M342W 3/1/21

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

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

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

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

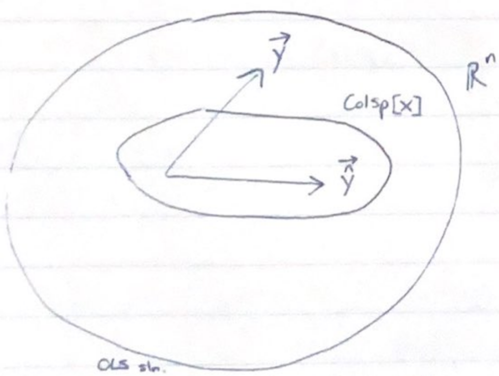
$$\overset{p+1}{\overset{\shortparallel}{\text{rank}[X]}} = \dim[\text{Colsp}[X]]$$

$$\text{Colsp}[X] := \text{Span}[\vec{1}, \vec{x}_{\cdot 1}, \ldots, \vec{x}_{\cdot p}] = \{w_0 \vec{1}_n + w_1 \vec{x}_{\cdot 1} + \ldots + w_p \vec{x}_{\cdot p} : \underset{\vec{w} \in \mathbb{R}^{p+1}}{w_0, w_1, \ldots, w_p \in \mathbb{R}}\}$$

$\Downarrow$

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

$\hat{\vec{y}} \in \text{Colsp}[X]$? Yes $\vec{w} \Leftrightarrow \vec{b}$.



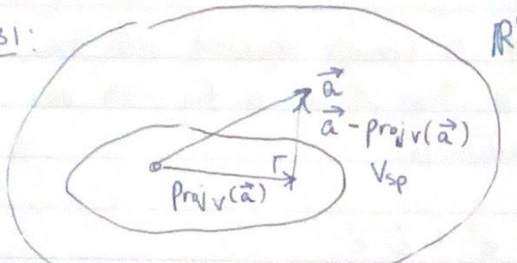
$$\hat{\vec{y}} = X\vec{b} = \underbrace{\overset{\text{OLS sln.}}{X(X^T X)^{-1} X^T}}_{H \text{ for "hat" matrix, the linear operator turning } \vec{y} \text{ into } \hat{\vec{y}}.} \vec{y} = H\vec{y} \qquad H \in \mathbb{R}^{n \times n}$$

$X\vec{b} \in \text{colsp}[X]$

$\Updownarrow$

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

231:



V is a K-dim subspace of the n-dim full space.
We want to "project" $\vec{a}$ onto V s.t. the difference between $\vec{a}$ and its projection is perpendicular.
This is called an "orthogonal projection".

$V_{sp} = \text{span}\{\vec{v}_1, \dots, \vec{v}_K\}$ $K < n$
$\text{proj}_V(\vec{a}) \in \text{span}\{\vec{v}_1, \dots, \vec{v}_K\}$
$\Rightarrow \text{proj}_V(\vec{a}) = w_1\vec{v}_1 + \dots + w_K\vec{v}_K$
$(\exists \vec{w}) \quad = V\vec{w}$

We want a formula for this projection as a function of the space V.

s.t. $\forall\, V = [\vec{v}_1 | \cdots | \vec{v}_K]$, $\vec{w} \in \mathbb{R}^K$

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

$\Rightarrow \cancel{\forall_j} \quad (\vec{a} - V\vec{w})^T v_j = 0 \quad \forall j \iff v_j^T(\vec{a} - V\vec{w}) = 0 \quad \forall j$

$$\Rightarrow \begin{array}{l} v_1^T(\vec{a} - V\vec{w}) = 0 \\ v_2^T(\vec{a} - V\vec{w}) = 0 \\ \vdots \\ v_K^T(\vec{a} - V\vec{w}) = 0 \end{array} \Rightarrow \begin{bmatrix} \vec{v}_1^T \\ \vec{v}_2^T \\ \vdots \\ \vec{v}_K^T \end{bmatrix} (\vec{a} - V\vec{w}) = \vec{0}_K \Rightarrow V^T(\vec{a} - V\vec{w}) = \vec{0}_K$$

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

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

We call H (an n×n matrix), the orthogonal projection matrix onto the subspace $V_{sp} = \text{colsp}[V]$

$\uparrow$
Orthogonal projection onto colsp[V]

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

Properties of orthogonal projection matrices, H

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

$$H^T = (V(V^TV)^{-1}V^T)^T = V^{T^T}((V^TV)^{-1})^TV^T = V((V^TV)^{-1})^TV^T = V((\overbrace{V^TV}^{V^TV})^T)^{-1}V^T = H$$

~~2) H is idempotent, i.e. HH = H~~

- Let A be square, invertible, and symmetric.

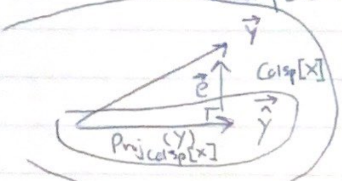
$A^{-1}A = I \Rightarrow (A^{-1}A)^T = I^T = I \Rightarrow A^T(A^{-1})^T = I \Rightarrow (A^{-1})^T = (A^T)^{-1}$

2) H is idempotent, i.e. $HH = H$

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

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

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



$\mathbb{R}^n$ $\quad y = \hat{\vec{y}} + \vec{e}, \quad \hat{\vec{y}} \cdot \vec{e} = 0$

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

$$\hat{\vec{y}} \cdot \vec{e} = (H\vec{y})^T(I-H)\vec{y} = \vec{y}^T H^T(I\vec{y} - H\vec{y})$$
$$= \vec{y}^T H(I\vec{y} - H\vec{y}) = \vec{y}^T HI\vec{y} - \vec{y}^T HH\vec{y}$$
$$= \vec{y}^T H\vec{y} - \vec{y}^T H\vec{y} = 0$$

- If 1), 2) then matrix is an orthogonal projection matrix H.

Let's verify $I-H$ is a projection matrix by demonstrating that it is symmetric and idempotent.

1) $(I-H)^T = I^T - H^T = I - H$ ✓

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

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

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

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

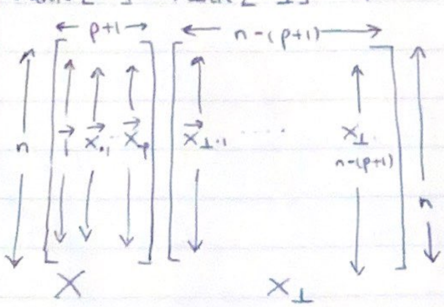
$H\vec{y} = \hat{\vec{y}}$

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

the "residual space" since it is the space the residuals $\vec{e}$ live inside

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

$\text{rank}[X] + \text{rank}[X_\perp] = n$ $\quad$ 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.