3/12/18

Lecture 11

Extrapolation vs Interpolation
Extrapolation can lead to "dicey" territory. (Dangerous)
You should always check that you are in the range of $\mathbb{D}$.
Silver calls Extrapolation "Out of Sample"

Extrapolation

$$x^* \notin \text{Range}[X] := [X_{\cdot 1,min}, X_{\cdot 1,max}] \times [X_{\cdot 2,min}, X_{\cdot 2,max}] \times \cdots \times [X_{\cdot p,min}, X_{\cdot p,max}]$$

$X_1 \sim f_1(x)$, $X_2 \sim f_1(x)$, ...
if x^* is in the "tails" of the distr's $\Rightarrow$ extr.

(p+1 dimensional entity)

$\vec{y} \in \mathbb{R}^n$ $X \in \mathbb{R}^{n \times (p+1)}$ s.t. $\text{rank}[X] = p+1$ linear, multivariate, least squares fitting

$\underset{(p\times 1)\times 1}{\vec{b}} = (X^T X)^{-1} X^T \vec{y}$ $\hat{y}^* = g(\vec{x}^*) = \vec{x}^* \vec{b} = b_0 + b_1 x_1^* + \cdots + b_p x_p^*$ $\hat{\vec{y}} = X\vec{b}$

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

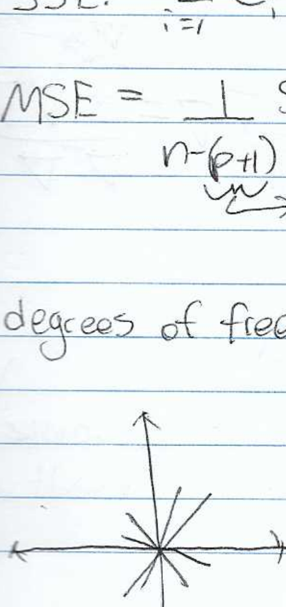
$\hat{\vec{y}} = H\vec{y}$ $\vec{y} = g(\vec{x}) + \vec{e}$ $\vec{y} = \hat{\vec{y}} + \vec{e} \Rightarrow \vec{e} = \vec{y} - \hat{\vec{y}}$
$\hookrightarrow$ dimension $(n \times 1)$

$$SSE := \sum_{i=1}^{n} e_i^2 = \|\vec{e}\|^2$$

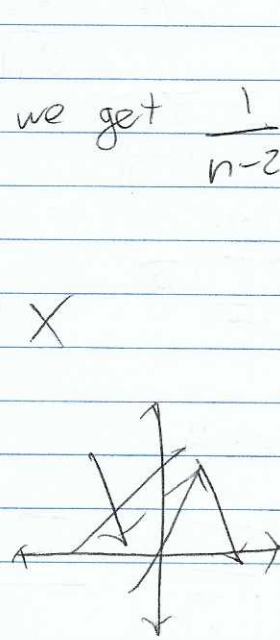
$$MSE = \frac{1}{n - (p+1)} SSE$$ $\left(\text{If } p=1 \text{ then we get } \frac{1}{n-2} SSE\right)$

$(p+1) \rightarrow$ (degrees of freedom)

degrees of freedom will be rank of X



If intercept is fixed then we can choose the slope: degrees of freedom=1



If we can choose intercept and slope, degrees of freedom = 2

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n-(p+1)} SSE} = \frac{\|\vec{e}\|}{\sqrt{n-p+1}}$$ $$R^2 := \frac{S_y^2 - S_e^2}{S_y^2}$$

(Data frames and matrices are not the same)

(Not being linearly independent is rarely a problem in real data)

R^2 = percent sample variance reduction

ANOVA = analysis of variance ANCOVA = analysis of covariance

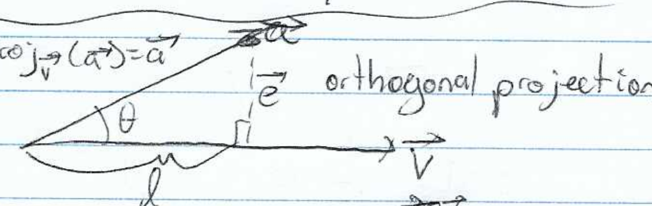
You dummify categorical variables

Calculation of R^2 works only if you have an intercept

If you add all the columns of airbags you get a column of 1's
|| || cars || ||
So if this is the case then we will have linear dependence, so one airbag type or one car type is unnecessary.

Linear Algebra

$\vec{e} + \text{proj}_{\vec{v}}(\vec{a}) = \vec{a}$



Law of Cosines: $\cos(\theta) = \frac{\vec{a} \cdot \vec{v}}{\|\vec{a}\|\|\vec{v}\|}$

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

$$\text{proj}_{\vec{v}}(\vec{a}) = l\vec{v}_0 = \frac{\vec{a} \cdot \vec{v}}{\|\vec{v}\|^2}\vec{v} = \frac{\overbrace{\vec{a}^T\vec{v}}^{\vec{v}^T\vec{a}}\,\vec{v}}{\vec{v}^T\vec{v}}$$

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

$$\text{proj}_{\vec{v}}(\vec{a}) = \left(\frac{\overset{n\times 1}{\vec{v}}\,\overset{1\times n}{\vec{v}^T}}{\vec{v}^T\vec{v}}\right)\vec{a}$$

||

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

$$\left(\begin{array}{l} HHHH(H\vec{a}) \\ \text{already projected it, so same result} \end{array}\right)$$

$$HH = \left(\frac{\vec{v}\vec{v}^T}{\|\vec{v}\|^2}\right)\left(\frac{\vec{v}\vec{v}^T}{\|\vec{v}\|^2}\right) = \frac{\vec{v}\cancel{\|\vec{v}\|^2}\vec{v}^T}{(\|\vec{v}\|^2)^{\cancel{2}}} = \frac{\vec{v}\vec{v}^T}{\|\vec{v}\|^2} = H$$