Announcements:

- All HW (from HW 3 on) pushed back 1 week
- We may or may not have HW 6, TBD.

Multiple Linear Regression Model

There are some variations on how to specify the MLR model (e.g. index order, # of predictors as p or p−1).

In order to all be on the same page, let's all use:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_{p-1} X_{(p-1)i} + \varepsilon_i \qquad i = 1, \ldots, n$$

$$\Updownarrow$$

$$\boxed{Y = \overbrace{X\beta}^{\text{const.}} + \overbrace{\varepsilon}^{\text{random}}}$$

where

$$\underset{n \times p}{X} = \underbrace{\begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{(p-1)1} \\ 1 & X_{12} & X_{22} & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{(p-1)n} \end{bmatrix}}_{\text{design}} \quad \& \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}$$

LS EST:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}}\, Q(\beta) = \underset{\beta}{\operatorname{argmin}}\, (Y - X\beta)^T (Y - X\beta)$$

$$= (X^T X)^{-1} X^T Y \quad \text{if } X \text{ has full rank}$$

Fitted Values:

"hat matrix"

$$\hat{Y} = X\hat{\beta} = \underline{X(X^T X)^{-1} X^T} Y = HY$$

Residuals:

$$e = Y - \hat{Y} = Y - HY = (I - H)Y$$

Sum of Squared Error:

$$SSE = \sum_{i=1}^{n} e_i^2 = e^T e = \|e\|^2$$

random vector:

$$W \sim N(\mu, \Sigma)$$

$$E(AW) = AE(W) = A\mu$$

$$Var(AW) = A\, Var(W)\, A^T = A \Sigma A^T$$

$$\boxed{AW \sim N(A\mu, A\Sigma A^T)}$$

$\longrightarrow \mathrm{Var}(W) = E\big[\underset{n\times 1}{(W-\mu)}\underset{1\times n}{(W-\mu)^T}\big]$

$$
\begin{aligned}
\mathrm{Var}(AW) &= E\left((AW - A\mu)(AW - A\mu)^T\right) \\
&= E\left(A(W-\mu)(A(W-\mu))^T\right) \\
&= E\left(A(W-\mu)(W-\mu)^T A^T\right) \\
&= A\,E\left[(W-\mu)(W-\mu)^T\right]A^T \\
&= A\,\mathrm{Var}(W)A^T
\end{aligned}
$$

Distribution of Quantities: Why is $E(Y) =$

$$Y \sim N(\overset{\mu}{\overset{\parallel}{X\beta}}, \sigma^2 I_n)$$

$$\hat{\beta} = \underbrace{(X^TX)^{-1}X^T}_{A} Y \sim N\Big(\underbrace{\underbrace{(X^TX)^{-1}X^T}_{A}\underbrace{X\beta}_{\mu}}_{=}, (X^TX)^{-1}X^T \sigma^2 I \left((X^TX)^{-1}X^T\right)^T\Big)$$

$$\left((X^TX)^{-1}X^T\right)^T = \left(X\left((X^TX)^{-1}\right)^T\right)$$

$$N\left((X^TX)^{-1}(X^TX)\beta, \sigma^2 \cancel{(X^TX)^{-1}} \cancel{X^T} X \left[(X^TX)^{-1}\right]^T\right)$$

$$N\left(\beta, \sigma^2\left((X^TX)^{-1}\right)^T\right)$$

$$= N\left(\beta, \sigma^2\left((X^TX)^T\right)^{-1}\right)$$

$$= N\left(\beta, \sigma^2(X^TX)^{-1}\right)$$

$$AA^{-1} = I$$

$$(AA^{-1})^T = I$$

$$\Rightarrow (A^{-1})^T A^T = I$$

$$\Rightarrow (A^{-1})^T \cancel{(A^T)(A^T)^{-1}} = (A^T)^{-1}$$

if A^{-1} exists, u can swap ^{T} & $^{-1}$

How about $\hat{Y}$'s distribution?

$$Y \sim N(X\beta, \sigma^2 I)$$

$$\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\bar{X}\sigma^2}{SSX}$$

$$\hat{Y} = HY \sim N(\underbrace{HX\beta}_{E(\hat{Y})}, \underbrace{H\sigma^2 I H^T}_{Var(\hat{Y})})$$

$$E(\hat{Y}) = HX\beta = X\cancel{(X^TX)^{-1}}\,\cancel{X^T}X\beta$$

$$= X\beta$$

$$Var(\hat{Y}) = \sigma^2 HH^T \overset{?}{=} \sigma^2 H$$

Properties of H:

Any matrix A is a projection matrix if:

1. A is symmetric
2. A is idempotent

claim: H is a projection matrix

$$(ABC)^T = C^T B^T A^T$$

① $H^T = \left(X (X^TX)^{-1} X^T \right)^T = (X) \left((X^TX)^{-1} \right)^T X^T =$

$= X \left((X^TX)^T \right)^{-1} X^T$

$= X \ (X^TX)^{-1} X^T = H.$

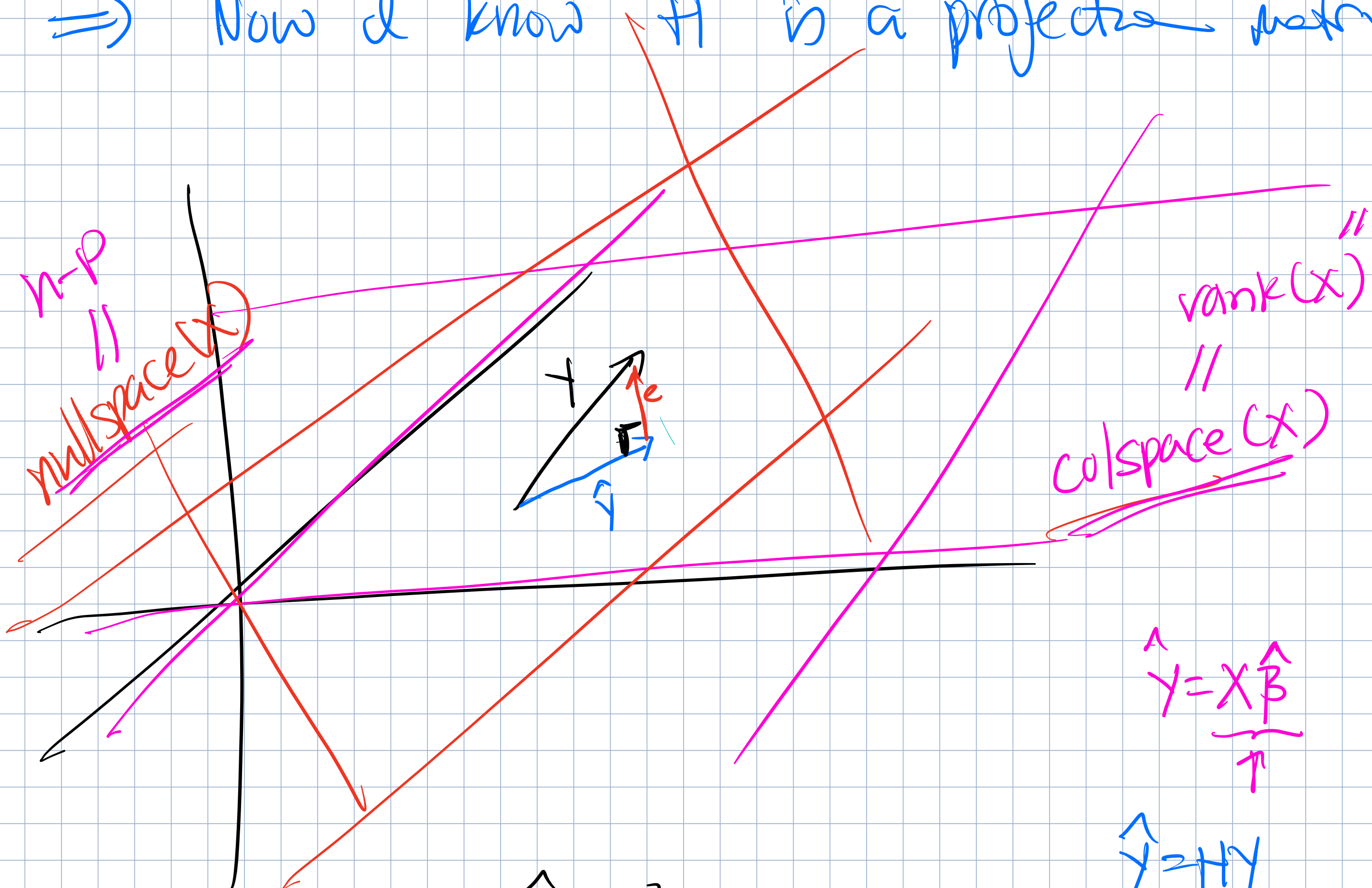
② idempotent $\Leftrightarrow A^2 = A$

can we show H is idempotent?

$H^2 = H \cdot H = X(X^TX)^{-1} \cancel{X^T X (X^TX)^{-1}} X^T$

$= X(X^TX)^{-1}X^T = H.$

$\Rightarrow$ Now I know H is a projection matrix.

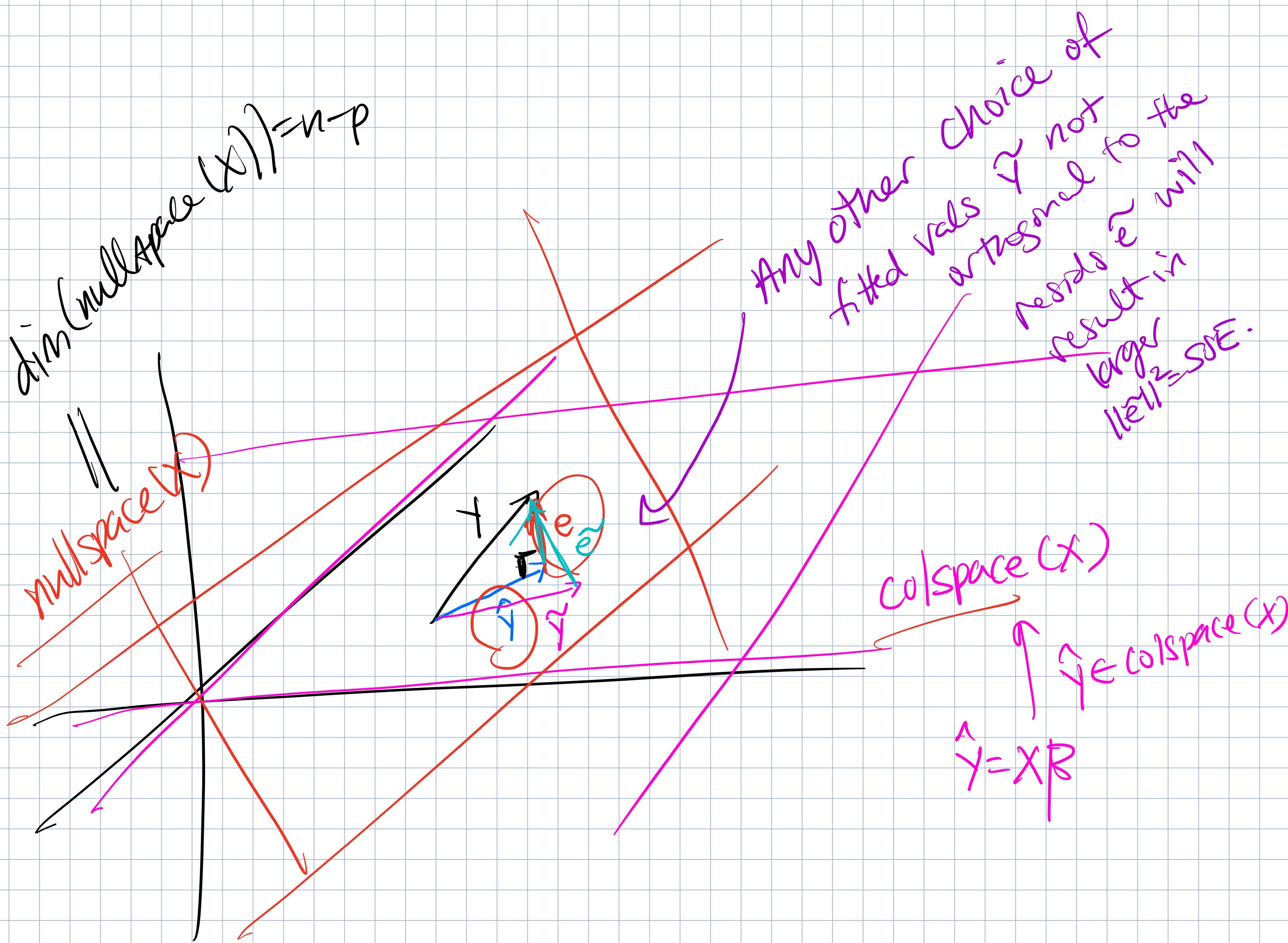


$\hat{Y} = X\hat{\beta}$

$\hat{Y} = HY$

$\langle e, \hat{Y} \rangle = \sum_i e_i \hat{Y}_i = 0$

$$\langle e, \hat{Y} \rangle = e^T \hat{Y} = ((I-H)Y)^T (HY)$$
$$= Y^T (I-H)^T H Y$$
$$= Y^T (I-H) H Y$$
$$= Y^T (H - H^2) Y$$
$$= Y^T (H-H) Y = Y^T 0 \cdot Y = 0$$



H was a projection matrix

I-H is also a projection ".

Exercise

How can we estimate σ^2?

$$\text{MSE} = \hat{\sigma}^2 = \frac{SSE}{n-p} = \frac{e^Te}{n-p}.$$

Why is this a good estimator?

We want to know something about the dist of $\|e\|^2 = e^Te$.

$$Y \sim N(X\beta, \sigma^2 I_n)$$

$$\underline{(I-H)}\, Y = Y - \hat{Y} = \underline{e} \sim N\Big(\underbrace{(I-H)X\beta}_{=\,0},\ \underline{\sigma^2(I-H)}\Big)$$

AIA^T

$(I-H)X\beta$

$= (X - HX)\beta$

$$\boxed{e \sim N(0, \sigma^2(I-H))}$$

$$= (X - X\cancel{(X^TX)^{-1}X^TX})\beta$$

$$= (X - X)\beta = 0$$

$$\text{Var}((I-H)Y) = (I-H)\text{Var}(Y)(I-H)^T$$

$$= (I-H)\sigma^2 I(I-H)$$

$$= \sigma^2(I-H)(I-H)$$

$$= \sigma^2(I - IH - HI + H^2)$$

$$= \sigma^2(I - H - \cancel{H} + \cancel{H})$$

What about the distribution of e^Te?

$$Y \sim N(X\beta, \sigma^2 I)$$

$$Y - X\beta \sim N(0, \sigma^2 I)$$

$$\frac{Y - X\beta}{\sigma} \sim \underline{N(0, I)}$$

$$\frac{(I-H)(Y-X\beta)}{\sigma} \sim N(0, \underline{I-H})$$

$$\left\| \frac{(I-H)(Y-X\beta)}{\sigma^2} \right\|^2 =$$

$$\frac{(Y-X\beta)^T (I-H)^T (I-H)(Y-X\beta)}{\sigma^2}$$

quadratic form of Y.

$$\frac{(Y-X\beta)^T (I-H)(Y-X\beta)}{\sigma^2} = \frac{Y^T(I-H)Y}{\sigma^2} = \frac{e^Te}{\sigma^2} = \frac{SSE}{\sigma^2}$$

$$+ \cancel{\frac{Y^T(I-H)X\beta}{\sigma^2}} \to 0$$

$$\boxed{\frac{Y^T(I-H)Y}{\sigma^2} \sim \chi^2_{\mathrm{rank}(I-H) = n-p}}$$

trace is cyclic:

$tr(ABC) = tr(BCA)$

$= tr(CAB)$

$= tr(ABC)$

For idempotent matrices,

$$\mathrm{rank}(I-H) = \mathrm{trace}(I-H)$$

$$= \mathrm{trace}(I) - tr(H)$$

$$= n - tr(X(X^TX)^{-1}X^T)$$

$$= n - tr(\underbrace{(X^TX)^{-1}X^TX}_{p\times n\,n\times p})$$

$$= n - tr(I_p)$$

$= n-p$

Idempotent matrices have very special eigenvals.

They're always 0 or 1.

$H = \underline{U \Lambda U^T}$ ← eigendecomp where Λ holds eigenvals, U holds the eigenvectors

$\|$

$H^2 = (U\Lambda U^T)(U\Lambda U^T)$

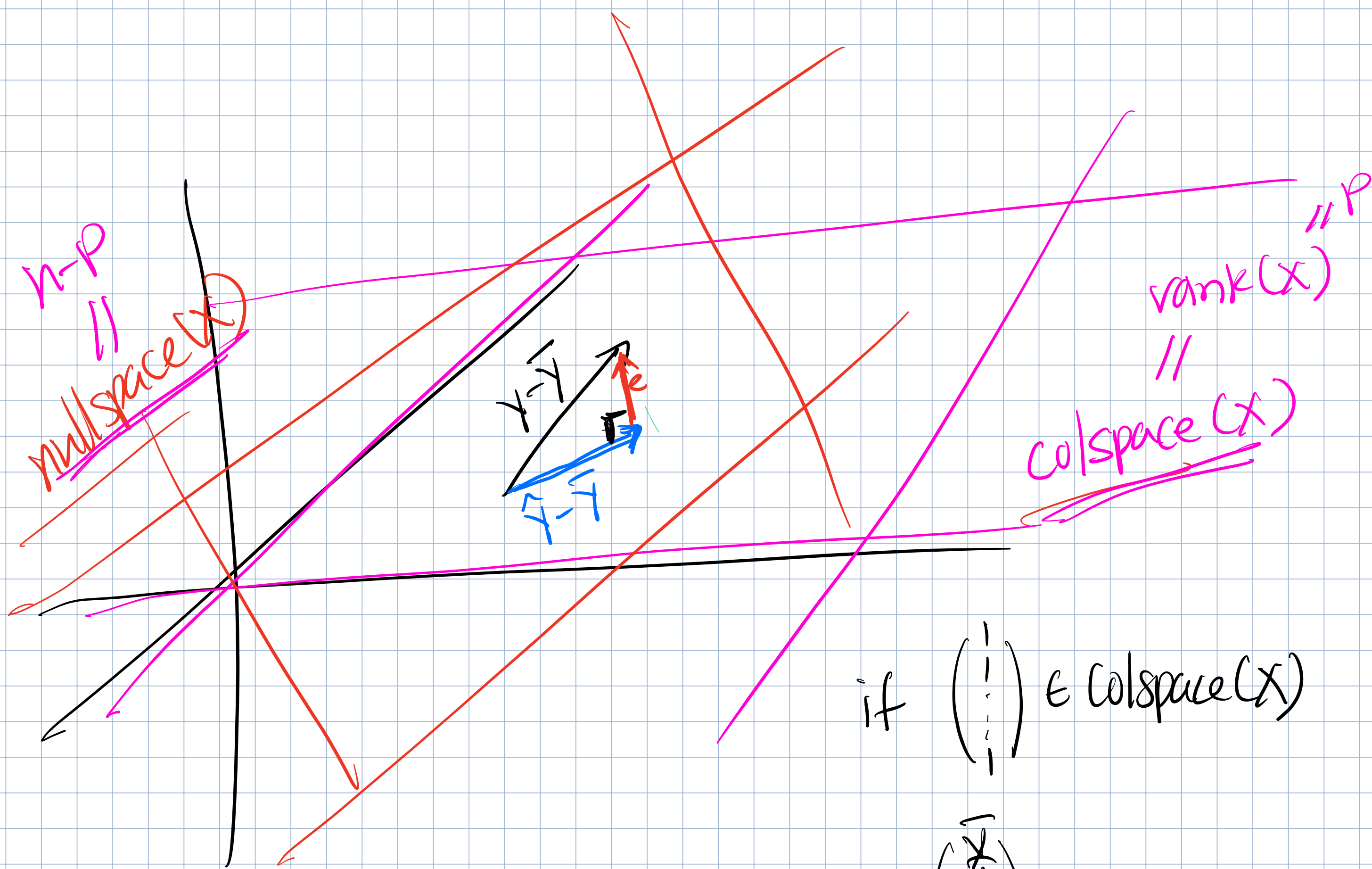
$= U\Lambda \cancel{U^T U} \Lambda U^T$

$= \underline{U\Lambda^2 U^T}$ $\Rightarrow$ $\boxed{\Lambda^2 = \Lambda}$

$\lambda_i^2 = \lambda_i$

$\Rightarrow \lambda_i = 0$ or 1

SS Decomp $\longleftrightarrow$ Projection



if $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in$ colspace(X)

$\begin{pmatrix} \bar{Y} \\ \vdots \\ \bar{Y} \end{pmatrix}$

$(Y-\bar{Y}) - (\hat{Y}-\bar{Y}) = Y - \cancel{\bar{Y}} - \hat{Y} + \cancel{\bar{Y}} = e$

Pyth. Theorem:

$$\|Y-\bar{Y}\|^2 = \|Y-\hat{Y}\|^2 + \|\hat{Y}-\bar{Y}\|^2$$

$$SST = SSE + SSR$$