ACS234
Maths and Data Modelling

Tutorial 2
Wednesday 1pm LT04

https://github.com/ineskris/ACS234/tree/master/Tutorial2

Done in Lecture (week 3 - week 4)

- Newton Interpolation
- Simple Linear Regression - Least Squares
- Introduction Multiple Linear Regression

Newton Interpolation

Exercice 1

The data

x	-5	0	4	5
f(x)	2	1	10	1

a) Write the cubic interpolating polynomial in the Newton form.

b) Can you write a Matlab and a Python code to solve this problem and check your results.

Simple Linear Regression

Simple linear regression allows us to study the relationship between only two variables.

Model $y = a_0 + a_1 x + e$

Prediction $\hat{y} = \hat{a}_0 + \hat{a}_1 x$

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

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

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

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

For a simple linear regression $m = 1$

Exercice 2 - Calculate the coefficient of determination and the standard error of estimate of this dataset with the model $y = 3.1 - x$

x	-5	0	5	10
f(x)	10.1	3.5	-1.2	-6.8

Exercice 2bis

a) What model is the best to use for the dataset below (calculate the Mean square Error)

$$y = -2 + x \qquad \textbf{Or} \qquad y = -3 + 2x$$

Height	1	3	11	13
Mass	-1	3	17	21

b) We can find the exact model that minimises the MSE.

We need to find the two coefficients α and β for the model $y = \alpha + \beta x$.

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Calculate these two coefficients with the dataset above ?

Calculate the coefficient of determination for the model selected in (a) and the model defined in (b). Comment the result.

Multiple Linear Regression

$$y = a_0 + a_1x_1 + a_2x_2 + \ldots + a_mx_m + e$$

Response

Coefficients

Variables

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.