Lec 11 Math 390.4 3/12/18

Prediction in lin regr models:

$$g(\vec{x}^*) = \vec{x}^* \vec{b}$$

make sure to take the (or T) since obs's are row vectors and $\vec{b}$ is a col. vec.

Extrapolation.

What if $x^* \notin \text{Range}[X]$?

Define $\text{Range}[X] = \underbrace{[x_{\cdot 1, \min}, x_{\cdot 1, \max}]}_{\text{min/max of first predictor}} \times \underbrace{[x_{\cdot 2, \min}, x_{\cdot 2, \max}]}_{\text{min/max of second predictor}} \times \ldots \times \underbrace{[x_{\cdot p, \min}, x_{\cdot p, \max}]}_{\text{min/max of last predictor}}$

If you run g on such an $\vec{x}^*$, you will likely get a wrong answer. (demo) It is suspect if it's towards the tails of the data of the predictors too. It depends on how thin the tails are.

Warning: most people ignore this!!

Review of Lec10: $\vec{y} \in \mathbb{R}^n$, $X \in \mathbb{R}^{n \times (p+1)}$. The least squares linear model fit is given by $\vec{b} = (X^\top X)^{-1} X^\top \vec{y}$

$$\hat{\vec{y}} = \begin{bmatrix} \leftarrow \vec{x}_1 \rightarrow \\ \leftarrow \vec{x}_2 \rightarrow \\ \vdots \\ \leftarrow \vec{x}_n \rightarrow \end{bmatrix} \vec{b} = X\vec{b} = \underbrace{X(X^\top X)^{-1}X^\top}_{H} \vec{y} = H\vec{y}$$

"hat matrix" because it turns $\vec{y} \rightarrow \hat{\vec{y}}$

What is rank(H)? $p+1$. Why? $\hat{y}_i = b_0 + b_1 x_{i1} + \ldots + b_p x_{ip}$ which lives in a $p+1$ dim. space. Proof of this comes later.

How to get R^2, RMSE?

$p+1$ = degrees of freedom. Why? Dim(Colsp(X))! How many dim does model represent?

Recall $y = g(\vec{x}) + e \Rightarrow \vec{y} = \hat{\vec{y}} + \vec{e} \Rightarrow \vec{e} = \vec{y} - \hat{\vec{y}}$

$SSE := \|\vec{e}\|^2 = \sum_{i=1}^n e_i^2$, $MSE = \frac{1}{n-(p+1)} SSE \Rightarrow RMSE = \sqrt{MSE}$

$R^2 = \frac{s_y^2 - s_e^2}{s_y^2}$ doesn't change!

$\boxed{\text{DEMO}}$

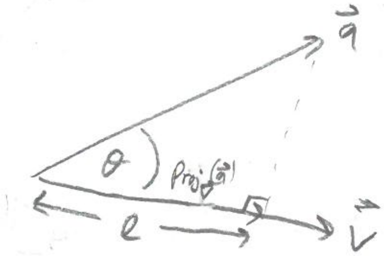
$$\hat{\vec{y}} = \begin{bmatrix} \leftarrow x_{1\cdot} \rightarrow \\ \leftarrow x_{2\cdot} \rightarrow \\ \vdots \\ \leftarrow x_{n\cdot} \rightarrow \end{bmatrix} \vec{b} = X\vec{b} = \underbrace{X(X^TX)^{-1}X^T}_{H \in \mathbb{R}^{n \times n}} \vec{y} = H\vec{y}$$

"the hat matrix"

What is rank(H)? H takes $\vec{y}$ which is n-dim and lays it onto a $p+1$-dim subspace $\Rightarrow$ rank(H) $= p+1$. How do we know? Let's do some lin. alg.

Let's do some pure lin algebra...

Let's project two vectors $\in \mathbb{R}^d$



"orthogonal projection"

perpendicular

By law of cosines, $\cos(\theta) = \dfrac{\vec{a}\cdot\vec{v}}{\|\vec{a}\|\|\vec{v}\|}$

Let l be length of projection. Using 9th grade trig,

$$l = \|\vec{a}\|\cos(\theta) = \frac{\vec{a}\cdot\vec{v}}{\|\vec{v}\|}$$

Now we take l and give it the correct direction:

$$\text{Proj}_{\vec{v}}(\vec{a}) = l\underbrace{\frac{\vec{v}}{\|\vec{v}\|}}_{\text{unit vector in } \vec{v} \text{ dir}} = \frac{\vec{a}\cdot\vec{v}}{\|\vec{v}\|^2}\vec{v} = \frac{\vec{a}\cdot\vec{v}}{\vec{v}\cdot\vec{v}}\vec{v} = \frac{\vec{v}\vec{v}^T\vec{a}}{\vec{v}^T\vec{v}} = \underbrace{\frac{\vec{v}\vec{v}^T}{\vec{v}^T\vec{v}}}_{\in \mathbb{R}^{d\times d}}\vec{a} = H\vec{a}$$

projection matrix

Note if $\|\vec{v}\| = 1 \Rightarrow H = \vec{v}\vec{v}^T$

Note $HH = \dfrac{\vec{v}\vec{v}^T\vec{v}\vec{v}^T}{\|\vec{v}\|^2\|\vec{v}\|^2} = H$ idempotency!

Let's project $\vec{a}$ onto $\mathcal{V} = \text{colsp}(V)$

all lin ind.

let $V = [\vec{v}_1 \cdots \vec{v}_K]$

You know $\text{Proj}_{\mathcal{V}}(\vec{a}) = V\vec{w}$ since it must be $\in \text{colsp}(v_1, \ldots v_K)$

$V\vec{w} - \vec{a} \perp$ each $\vec{v}_1, \ldots \vec{v}_K$

$$\vec{v}_1^T(V\vec{w} - \vec{a}) = 0$$
$$\vec{v}_2^T(V\vec{w} - \vec{a}) = 0$$
$$\vdots$$
$$\vec{v}_d^T(V\vec{w} - \vec{a}) = 0$$

$$\Rightarrow V^T(V\vec{w} - \vec{a}) = 0$$
$$\Rightarrow V^TV\vec{w} = V^T\vec{a}$$
$$\Rightarrow \vec{w} = (V^TV)^{-1}V^T\vec{a}$$
$$\Rightarrow V\vec{w} = V(V^TV)^{-1}V^T\vec{a}$$
$$\Rightarrow \text{Proj}_{\mathcal{V}}(\vec{a}) = \underbrace{V(V^TV)^{-1}V^T}_{H}\vec{a}$$

H is called a proj. matrix.

Properties:

(a) symmetric

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

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

(b) idempotent $HH = H$

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