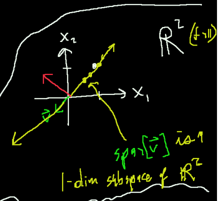
$\vec{b} = (X^\top X)^{-1} X^\top \vec{y}$, the OLS linear model , $\hat{\vec{y}} = X\vec{b}$, $g(\vec{x}_*) = \hat{y}_* = \vec{x}_* \vec{b}$

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

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

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



$$\overset{p+1}{\overset{\|}{\text{rank}[X]}} = \dim[\text{colsp}[X]]$$

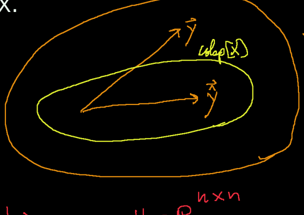
$$\text{colsp}[X] := \text{span}[\vec{1}, \vec{x}_{\cdot 1}, \ldots, \vec{x}_{\cdot p}] := \{ w_0 \vec{1}_n + w_1 \vec{x}_{\cdot 1} + \ldots + w_p \vec{x}_{\cdot p} : \underbrace{w_0, w_1, \ldots, w_p \in \mathbb{R}}_{\vec{w} \in \mathbb{R}^{p+1}} \}$$

n-dim col. vecs

$\Downarrow$

p+1 < n 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



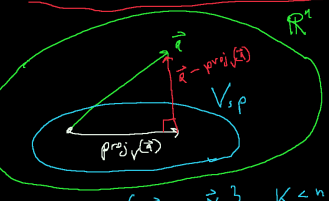
$$\hat{\vec{y}} = X\vec{b} \overset{\text{OLS sol.}}{=} \underbrace{X(X^\top X)^{-1}X^\top}_{H} \vec{y} = H\vec{y} \qquad H \in \mathbb{R}^{n \times n}$$

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

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

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

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

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

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

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

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

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

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

orthogonal projection onto colsp[V]

We call the n x n matrix H the orthogonal projection matrix onto the subspace Vsp = colsp[V].

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

Properties that define orthogonal projection matrices, H

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

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

Let A be square, invertible and symmetric

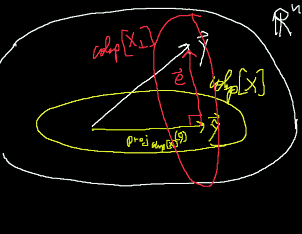
$A^{-1}A = I \Rightarrow (A^{-1}A)^\top = I^\top = I \Rightarrow A^\top \underbrace{(A^{-1})^\top} = I \quad \Rightarrow (A^{-1})^\top = (A^\top)^{-1}$

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

$$HH = (V(V^\top V)^{-1}V^\top)(V(V^\top V)^{-1}V^\top) = V\cancel{(V^\top V)^{-1}}\cancel{(V^\top V)}(V^\top V)^{-1}V^\top = V(V^\top V)^{-1}V^\top = H \checkmark$$

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

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



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

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

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

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

$(I - H)^\top = I^\top - H^\top = I - H \checkmark$

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

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

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

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

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

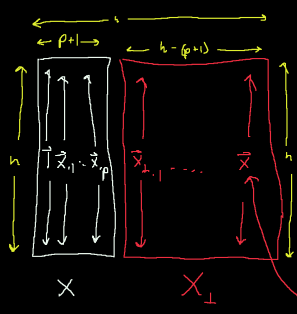
$$\text{colsp}[X] \oplus \underbrace{\text{colsp}[X_\perp]} = \mathbb{R}^n$$

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

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

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

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.

$\vec{x}_{\perp \cdot n-(p+1)}$ the last column in X_perp