MON NOV 6

PCA REVIEW

HW DUE TOMORROW DISC

LINEAR MODELS

$$\underset{n\times 1}{Z} = \begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix}$$

$$\operatorname{cov}(Z)_{i,j} = \operatorname{cov}(Z_i, Z_j) = \mathbb{E}\left[(Z_i - \mathbb{E}Z_i)(Z_j - \mathbb{E}Z_j)\right]$$

$\underset{n\times 1}{Z}$ is a random vector A is $k\times n$ matrix

$$\operatorname{cov}(\underbrace{\underset{k\times n}{A}\ \underset{n\times 1}{Z}}_{k\times 1 \text{ random vector}}) = \underset{k\times n}{A}\ \underbrace{\operatorname{cov}(Z)}_{n\times n}\ \underset{n\times k}{A^T}$$

$$Z = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_5 \end{bmatrix} \qquad Z_1, Z_2, Z_3, Z_4, Z_5 \text{ independ } N(0,1)$$

$$\Sigma = \operatorname{cov}(Z)_{i,j} = 0 \qquad i \neq j$$

$$\operatorname{cov}(AZ) = A\,\Sigma\,A^T$$

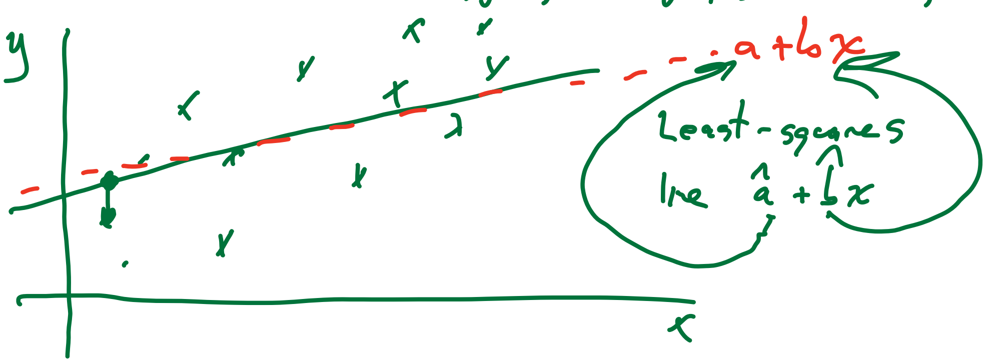
$$\underset{\sim}{X} = \begin{bmatrix} X_{11} & \cdots & X_{15} \\ X_{21} & & X_{25} \\ & & \\ X_{r1} & & X_{r5} \end{bmatrix}$$

LINEAR MODELS

(X, Y) HAS SOME JOINT DISTRIBUTION

BEFORE: MOTIVATION

BIVARIATE DATA $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$



$r_i = y_i - a + bx_i$

least-squares line: pick a, b to minimize $\sum_{i=1}^{n} r_i^2 = f(a,b)$

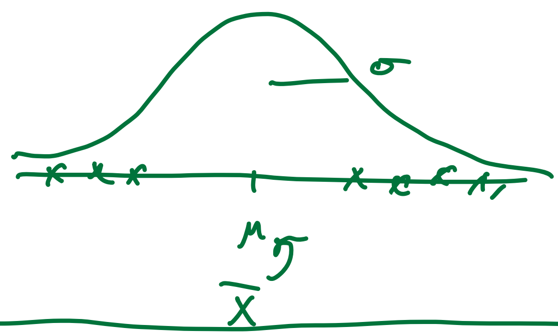
least-square line is just data summary

ONE MODEL FOR DATA a, b unknown to us

$$Y_i = a + bx_i + \varepsilon_i \qquad \varepsilon_1, \varepsilon_2, --- \text{ IID } N(0, \sigma^2)$$

unobserved

Goal: use the data to estimate a, b



Random variable world (X, Y) some joint distribution

$(Y = \alpha + \beta X + \varepsilon)$

Find $\hat{Y} = a + bX$ so that $\hat{Y}$ is a good guess at Y

$$E(Y - \hat{Y})^2 = MSE$$

we ~~What~~ should (design $\hat{Y}$) so that $E\hat{Y} = EY$

Suppose $\tilde{Y}$ is an estimator of Y

$$\boxed{E\tilde{Y} - EY = \beta} \neq 0$$

let $\hat{Y} = \tilde{Y} - \beta$ $E\hat{Y} = E\tilde{Y} - \beta = -(E\tilde{Y} - EY)$

$\hat{Y} - \tilde{Y} = -\beta$ $= EY$

CLAIM: $\mathrm{MSE}(\hat{Y}) \le \mathrm{MSE}(\tilde{Y})$

$$\begin{aligned}
\mathrm{MSE}(\hat{Y}) &= \mathbb{E}(\hat{Y} - Y)^2 \\
&= \mathbb{E}[(\hat{Y} - \tilde{Y} + \tilde{Y} - Y)^2] \\
&= \mathbb{E}[(\hat{Y} - \tilde{Y})^2 + 2(\hat{Y} - \tilde{Y})(\tilde{Y} - Y) + (\tilde{Y} - Y)^2] \\
&= \mathbb{E}[(\hat{Y} - \tilde{Y})^2] + 2\mathbb{E}[(\hat{Y} - \tilde{Y})(\tilde{Y} - Y)] \\
&\quad + \mathbb{E}[(\tilde{Y} - Y)^2] \\
&= \beta^2 - 2\beta\,\mathbb{E}[\tilde{Y} - Y] + \mathrm{MSE}(\tilde{Y}) \\
&\quad - 2\beta(\mathbb{E}[\tilde{Y}] - \mathbb{E}[Y]) \\
&= \beta^2 - 2\beta^2 + \mathrm{MSE}(\tilde{Y})
\end{aligned}$$

$$\mathrm{MSE}(\hat{Y}) = -\beta^2 + \mathrm{MSE}(\tilde{Y})$$

$$\Rightarrow \mathrm{MSE}(\hat{Y}) \le \mathrm{MSE}(\tilde{Y})$$

Moral: we should always take $\hat{Y}$ so that ~~...~~ $\mathbb{E}\hat{Y} = \mathbb{E}Y$

without loss of generality assume $\mathbb{E}X = \mathbb{E}Y = 0$
(Ex: Show that general case follows)

$$\hat{Y} = aX + b \qquad \mathbb{E}\hat{Y} = \mathbb{E}Y = 0$$

$$\begin{aligned}
\mathbb{E}\hat{Y} &= a\,\mathbb{E}(X) + b \\
&= b
\end{aligned}$$

$$\Rightarrow b = 0$$

$$\hat{Y} = aX$$

Question: How to choose a?

$$MSE(\hat{Y}) = \mathbb{E}[(aX - Y)^2]$$
$$= \mathbb{E}[a^2X^2 - 2aXY + Y^2]$$
$$= a^2\,\mathbb{E}X^2 - 2a\,\mathbb{E}[XY] + \mathbb{E}Y^2$$
$$= a^2\,Var(X) - 2a\,Cov(X,Y) + Var(Y)$$

minimize wrt a

$$a = \frac{SD(Y)}{SD(X)}\,COR(X,Y) \quad \frac{Cov(X,Y)}{SD(X)\,SD(Y)}$$

Empirical version of this: $(x_1,y_1), \ldots, (x_n,y_n)$

It should not surprise you that the least-squares line has

sample SD of $y_1, \ldots, y_n$

$$\hat{a} = \frac{s_y}{s_x}\,\rho_{xy}$$

sample SD of $x_1, \ldots, x_n$

sample correlation

$$\hat{b} = \bar{y} - a\bar{x}$$

$$\frac{\frac{1}{n-1}\sum_i (x_i - \bar{x})(y_i - \bar{y})^2}{SD_x \quad SD_y}$$

$\hat{y}_i$ = fitted value

$= \hat{a} + \hat{b}x_i$

ρ_{xy} = strength linear relationship

$\rho_{xy} = \pm 1$ if data on line

$|\rho_{xy}| < 1$