Assignment 1

1) Let X1, X2 and X3 be the three different vectors and each one has 6 samples. Find the correlations between the variable vectors: X1 and X2, X1 and X3, X2 and X3 using correlation coefficient approach. The sample values of three vectors are given as

X1 = {2.5, 3.6, 1.2, 0.8, 4.0, 3.4},
X2 = {1.2, 1.0, 1.8, 0.9, 3.0, 2.2},
X3 = {8.0, 15.0, 12.0, 6.0, 8.0, 10.0}

2) *k* nearest neighbour (called *X*) has been used to estimate the price of the houses based on three different variables which are the number of rooms, size of house in m^2, and age of houses. The training set is given below which is used to train the algorithm. Use the test dataset to estimate the price of the houses based on the *k* nearest neighbour algorithm (for *k*=1 and *k*=2).

Training dataset

Sample No.	y: Price of the house	x_1: Number of rooms	x_2: Size of house (m^2)	Age of House (year)
1	500.000 $	2	45	25
2	800.000 $	3	65	30
3	1.000.000 $	6	100	40
4	350.000 $	2	30	20
5	100.000 $	2	25	20

Testing dataset

Sample No.	y: Price of the house	x_1: Number of rooms	x_2: Size of house (m^2)	Age of House (year)
1	?	4	100	25
2	?	1	60	20

IMPORTANT

* The submission deadline of your results is ***19th of February until 23:00***. You can solve questions manually or use PC.

*Please send your document including **your answers (show how to solve the questions), your name** and **personal number** to my email address: hku@bth.se

* The outcome of your results will be **pass** or **fail**.