ACS234
Maths and Data Modelling

Tutorial 3
Wednesday 1pm LT04

https://github.com/ineskris/ACS234/tree/master/Tutorial3

Done in Lecture (week 5)

- Multiple Linear Regression

Matrix - Basics

Matrix 2 x 2

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Transpose

$$A^T = A' = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

Inverse

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Multiplication Matrix 2 x 2

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \times \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{12} & a_{21}b_{21} + a_{22}b_{22} \end{pmatrix}$$

Exercice 1 - Let A, B, D be the matrices below.

$$A = \begin{pmatrix} 1 & -2 \\ 3 & 0 \end{pmatrix} \quad B = \begin{pmatrix} 4 & 4 \\ -1 & 1 \end{pmatrix} \quad C = \begin{pmatrix} 7 \\ 0 \end{pmatrix} \quad D = \begin{pmatrix} -3 & 1 \end{pmatrix}$$

Compute AB, BA, BB’, B’B, AC, BD’ and A^{-1}

Exercice 1 - Solution

$$AB = \begin{pmatrix} 6 & 2 \\ 12 & 12 \end{pmatrix} \qquad BA = \begin{pmatrix} 16 & -8 \\ 2 & 2 \end{pmatrix} \qquad BB' = \begin{pmatrix} 32 & 0 \\ 0 & 2 \end{pmatrix} \qquad B'B = \begin{pmatrix} 17 & 15 \\ 15 & 17 \end{pmatrix}$$

$$AC = \begin{pmatrix} 7 \\ 21 \end{pmatrix} \qquad BD' = \begin{pmatrix} -8 \\ 4 \end{pmatrix}$$

$$A^{-1} = \begin{pmatrix} 0 & \frac{1}{3} \\ -\frac{1}{2} & \frac{1}{6} \end{pmatrix}$$

Multiple Linear Regression

Multiple regression is like linear regression, but with more than one independent value, meaning that we try to predict a value based on **two or more** variables.

$$y = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_m x_m + e$$

Response

Coefficients

Variables

Prediction $$Y = \hat{a}X + e \qquad \hat{a} = (X'X)^{-1}X'Y$$

Coefficient of determination $R^2 = 1 - \frac{S_r}{S_t}$

Sum of squared deviations $S_t = \sum_{i=1}^{n} (y_i - \bar{y})^2$

Standard Error of Estimate $S_{y/x} = \sqrt{\frac{S_r}{n - (m+1)}}$

Sum of squares of the errors $S_r = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

Case Study

Can you use a multiple regression model to predict the housing price in Boston ?

https://github.com/ineskris/ACS234/blob/master/Tutorial3/Tutorial3.ipynb

Exercice 2 - Find the coefficients of the model a1 and a2 as well as R^2 using only the data in the red box.

Wgt	Gest	Smoke
2940	38	yes
3130	38	no
2420	36	yes
2450	34	no
2760	39	yes
2440	35	yes
3226	40	no
3301	42	yes
2729	37	no
3410	40	no
2715	36	yes
3095	39	no
3130	39	yes
3244	39	no
2520	35	no
2928	39	yes
3523	41	no
3446	42	yes
2920	38	no
2957	39	yes
3530	42	no
2580	38	yes
3040	37	no
3500	42	yes
3200	41	yes
3322	39	no
3459	40	no
3346	42	yes
2619	35	no
3175	41	yes
2740	38	yes
2841	36	no

$$\textbf{Weight} = a_1 \times \textbf{Gestation} + a_2 \times \textbf{Smoke}$$

Exercice 2 - Solution

$$X = \begin{pmatrix} 35 & 1 \\ 40 & 0 \\ 42 & 1 \\ 37 & 0 \end{pmatrix} \quad X'X = \begin{pmatrix} 5958 & 77 \\ 77 & 2 \end{pmatrix} \quad (X'X)^{-1} = \begin{pmatrix} 3.34057124e-04 & -1.28611993e-02 \\ -1.28611993e-02 & 9.95156172e-01 \end{pmatrix}$$

$$X'y = (454055 \quad 5741)$$

$$\hat{a} = (X'X)^{-1}X'Y = \begin{pmatrix} 77.844 \\ -126.500 \end{pmatrix}$$

$$a_1 = 77.8$$
$$a_2 = -126.5$$

$$R^2 = 0.83$$