Mini Project # 1 **Due: Sep. 10/2025**

1.

a. A system has generated outputs marked by vector 'd' in response to inputs which are listed in vector 'x':

d = [6.0532, 7.3837, 10.0891, 11.0829, 13.2337, 12.6710, 12.7972, 11.6371] ;
x = [1 , 1.7143, 2.4286 , 3.1429, 3.8571, 4.5714, 5.4857, 6] ;

Use the theory of regression to fit a line to this data. Measure the cost function defined as the mean of squared errors. Plot your data points and the line that models the system's function.

b. Use the same data but this time fit a second order polynomial to these data points. What's the value of your cost function? Plot the second order curve.

c. Increase the order of the polynomial to 6 and fit the curve. How much error do you measure this time? Plot the 6^{th} order curve.

d. Remove one data point (d: 12.7772, x: 5.2857) from your set. Once again fit the 6^{th} order polynomial. After finding the polynomial bring that data point back and once again measure the cost function for all data points. How much the value of the cost function changed compared to the 6^{th} order polynomial where all data points were used for curve fitting? Is the 6^{th} order polynomial a case of over fitting? Plot the new 6^{th} order curve.

e. Plot the value of the cost function as a function of the polynomial order (from 1 to 10) using all data points. Based on this curve, which order is suitable for this dataset to avoid over or under fitting?

2.

Consider a 2-dimensional classification dataset with the given desired values for each point.

x1 = [0.5, 0.8, 0.9, 1.0, 1.1, 2.0, 2.2, 2.5, 2.8, 3.0];
x2 = [0.5, 0.2, 0.9, 0.8, 0.3, 2.5, 3.5, 1.8, 2.1, 3.2];
d = [0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 1.0, 1.0, 1.0, 1.0];

Here for (x1=0.5, x2=0.5) the desired value is 0.0 and for (x1=3.0, x2=3.2) the desired value is 1.0.

a. Use the logistic regression algorithm to design a supervised classifier that can perfectly separate these two sets. Plot these data points and your decision line.

b. Increase the order of the decision making curve to 2 and once again use the logistic regression to design the classifier. How do you compare these two classifiers?