Lec 13 Math 390.3 3/19/20

Project onto colsp[V] $V = [\vec{v}_1 | \cdots | \vec{v}_k]$

$$\text{proj}_V(\vec{a}) = V(V^TV)^{-1}V^T\vec{a} = \sum_{j=1}^{k} \frac{\vec{v}_j\vec{v}_j^T}{\|\vec{v}_j\|^2}\vec{a} = \sum_{j=1}^{k} \vec{v}_j\vec{v}_j^T\vec{a} = VV^T\vec{a}$$

If V orthog, if colns are normalized

$$\|\text{proj}_V(\vec{a})\|^2 = \sum_{j=1}^{k} \left\| \frac{\vec{v}_j\vec{v}_j^T}{\|\vec{v}_j\|^2}\vec{a}\right\|^2 = \sum_{j=1}^{k}\|\vec{v}_j\vec{v}_j^T\vec{a}\|^2$$

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

a complicated functional formula for the orth. proj. onto colsp[X].

If each of the cols of X are orthonormal ... will this become easier?

$X = QR$

from Gram Schmidt alg. — being careful to be able to reverse the alg.

$Q \in \mathbb{R}^{n \times (p+1)}$, $R = \mathbb{R}^{(p+1)\times(p+1)}$ square, upper tri. and full rank (otherwise Q must be deficient).

Imagine we wish to project onto $\text{colsp}[Q] = \text{colsp}(X)$? We showed last time that $H = QQ^T$ using the sum of the projections onto each orthonormal dimension

Another proof:

$$\hat{\vec{y}} = Q(Q^TQ)^{-1}Q^T\vec{y} = QQ^T\vec{y}$$

$$Q^TQ = \begin{bmatrix} \longleftarrow q_{\cdot 1} \longrightarrow \\ \longleftarrow q_{\cdot 2} \longrightarrow \\ \vdots \\ \longleftarrow q_{\cdot p+1} \longrightarrow \end{bmatrix}\begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ q_{\cdot 1} & q_{\cdot 2} & \cdots & q_{\cdot p+1} \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix} = I_{p+1}$$

$\vec{q}_{\cdot 1}^T\vec{q}_{\cdot 1} = \|\vec{q}_{\cdot 1}\|^2 = 1$

$q_{\cdot 1}^T\vec{q}_{\cdot 2} = 0$ why? orthog!

QR helps in L.S. Alg

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

$\Rightarrow X^TX\vec{b} = X^T\vec{y}$

$\Rightarrow (QR)^TQR\vec{b} = (QR)^T\vec{y}$

$\Rightarrow R^T\underbrace{Q^TQ}_{I}R\vec{b} = R^T\underbrace{Q^T\vec{y}}_{\vec{z}}$

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

$\Rightarrow R^TR\vec{b} = R^T\vec{z}$,

$\Rightarrow R\vec{b} = \vec{z}$

$\Rightarrow$ since R is upper triangular, this eq is easily solved via back-subst e.g.

$$\begin{bmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$\Rightarrow fy_3 = x_3 \Rightarrow y_3 = \frac{x_3}{f}$

$\Rightarrow dy_2 + ey_3 = x_2 \Rightarrow dy_2 = x_2 - e\frac{x_3}{f} \Rightarrow y_2 = \frac{1}{d}\left(x_2 - e\frac{x_3}{f}\right)$

etc...

Recall $SST = SSR + SSE \Rightarrow$ $SSR\uparrow \Rightarrow SSE\downarrow \Rightarrow R^2\uparrow$, $RMSE\downarrow$

$SSR\downarrow \Rightarrow SSE\uparrow \Rightarrow R^2\downarrow$, $RMSE\uparrow$

$f(\vec{x})$ only doesn't change!

Now we have a new expression for $\hat{\vec{y}}$:

$$\hat{\vec{y}} = H\vec{y} = QQ^T\vec{y} = \text{proj}_Q(\vec{y}) = \sum_{j=1}^{p} \text{proj}_{\vec{q}_j}(\vec{y})$$

by Pythagorean thm!

$$\sum \hat{y}_i^2 = \|\hat{\vec{y}}\|^2 = \left\| \sum_{j=1}^{p} \text{proj}_{\vec{q}_j}(\vec{y}) \right\|^2 = \sum_{j=1}^{p} \|\text{proj}_{\vec{q}_j}(\vec{y})\|^2$$

$$SSR = \sum \hat{y}_i^2 - n\bar{y}^2 = \sum_{j=1}^{p} \|\text{proj}_{\vec{q}_j}(\vec{y})\|^2 - n\bar{y}^2$$

What if we added a new lin. ind. predictor $\vec{x}_{\cdot new}$ to X. The new rank$(X) = p+1+1$

What happens to SSR?

the new predictor orthonormalized

$$SSR_{new} = \left(\sum_{j=1}^{p} \|\text{proj}_{\vec{q}_j}(\vec{y})\|^2 \right) + \underbrace{\|\text{proj}_{\vec{q}_{new}}(\vec{y})\|^2}_{>0} - n\bar{y}^2$$

$\Rightarrow SSR_{new} > SSR \Rightarrow SSE_{new} < SSE$

What happens to R^2

$$R^2_{new} = \frac{SSR}{SST} > R^2 = \frac{SSR}{SST}$$ since $SST := \sum(y_i - \bar{y})^2$ and does not change

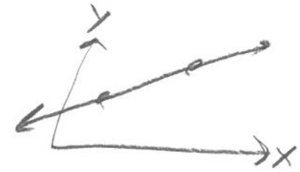
$\Rightarrow SSE_{new} < SSE \Rightarrow RMSE_{new} < RMSE$.

We can increase R^2 (and lower RMSE) just by adding a predictor.

Now can it be a made up vector? Yes!! $\boxed{\text{DEMO}}$

Taking this further... what happens if we add so many predictors that $p+1 = n$ i.e. the # of rows? $\Rightarrow$ X lin. ind. is now $\mathbb{R}^{n\times(p+1)} = \mathbb{R}^{n\times n}$ square and full rank

What is H? $(AB)^{-1} = B^{-1}A^{-1}$



try to picture this with $n=100$!

$$H = X(X^\top X)^{-1}X^\top = \overbrace{XX^{-1}}^{I}\overbrace{(X^\top)^{-1}X^\top}^{I} = I$$

$$\Rightarrow \underbrace{\hat{\vec{y}}}_{X\vec{b}} = H\vec{y} = I\vec{y} = \vec{y}. \Rightarrow X\vec{b} = \vec{y} \Rightarrow \vec{b} = X^{-1}\vec{y}$$ linear equation! Can be solved!

$$\Rightarrow \vec{e} = \vec{y} - \hat{\vec{y}} = \vec{0} \Rightarrow R^2 = 100\%,\ RMSE = 0.$$

This can't be real! **It's NOT.**

What is going on? This is called **overfitting**. There are many real world causes. ← MAJOR CLASS CONCEPT

Mindlessly adding predictors is **not** something you would do! It's just an illustration. ~~We will be covering overfitting in detail later.~~

~~What we should be able to see now is the SSE / R^2 / RMSE —~~

all ~~of our errors a~~

What happened? Recall

$$y = h^*(\vec{x}) + \varepsilon$$

$\uparrow$ $x_1, \ldots, x_p$

$h^* \in \mathcal{H}$, the "best" model in the class. The $= \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$

$$\mathcal{H} = \{\vec{w}\cdot\vec{x} : \vec{w} \in \mathbb{R}^{p+1}\}$$

y only depends on $x_1, \ldots, x_p$. If we fabricate $\{x_{p+1}, \ldots, x_{n-1}\}$ the new predictors, then we fit a new model with a new hypothesis set

$$\mathcal{H} \subset \mathcal{H}_{new} = \{\vec{w}\cdot\vec{x} : \vec{w} \in \mathbb{R}^{n}\}$$

but h^* hasn't changed!

$$h^* = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p + \overbrace{0x_{p+1} + 0x_{p+2} + \ldots + 0x_{n-1}}^{\text{these should all be zero!}}$$