Review

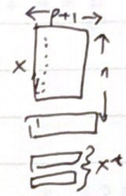
(Lecture 8) 02/26/19

rank $[X] = p+1$ $\Rightarrow \vec{b} = (X^T X)^{-1} X^T \vec{y}$

full rank

"design matrix"

$(p+1)\times n$ $\quad (p+1)\times n \quad (p+1)\times n \quad n\times 1$



OLS estimates

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

prediction

$\hat{y}_* = g(\vec{x}_*) = b_0 + b_1 x_{*,1} + \ldots + b_p x_{*,p} = \vec{x}_* \vec{b}$

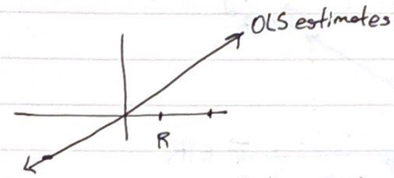
What if $\vec{x}^*$ is far from all x?

$$\text{Range}[X] = \underbrace{[X_{\cdot 1,min}, X_{\cdot 1,max}] \times [X_{\cdot 2,min}, X_{\cdot 2,max}] \times \ldots \times [X_{p,min}, X_{p,max}}_{\text{Continuous}}$$

$\times \{\text{all levels of } x_j\} \times$ cathegorical

$\vec{x}^* \notin$ Range $[X]$ historical training data?

$\hat{y} = g(x^*)$ extrapolation is risky



if $\vec{x}^* \in$ Range $[X]$ prediction is an interpolation.

0, 1 are reference color.

if predicting for purple,

Note: supervised learning doesn't generally work when extrapolating. Further, different models extrapolate, very different.

How well does this model perform?

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum (y_i - \hat{y}_i)^2 = (\vec{y} - \hat{\vec{y}})^T (\vec{y} - \hat{\vec{y}}) = \|\vec{e}\|^2$$

$\hat{y}_i = g(\vec{x}_i)$

substitute OLS

$$\hat{\vec{y}} = X\vec{b} = \underbrace{X(X^T X)^{-1} X^T}_{H \text{ "hat matrix"}} \vec{y}$$

$\vec{y} \to \hat{\vec{y}}$

makes y $\quad$ $\hat{y}$-hat

$$\hat{\vec{y}} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} = \begin{matrix} \vec{x}_1 \vec{b} \\ \vec{x}_2 \vec{b} \\ \vdots \\ \vec{x}_n \vec{b} \end{matrix} = \begin{bmatrix} \leftarrow x_1 \rightarrow \\ \leftarrow x_2 \rightarrow \\ \vdots \\ \leftarrow x_n \rightarrow \end{bmatrix} \vec{b}$$

MAT 231: $A \in \mathbb{R}^{n\times n}$ (full rank)

$\vec{v} \in \mathbb{R}^n$

$z = A\vec{v} = \begin{bmatrix} a_{11}v_1 + a_{12}v_2 + \ldots + a_{1n}v_n \\ \vdots \\ \vdots \end{bmatrix}$

if rank$[A] = 1$

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

rank $[H] = p+1$

dimension of col space

"degrees of freedom" of model is (the numbere of dimension in the model)

rank of H

dimensionality of $\hat{y}$

$p+1$

rank 2

$\hat{y} = Hy$ $\qquad$ $\vec{\hat{y}} \in \text{span}\left\{ [1], [x_{11}], \dots \right\}$ weigh scalar constants

$\hat{y} = b_0 + b_1 x_1 + \dots + b_p x_p$, $\quad \begin{bmatrix} \uparrow \\ \vec{\hat{y}} \\ \downarrow \end{bmatrix} = b_0 \begin{bmatrix} \uparrow \\ 1 \\ \downarrow \end{bmatrix} + b_1 \begin{bmatrix} \uparrow \\ x_{11} \\ \downarrow \end{bmatrix} + \dots + b_p \begin{bmatrix} \uparrow \\ x_p \\ \downarrow \end{bmatrix}$

if linearly correlated.

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

rank[H]

unit of y

$$RMSE = \sqrt{MSE}$$

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

unitless

multivariable regression performance is evaluated same

indicating intercept.
No intercept.

Line 93 => 0 +

r^2 => 2 variables.
R^2 => function of residuals.

practice Lab

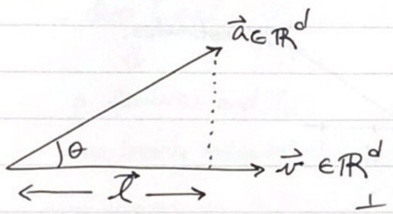
Line 172 cbind(1, boston)
intercept. dataframe needs to coerce it as matrix.

Line 212
solve(X^TX) if it's not full rank it's not invertible

Line 230 get column vector.
303 74% R^2 look like.
~. (take everything in Boston data)

Line 344 raw_mod = lm.fit(Xmm, y)
equals to $\vec{b} = (X^TX)^{-1}X^Ty$

Next Unit



$\vec{\ell}$: $\text{proj}_{\vec{v}}(\vec{a})$ "the orthogonal projection of $\vec{a}$ onto $\vec{v}$"

Using the law of cosine => $\cos(\theta) = \frac{\vec{a} \cdot \vec{v}}{\|\vec{a}\| \|\vec{v}\|} = \frac{\|\vec{\ell}\|}{\|\vec{a}\|}$

$$\Rightarrow \frac{\vec{a} \cdot \vec{v}}{\|\vec{v}\|} = \|\vec{\ell}\|$$

$$\Rightarrow \vec{\ell} = \|\vec{\ell}\| \frac{\vec{v}}{\|\vec{v}\|}$$

$$\Rightarrow \text{proj}_{\vec{v}}(\vec{a}) = \frac{\vec{a} \cdot \vec{v}}{\|\vec{v}\|^2} \vec{v} = \underbrace{\left(\frac{1}{\|\vec{v}\|^2} \vec{v} \vec{v}^T \right)}_{H \ d \times d} \vec{a}$$

Null space

$$\Rightarrow H\vec{a} = c \frac{\vec{v}}{\|\vec{v}\|} \text{ for } c \in \mathbb{R}$$

$\in \mathbb{R}^d$

"a projection matrix" of rank = 1