02/26

$\text{rank}(X) = p+1$ (dimention)

full rank

$$\vec{b} = \underset{(p+1)\times n}{(X^T X)^{-1}} \underset{(p+1)\times n}{X^T} \underset{n\times 1}{\vec{y}} \in \mathbb{R}^{p+1}$$

OLS estma

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

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

$$\text{Range}[X] = [X_{\cdot 1,min}, X_{\cdot 1,max}] \times [X_{\cdot 2\,min}, X_{\cdot 2,max}] \times \dots \times [X_{p,min}, X_{p,Max}]$$

Continuous

$\mathbb{D}$

$\vec{x}^* \in \text{Range}[X]$

* {all levels of x_j} Cathegorical

new x not in the range of $\mathbb{D}$

if it is in the range interpolation $\vec{x}^* \in \text{Range}(X)$

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

extrapolation

extrapolating if x^* is out of range

Reference back category will add to ~~red~~

Purple will be extrapolation

* Supervise Learning does not generally work ~~for~~ when extrapolating further different models extrapolate very differently

3 thing 2 colums

red 0, blue 1

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

How well does this model perform?

this is not in extrapolation !!!

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum (y_i - \hat{y}_i)^2 \rightarrow g(\vec{x_i})$$

$$= (\vec{y} - \vec{\hat{y}})^T (\vec{y} - \vec{\hat{y}}) = \|\vec{e}\|^2$$

matrix

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

$$\Rightarrow \underbrace{X (X^T X)^{-1} X^T}_{H} \vec{y}$$

H → "hat matrix"

$$\vec{y} \rightarrow \vec{\hat{y}}$$

makes $\vec{y}$ into $\vec{\hat{y}}$

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

$$\text{rank}[H] = p+1$$

$$\hat{y} = Hy$$

$$\hat{y} = b_0 + b_1 x_1 + \cdots b_p x_p$$

this is (1) → 1 vector dimention $p+1$

$\text{rank}[A] = 1$

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

$$\vec{\hat{y}} = b_0 \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} + b_1 \begin{bmatrix} \uparrow \\ x_1 \\ \downarrow \end{bmatrix} + b_p \begin{bmatrix} \uparrow \\ x_p \\ \downarrow \end{bmatrix}$$

$$\Downarrow$$

$$\vec{y} \in \text{span}\{\ \}$$

# of linearly Independent → $p+1$

# of dimention

"The degrees of freedom of the model g is . . . > $p+1$

is where they live

Span # of linear independent or linear combination

$\vec{y} \in \text{span}\left\{ \begin{bmatrix} \uparrow \\ 1 \\ \downarrow \end{bmatrix} \begin{bmatrix} \uparrow \\ X_{\cdot 1} \\ \downarrow \end{bmatrix} \right\}$ is this 2 ?

span

$$\vec{y} = b_0 \begin{bmatrix} \uparrow \\ 1 \\ \downarrow \end{bmatrix} + b_1 \begin{bmatrix} \uparrow \\ X_{\cdot 1} \\ \downarrow \end{bmatrix}$$

\# The degrees of freedom the dimentionality of $\mathcal{H}$

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

$$RMSE = \sqrt{MSE}$$

$$R^2 = 1 - \frac{SSE}{SST} = \frac{S_y^2 - S_e^2}{S_y^2}$$

Multivariable predicted performance is evaluated the same as simple regression

everithin relative to Van

ref=Van

in R remove intersept

lm (Price ~ 0 + origin, data = cars)

doing this you get the price of both cars

~~to~~ it is data frame

make it a matrix.

Solve = inverse

Lm(medv ~ ⊙ , Boston)

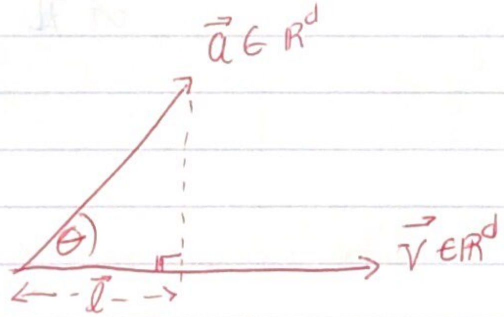
⊙ → everithing, Boston → dataset

lm.fit → look for this

Same as Lm but faster.

what is $\hat{y}$ when all $x = 0$

intersept ↑



"the orthogonal projection of a onto $\vec{v}$"

$\vec{\ell} = proj_{\vec{v}}(\vec{a})$

Using law of cosines

$\Rightarrow \cos(\theta) = \dfrac{\vec{a} \cdot \vec{v}}{\|\vec{a}\| \, \|\vec{v}\|} = \dfrac{\|\vec{\ell}\|}{\|\vec{a}\|}$

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

$\Rightarrow \vec{\ell} = \|\vec{\ell}\| \dfrac{\vec{v}}{\|\vec{v}\|} \Rightarrow proj_{\vec{v}}(\vec{a}) = \dfrac{\vec{a} \cdot \vec{v}}{\|\vec{v}\|^2} \vec{v}$

$\frac{\vec{v}}{\|\vec{v}\|}$ → unit vector supporting

doing this we normalize it

"a projection matrix"

$$\frac{\vec{a} \cdot \vec{v}}{\|\vec{v}\|^2} \vec{v} = \underbrace{\left(\frac{1}{\|\vec{v}\|^2} \vec{v}\vec{v}^T \right)}_{H \to \text{matrix}} \vec{a} = H\vec{a} = c\vec{v} \text{ for } c \in \mathbb{R}$$

$c\vec{v}$ → one portion of $\vec{v}$

$|\text{rank}| = 1$

d: dimentional vector