Lecture 09   3/1/21

$\vec{b} = (X^\top X)^{-1} X^\top \vec{y}$, the OLS linear model, $\hat{\vec{y}} = X\vec{b}$,

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

What if we have no features? i.e. the null model case. Is this an OLS solution?

$$X = [\vec{1}_n] = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$\vec{b} = b_0 = (\underbrace{X^\top X}_{\underbrace{\underbrace{[1 \cdots 1]\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}}_{n}}_{\frac{1}{n}}})^{-1} \underbrace{X^\top \vec{y}}_{\underbrace{[1 \cdots 1]\vec{y}}_{\sum y_i}} = \frac{\sum y_i}{n} = \bar{y} = g_0$$

$\mathbb{R}^2$ (of null)

span$[\vec{v}]$ is a 1-dim subspace of $\mathbb{R}^2$

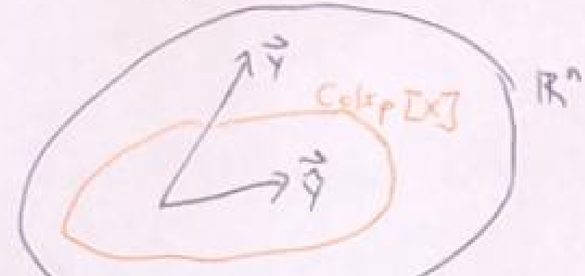
$$\underset{p+1}{\operatorname{rank}}[X] = \dim[\operatorname{colsp}[X]]$$

$$\operatorname{colsp}[X] := \operatorname{span}[\vec{1}, \vec{x}_{\cdot 1}, \ldots, \vec{x}_{\cdot p}] := \quad \vec{w} \in \mathbb{R}^{p+1}$$

$$= \left\{ w_0 \vec{1}_n + w_1 \vec{x}_{\cdot 1} + \ldots + w_p \vec{x}_{\cdot p} : w_0, w_1, \ldots, w_p \in \mathbb{R} \right\}$$

p+1 dimensional subspace of the entire n-dimensional "full space" (the number of dimensions of y which is n, the number of rows of X.

$\vec{y} \in \text{Colsp}[X]$? YES

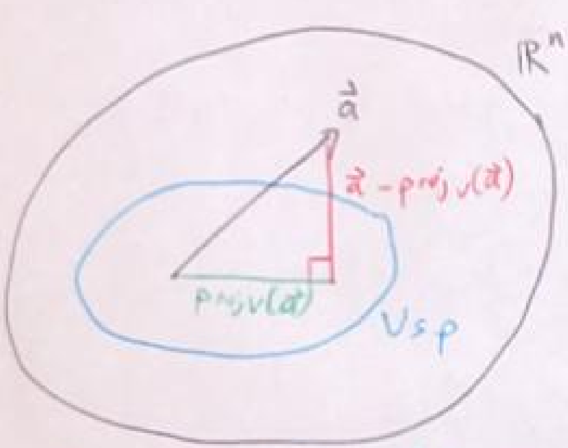


$$\hat{\vec{y}} = X\vec{b} \overset{\text{OLS solution}}{=} \underbrace{X(X^\top X)^{-1}X^\top}_{H} \vec{y} = H\vec{y}$$

$H \in \mathbb{R}^{n \times n}$

H for "hat" matrix, the linear operator turning y-vec into y-hat-vec.

$$X\vec{b} \in \text{colsp}[X]$$
$$\Updownarrow$$
$$H\vec{y} \in \text{colsp}[X] \Rightarrow \text{rank}[H] = p+1 \Rightarrow H \text{ is not invertible}$$



V is a K-dim subspace of the n-dim full space.

We want to "project" a-vec onto V such that the difference between a-vec and its projection is ~~perpendicular~~ perpendicular. This is called an "orthogonal projection." We want a formula for this projection as a function of the space V.

$Vsp = \text{span}\{\vec{v}_1, \ldots, \vec{v}_K\}, K < n$

$\text{proj}_V(\vec{a}) \in \text{span}\{\vec{v}_1, \ldots, \vec{v}_K\} \Rightarrow \exists \vec{w}$

$\text{proj}_V(\vec{a}) = w_1\vec{v}_1 + \ldots + w_K\vec{v}_K = V\vec{w}$

Such that $V = [\vec{v}_1 | \ldots | \vec{v}_K]$, $\vec{w} \in \mathbb{R}^K$

due to orthogonal constraint, $\vec{a} - \text{proj}_V(\vec{a}) \perp \vec{v}_j \; \forall j$

$\Rightarrow (\vec{a} - V\vec{w})^\top \vec{v}_j = 0 \;\; \forall j \Leftrightarrow \vec{v}_j^\top(\vec{a} - V\vec{w}) = 0 \;\; \forall j$

$$\Rightarrow \begin{aligned} \vec{v}_1^\top(\vec{a} - V\vec{w}) &= 0 \\ \vec{v}_2^\top(\vec{a} - V\vec{w}) &= 0 \\ &\vdots \\ \vec{v}_K^\top(\vec{a} - V\vec{w}) &= 0 \end{aligned} \Rightarrow \begin{bmatrix} \vec{v}_1^\top \\ \vec{v}_2^\top \\ \vdots \\ \vec{v}_K^\top \end{bmatrix} (\vec{a} - V\vec{w}) = \vec{0}_K \Rightarrow V^\top(\vec{a} - V\vec{w}) = \vec{0}_K$$

$\Rightarrow V^T\vec{a} - V^TV\vec{\omega} = \vec{0}_K \Rightarrow (V^TV)^{-1}V^TV\vec{\omega} = (V^TV)^{-1}V^T\vec{a} \Rightarrow \vec{\omega} = (V^TV)^{-1}V^T\vec{a}$

$\text{proj}_V(\vec{a}) = V\vec{\omega} = \underbrace{V(V^TV)^{-1}V^T}_{H}\vec{a} = H\vec{a}$

orthogonal projection onto colsp[V]

We call the $n \times n$ matrix H, the orthogonal projection matrix onto the subspace $V_{sp} = \text{colsp}[V]$.

$H = X(X^TX)^{-1}X^T$ is the orthogonal projection matrix onto colsp[X].

Properties that define orthogonal projection matrices, H

(1) H is symmetric, $H^T = H$

$H^T = \left(V(V^TV)^{-1}V^T\right)^T = V^{TT}\left((V^TV)^{-1}\right)^T V^T = V\left((V^TV)^{-1}\right)^T V^T = V\left((\overbrace{V^TV}^{V^TV})^{-1}\right)^T V^T = H$ ✓

Let A be square, invertible, and symmetric.

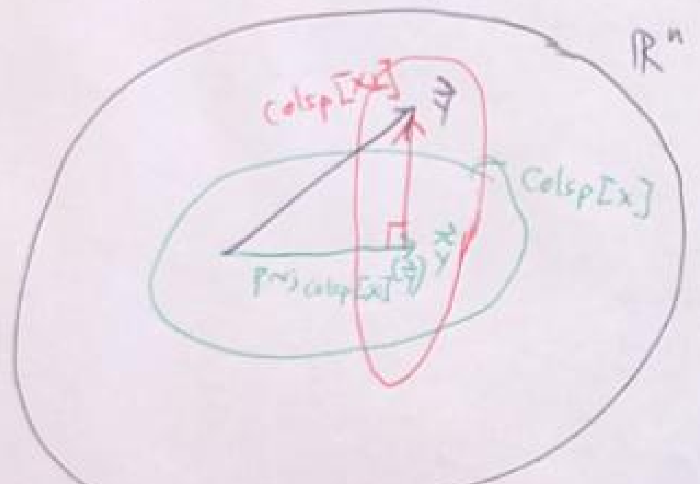
$A^{-1}A = I \Rightarrow (A^{-1}A)^T = I^T = I \Rightarrow A^T(A^{-1})^T = I \Rightarrow (A^{-1})^T = (A^T)^{-1}$

(2) H ~~is independent~~ is idompetent i.e. $HH = H$

$HH = \left(V(V^TV)^{-1}V^T\right)\left(V(V^TV)^{-1}V^T\right) = V\cancel{(V^TV)^{-1}}\cancel{(V^TV)}(V^TV)^{-1}V^T = V(V^TV)^{-1}V^T = H$ ✓

$\text{proj}_V(\text{proj}_V\vec{a}) = \text{proj}_V(H\vec{a}) = HH\vec{a} = H\vec{a} = \text{proj}_V(\vec{a})$

$\hat{y} = H\vec{y} = \text{proj}_{\text{colsp}[X]}(\vec{y})$



$\vec{y} = \vec{\hat{y}} + \vec{e}, \quad \vec{\hat{y}} \cdot \vec{e} = 0$

$\vec{e} = \vec{y} - \vec{\hat{y}} = \vec{y} - H\vec{y} = I\vec{y} - H\vec{y} = (I - H)\vec{y}$

$\vec{\hat{y}} \cdot \vec{e} = (H\vec{y})^T(I - H)\vec{y} = \vec{y}^TH^T(I\vec{y} - H\vec{y})$

$= \vec{y}^TH(I\vec{y} - H\vec{y}) = \vec{y}^THI\vec{y} - \vec{y}^THH\vec{y}$

$= \vec{y}^TH\vec{y} - \vec{y}^TH\vec{y} = 0$

Lets verify $I - H$ is a projection matrix by demonstrating that it is (1) symmetric and (2) idempotent.

$(I - H)^T = I^T - H^T = I - H$ ✓

$(I - H)(I - H) = II - IH - HI + HH = I - H \cancel{- H + H} = I - H$

$(I - H)\vec{e} = \vec{e}$

$H\vec{e} = \vec{0}_n$

$(I - H)\hat{\vec{y}} = \vec{0}_n$

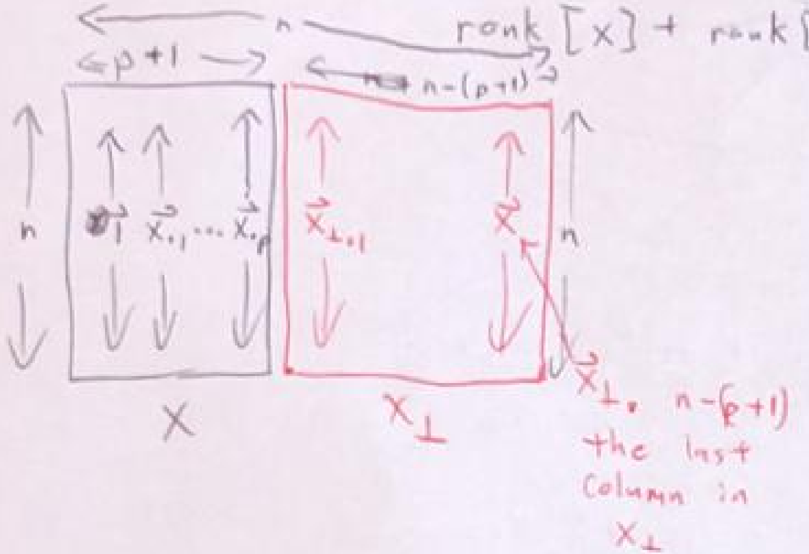
$H\hat{\vec{y}} = \hat{\vec{y}}$

$\text{colsp}[X] \oplus \underbrace{\text{colsp}[X_\perp]} = \mathbb{R}^n$

the "residual space" since it's the space the residuals e-vec live inside.

$\text{rank}[X] = p+1, \quad \text{rank}[X_\perp] = n-(p+1)$

$\text{rank}[X] + \text{rank}[X_\perp] = n$ degrees of freedom of the residuals.



The column vectors in $X_\perp$ are vectors that span the "rest of the space." They're not unique. And you can construct them computationally.