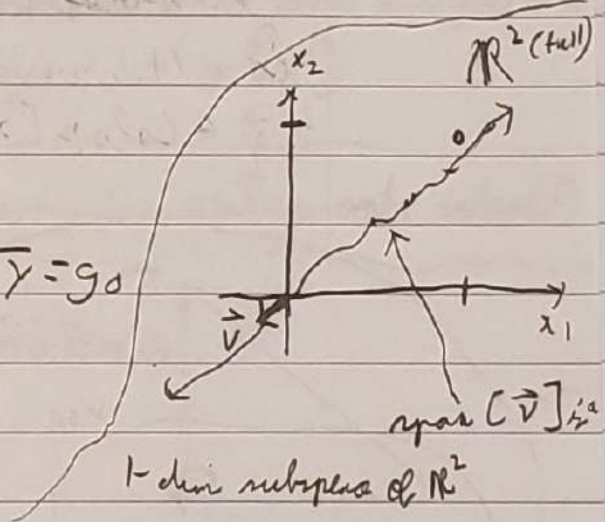
3/1/21

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

What if we have no features? ie the null model case. Is there an OLS solution?

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

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



p+1

rank $[X]$ = dim [Colsp $[X]$]

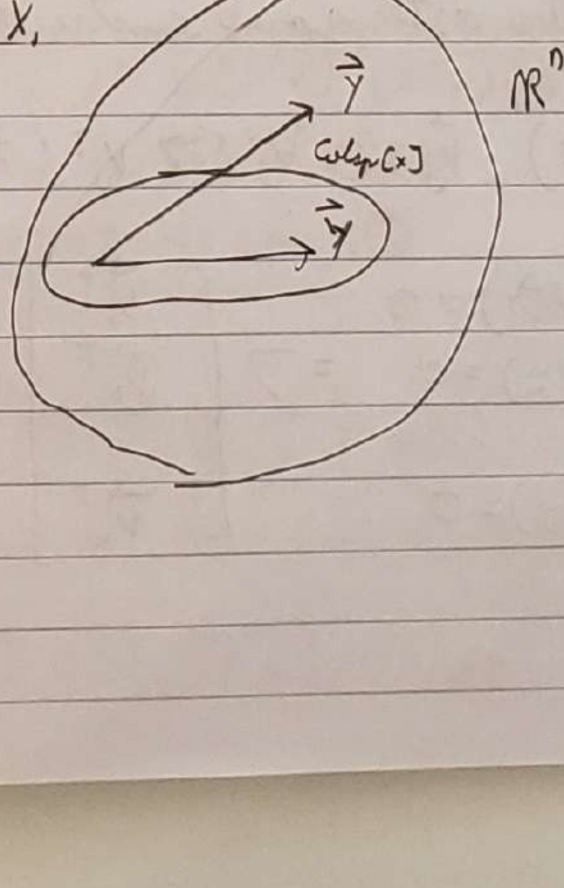
n-dim col. vector

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

"

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.

$\hat{\vec y} \in$ Colsp $[X]$? yes



OLS sol.

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

$$\vec{\hat{y}} = X\vec{b} = \underbrace{X(X^TX)^{-1}X^T}\vec{y} = H\vec{y}$$

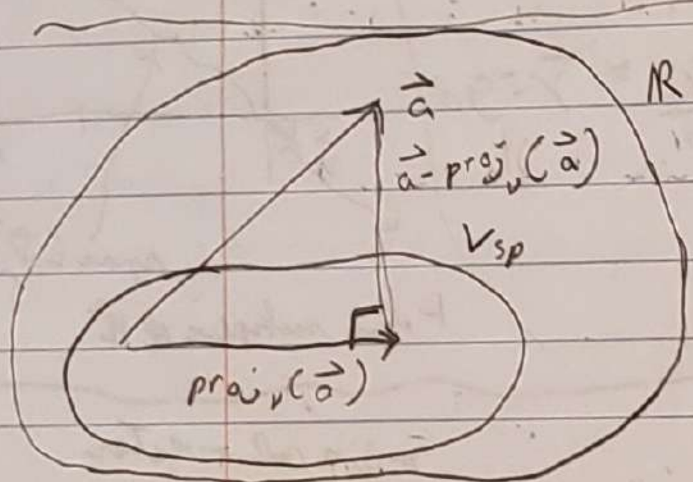
H for "hat matrix", the linear operator turning y-vec into $\hat{y}$-vec

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

$\Updownarrow$

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

---



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. This is called an "orthogonal projection". We want a formula for this projection as a function of the space V.

$V = \text{Span}\{\vec{v}_1, \dots \vec{v}_K\}$, $K < n$

$\text{proj}_V(\vec{a}) \in \text{Span}\{\vec{v}_1 \dots \vec{v}_K\} \Rightarrow \exists \vec{w}$

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

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

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

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

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

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

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

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

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, $H^T = H$

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

Let A be square, invertible and symmetric

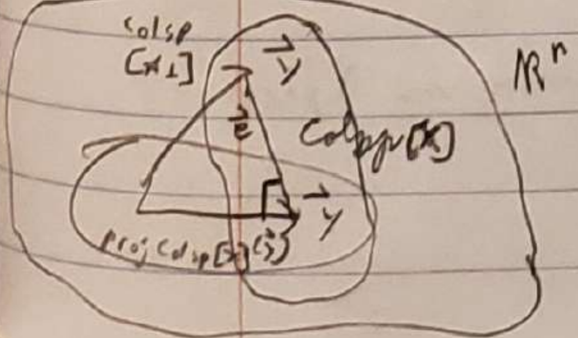
$$A^{-1}A = I \Rightarrow (A^{-1}A)^T = I^T = I \Rightarrow A^T(A^{-1})^T = I \Rightarrow (A^{-1})^T = (A^T)^{-1}$$

2 H is idempotent, ie. $HH = H$

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

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

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



$$\vec{y} = \hat{\vec{y}} + \vec{e}, \quad \hat{\vec{y}} \cdot \vec{e} = 0$$
$$\vec{e} = \vec{y} - \hat{y} = \vec{y} - H\vec{y} = I\vec{y} - H\vec{y}$$
$$= (I - H)\vec{y}$$

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

$=$

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

$= \vec{y}^T H\vec{y} - \vec{y}^T H\vec{y} = 0$

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

$(I-H)^T = I^T - H^T = I - H$ ✓

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

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

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

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

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

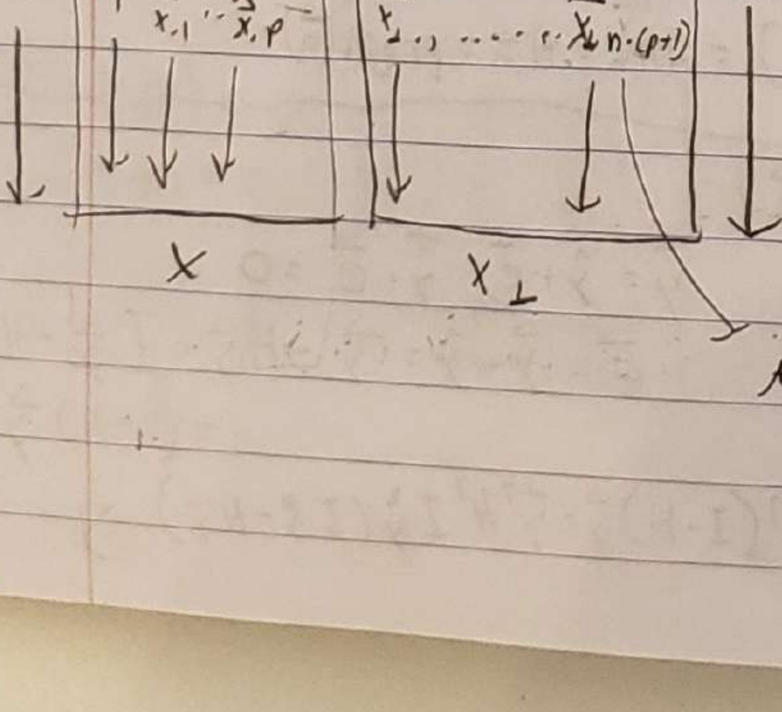
$\text{Colsp}[X] \oplus \underline{\text{Colsp}[X_\perp]} = \mathbb{R}^n$

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

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

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

degrees of freedom of the residuals



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

the last column in X_{prep}