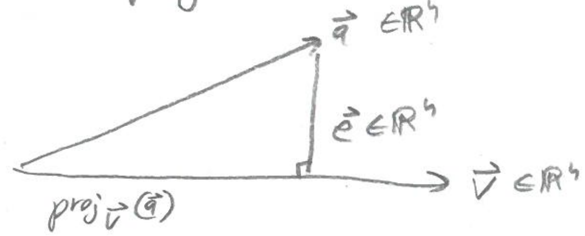
Lec 12 Math 340.4 3/14/18

orthogonal projection



$$= \frac{\vec v \vec v^T}{\|v\|^2}\vec a = H\vec a \qquad \frac{\vec v \vec v^T}{\|v\|^2} \in \mathbb{R}^{n\times n}$$

"H" is called a projection matrix. They are

Note:

① $\vec a = proj_{\vec v}(\vec a) + \vec e$ and ② $\underbrace{\vec e \cdot proj_{\vec v}(\vec a) = 0}_{\text{orthogonal}}$

$\Rightarrow \vec e = \vec a - proj_{\vec v}(\vec a)$

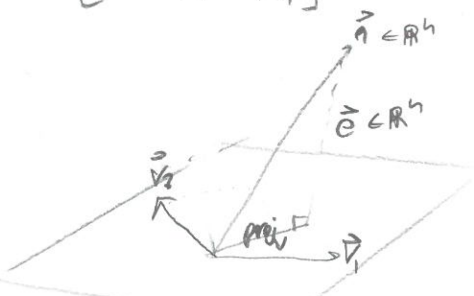
④ $proj_{\vec v}(\vec a) = c\,\vec v = \ell\,\vec v_0$ and ③ $\underbrace{\vec e \cdot \vec v = 0}_{\text{orthogonal}}$

$\in colsp\{\vec v\}$

i.e. all lin. comb. $\vec v$.

Instead of projecting onto one vector, project onto an entire space $\mathcal{V} := colsp\{V\}$

$$V = [\vec v_1 \mid \vec v_2 \mid \dots \mid \vec v_k] \in \mathbb{R}^{n\times k}$$



By Note 4:

$$proj_V(\vec a) \in colsp(V)$$

$$\Rightarrow proj_V(\vec a) = w_1\vec v_1 + w_2\vec v_2 + \dots + w_k\vec v_k = V\vec w$$

By Note 1:

$$\vec e = \vec a - V\vec w$$

By note 3:

$$\vec e\cdot\vec v_1 = 0,\ \vec e\cdot\vec v_2 = 0,\ \dots,\ \vec e\cdot\vec v_k = 0$$

$$\vec v_1^{\,T}(\vec a - V\vec w) = 0,\ \vec v_2^{\,T}(\vec a - V\vec w) = 0,\ \dots,\ \vec v_k^{\,T}(\vec a - V\vec w) = 0$$

$$\Rightarrow V^T(\vec a - V\vec w) = \vec 0_n$$

$$\Rightarrow V^T\vec a - V^TV\vec w = \vec 0_n \Rightarrow V^T\vec a = V^TV\vec w$$

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

Consider $AA^{-1} = I$

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

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

$\quad \parallel$

$(A^T)^{-1}$

Properties

(a) symmetric $\left(V(V^TV)^{-1}V^T\right)^T = V\left((V^TV)^{-1}\right)^T(V^T)^T = V(V^TV)^{-1}V^T$

(b) idempotent $HH = H$ Pf: $\left(V(V^TV)^{-1}V^T\right)\left(V(V^TV)^{-1}V^T\right) = H$ ✓

What did we learn? LS sol is the same as orthogonally projecting $\vec{y}$ onto the colsp of X. $HH = ?$

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

Once you project $\vec{y}$ onto colsp(X) it is already projected. Doing it again does nothing!

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

Fact $(A^{-1})^T = (A^T)^{-1}$

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

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

residuals..

What is $\vec{e}$? $\vec{e} = \vec{y} - \hat{\vec{y}} = \vec{y} - X(X^TX)^{-1}X^T\vec{y} = I\vec{y} - X(X^TX)^{-1}X^T\vec{y}$

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

Make sense?

$(I-H)(I-H) = I - HI - IH + HH = I - 2H + H = I - H$

also idempotent! It's a proj. onto the residual space (space orth. to X)

$\text{rank}(I-H) = n - (p+1)$ the leftover dimensions, colsp$(X^\perp)$.

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

$= H\vec{y} + I\vec{y} - H\vec{y} = I\vec{y} = \vec{y} \checkmark$

And $\hat{\vec{y}} \perp \vec{e}$

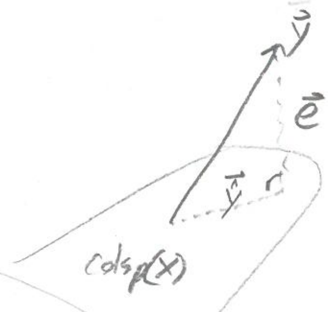
$\hat{\vec{y}} \cdot \vec{e}$

$= \hat{\vec{y}}^T\vec{e}$

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

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

$= \vec{y}^T(H - \underbrace{HH}_{H})\vec{y} = 0 \checkmark$



FYI LS: hope you can get a good model projecting on to $p+1 < n$ dimensions. Compression!! try to get same information with less space $\Rightarrow$ lose!

What is RMSE? Previously $RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n-2} SSE}$. The general formula is:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n-(p+1)} SSE}$$

R^2? Still

$$R^2 = \frac{s_y^2 - s_e^2}{s_y^2} = 1 - \frac{SSE}{SST}$$

doesn't change!

proved before beginning

Looking at this picture, we can invoke the Pythagorean Thm to show:

$$\|\vec{y}\|^2 = \|\hat{\vec{y}}\|^2 + \|\vec{e}\|^2$$

$$\Rightarrow \sum y_i^2 = \sum \hat{y}_i^2 + \sum e_i^2$$

Now subtract $n\bar{y}^2$ from both sides

$$\Rightarrow \sum y_i^2 - n\bar{y}^2 = \sum \hat{y}_i^2 - n\bar{y}^2 + \sum e_i^2$$

Note $\sum (y_i - \bar{y})^2 = \sum y_i^2 - \sum y_i\bar{y} - \sum \bar{y} y_i + \sum \bar{y}^2 = \sum y_i^2 - n\bar{y}^2 - n\bar{y}^2 + n\bar{y}^2 = \sum y_i^2 - n\bar{y}^2$

$\sum (\hat{y}_i - \bar{y})^2 = \sum \hat{y}_i^2 - 2\bar{y}\sum \hat{y}_i + n\bar{y}^2 = \sum \hat{y}_i^2 - n\bar{y}^2$

$$\sum \hat{y}_i = \hat{y} \cdot \vec{1}_n = \hat{y}^T \vec{1} = (H\vec{y})^T \vec{1} = \vec{y}^T H^T \vec{1} = \vec{y}^T \underbrace{H\vec{1}} = \vec{y}^T \vec{1} = \sum y_i = n\bar{y}$$

since $\vec{1} \in colsp(X)$ $H\vec{1} = \vec{1}$

$$\Rightarrow \underbrace{\sum (y_i - \bar{y})^2}_{SST} = \underbrace{\sum (\hat{y}_i - \bar{y})^2}_{SSR \text{ or } SSM} + \underbrace{\sum e_i^2}_{SSE}$$

the famous "sum of squares equivalence" for linear LS models

$$R^2 := \frac{SSR}{SST} = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST}$$

$$= \frac{\frac{1}{n-1}(\ )}{\frac{1}{n-1}(\ )} = \frac{s_y^2 - s_e^2}{s_y^2}$$

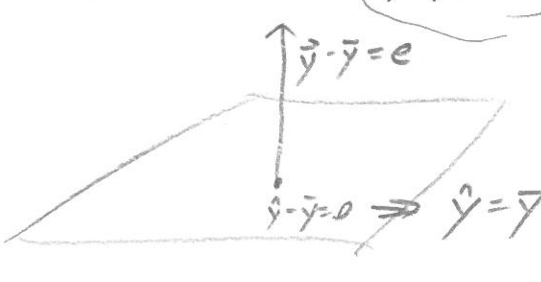
R^2 high can be thought of as:

$H(\vec{y} - \bar{y}) = H(\vec{y} - \bar{y}\vec{1})$
$= \hat{y} - \bar{y}H\vec{1}$
$= \hat{y} - \bar{y}\vec{1}$

these two vectors are about the same length

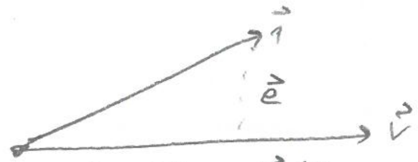
R^2 low is

R^2 zero is



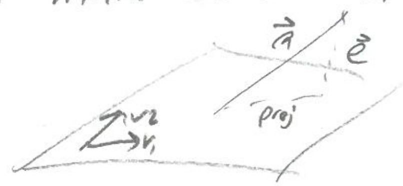
the X's are indep of the $\vec{y} - \bar{y}$!

Let's do some more lin. alg.



$$\text{Proj}_{\vec v}(\vec a) = \frac{\vec v \vec v^T}{\|\vec v\|^2}\vec a$$

What if we want to project onto $V = \text{colsp}(\vec v_1, \vec v_2)$ lin. indep. vectors. Let $V = [\vec v_1, \vec v_2$



$$\text{Proj}_V(\vec a) = \text{Proj}_{\vec v_1}(\vec a) + \text{Proj}_{\vec v_2}(\vec a)$$

Recall

$$\|a+b\|^2 = \|a\|^2 + \|b\|^2 + 2\|a\|\|b\|\cos(\theta)?$$

Note that:

$$\|\text{Proj}_V(\vec a)\|^2 = \|\text{Proj}_{v_1}(\vec a)\|^2 + \|\text{Proj}_{\vec v_2}(\vec a)\|^2 + \|\vec e\|^2 + 2\|\text{Proj}_{\vec v_1}(\vec a)\|\|\text{Proj}_{v_2}(\vec a)\|\cos(\theta_{v_1,v_2})$$
$$+ 2\|\text{Proj}_{\vec v_1}(\vec a)\|\|\vec e\|\underbrace{\cos(\theta_{v_1,e})}_{0} + 2\|\text{Proj}_{\vec v_2}(\vec a)\|\|\vec e\|\underbrace{\cos(\theta_{v_2,e})}_{0}$$

→ only zero if $v_1 \perp v_2$!

There may be a simplification to "$V(V^TV)^{-1}V^T$" if we don't have to worry about these cosine terms.

Assume V orthog.

$$\text{Proj}_V(\vec a) = \text{Proj}_{\vec v_1}(\vec a) + \text{Proj}_{\vec v_2}(\vec a)$$
$$= \frac{\vec v_1 \vec v_1^T}{\|\vec v_1\|^2}\vec a + \frac{\vec v_2 \vec v_2^T}{\|\vec v_2\|^2}\vec a = \left(\frac{\vec v_1 \vec v_1^T}{\|\vec v_1\|^2} + \frac{\vec v_2 \vec v_2^T}{\|\vec v_2\|^2}\right)\vec a$$

If V orthonormal = orthog. & all columns normalized to length 1

$$= \left(\vec v_1 \vec v_1^T + \vec v_2 \vec v_2^T\right)\vec a$$

$$\begin{bmatrix} v_{11} & v_{21} \\ v_{12} & v_{22} \\ \vdots & \vdots \\ v_{1n} & v_{2n} \end{bmatrix}\begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \end{bmatrix}$$

$$= \left(\begin{bmatrix} v_{11}^2 & v_{11}v_{12} & \cdots & v_{11}v_{1n} \\ v_{12}v_{11} & v_{12}^2 & \cdots & v_{12}v_{1n} \\ \vdots & & & \vdots \\ v_{1n}v_{11} & & \cdots & v_{1n}^2 \end{bmatrix} + \begin{bmatrix} v_{21}^2 & v_{21}v_{22} & \cdots & v_{21}v_{2n} \\ v_{22}v_{21} & v_{22}^2 & \cdots & v_{22}v_{2n} \\ \vdots & & & \\ v_{2n}v_{21} & v_{2n}v_{22} & \cdots & v_{2n}^2 \end{bmatrix}\right)\vec a = \Bigg(\underbrace{\begin{bmatrix} \uparrow & \uparrow \\ \vec v_1 & \vec v_2 \\ \downarrow & \downarrow \end{bmatrix}}_{Q}\underbrace{\begin{bmatrix} \leftarrow \vec v_1^T \rightarrow \\ \leftarrow \vec v_2^T \rightarrow \end{bmatrix}}_{Q^T}\Bigg)\vec a = QQ^T\vec a$$