c_n \\ = 2c_1 a_{1n} + 2c_n a_n + \cdots \end{bmatrix}$$

$$= 2\begin{bmatrix} \vec{a}_1^{\,T} c \\ \vec{a}_2^{\,T} c \\ \vdots \\ \vec{a}_n^{\,T} c \end{bmatrix} = 2A\vec{c}$$

$$\boxed{\frac{\partial}{\partial \vec{c}}\left[\vec{c}^{\,T} A \vec{C}\right] = 2A\vec{c}}$$

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

$$= \vec{0}_{p+1} - 2X^T Y + 2X^T X\vec{w} = \vec{0}_{p+1}$$

$$\Rightarrow X^T X\vec{w} = X^T\vec{Y}$$

$$(X^T X)X^T X\vec{w} = (X^T X)^{-1} X^T \vec{Y}$$

$$\underbrace{(X^TX)^{-1}(X^TX)}_{I_{p+1}}\vec{w} = (X^TX)^{-1}X^T\vec{y} \Rightarrow \boxed{\vec{b} = (X^TY)^{-1}X^TY}$$

only possible ~~x~~ if $\text{rank}(X^TY) = p+1$

$\Updownarrow$

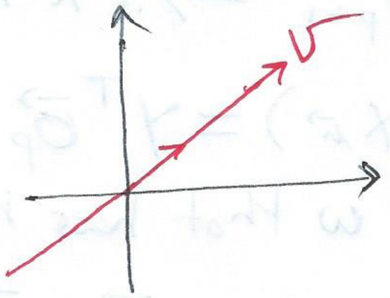
$\text{rank}[X] = p+1$

<u>Exple</u>: $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$

$V = \text{colop}(\vec{a}_1, \vec{a}_2)$

$\text{Kernel}(A) = 1$

$\text{rank}(A) = 1$



X

Full rank means is fully ~~peple~~ p+1 dimension.

<u>Exple</u>: Salary in year 2000 Sal year 2001 ~~Sal~~ Total ~~sa~~ 200

$X =$

<u>Exple</u>: Salary in yr 2000 height in ft height in presumy

$X =$

not full rank.

$\mathrm{rank}(X^T Y) = p+1$

$\Uparrow$

$\mathrm{rank}[X] = p+1$

$\Downarrow$

If $\mathrm{ran}(X) < p+1$ (not full rank)

$\Rightarrow$ a null space $\exists \vec{u} \neq 0 \in \mathbb{R}^{p+1}$

$X\vec{u} = \vec{0}_{p+1} \Rightarrow (X^T X)\vec{n}$

$\Rightarrow X^T(X\vec{u}) = Y^T \vec{0}_{p+1} = \vec{0} \Rightarrow X^T X$ not full rank.

We solve the ω that has the lowest.

$$\hat{y}^* = g(\bar{x}^*) = \vec{x}^* \cdot \vec{b}$$

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

So $p+1 < n$

Linear models look like: $b_0 + b_1 x_1^* + \cdots + b_p x_p^*$