Assignment 1

Note:

1. Solve all the problems using MATLAB or any other computer language.
2. After calculating the best model given your data, review the given data and you may remove any possible outliers (a value with possible error or high noise), then recalculate the model for cleaned data. Compare the two models. This exists only in problems 1 and 5.
3. Try by hand the linear and the logistic problems as well. You may use an excel sheet for this calculation.
4. In all your answers in each question, write down the equations of your solutions (after calculating its parameters).
5. Hint: it is better to take the model of lower number of parameters if the gain in R^2 is not high enough.

1. The numbers of insured persons y with an insurance company for the years 1987 to 1996 are shown in the table.

Year	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
y	13000	12700	12400	12000	1150	11000	10300	9800	9500	9100

Make a scatterplot of the data, letting x represent the number of years since 1987.

a) Fit linear, quadratic, cubic, by comparing the values of R^2. Determine the function that best fits the data. (Hint: take care of note 4 above)
b) In all your answers in each model, write down the equations of your solutions (after calculating its parameters).
c) Graph the function of best fit with the scatterplot of the data.
d) With the best function found in part (b), predict the average number of insured persons in 1997.

2. As Earth's population continues to grow, the solid waste generated by the population grows with it. Governments must plan for disposal and recycling of the growing amounts of solid waste. Planners can use data from the past to predict future waste generation and plan for enough facilities for disposing of and recycling the waste.

   Given the following data on the waste generated in Florida from 1990-1994, how can we construct a function to predict the waste that was generated in the years 1995-1999?

Year	Tons of Solid Waste Generated (in thousands)
1990	19,500
1991	19,700
1992	20,500
1993	21,700
1994	23,800

a) Fit linear, quadratic, cubic, and power functions to the data. By comparing the values of R^2, determine the function that best fits the data.
b) In all your answers in each model, write down the equations of your solutions (after calculating its parameters).
c) Graph the function of best fit with the scatterplot of the data.
d) With the best function found in part (b), predict the average tons of waste in 1996.

3. The following data was obtained by throwing a rubber ball.

Time (sec)	Height (m)
0.0000	1.03754
0.1080	1.40205
0.2150	1.63806
0.3225	1.77412
0.4300	1.80392
0.5375	1.71522
0.6450	1.50942
0.7525	1.21410
0.8600	0.83173

a) Fit linear, quadratic, cubic, and power functions to the data. By comparing the values of R^2, determine the function that best fits the data.
b) In all your answers in each model, write down the equations of your solutions (after calculating its parameters).
c) Graph the function of best fit with the scatterplot of the data.
d) Determine the maximum height of the ball (in meters).
e) With the model you selected in part (b), predict when the height of the ball is *at least* 1.5 meters.

4. Develop a model for estimating heating oil used for a single-family home in the month of January based on average temperature and amount of insulation in inches.

Oil	Temp F	Insulation
275	40	4
360	27	4
160	40	10
40	73	6
90	65	7
230	35	40
370	10	6
300	9	10
230	24	10
120	65	4
30	66	10
200	41	6
440	22	4
323	40	4
50	60	10

a) Fit linear, quadratic functions to the data. By comparing the values of R^2, determine the function that best fits the data.
b) In all your answers in each model, write down the equations of your solutions (after calculating its parameters).
c) Then use the regression models for the functions in b to predict the needed oil if the temperature is 10 Fahrenheit and the insulation is 5 attic insulations inches.
d) What is your recommendation for the company?
e) You may review the data and remove what is outside the reasonable range (outlier), then recalculate the results and compare.

Logistic Regression Problems

5. *Effect of Advertising* A company introduces a new software product. They advertised the product on television and found the following data relating the percent P of people who bought after x ads.

Number of Ads, x	% Who Bought, P
0	0.2
10	0.8
20	2.6
30	9.1
40	26
50	55.6
60	81.3
70	91.8
80	98.5
90	99.5

a) Fit linear, quadratic and logistic functions to the data. By comparing the values of R^2, determine the function that best fits the data.
b) In all your answers in each model, write down the equations of your solutions (after calculating its parameters).
c) Then use the regression models for the functions in b to predict the percent P of people who will buy the software after 100 ads. Comment on the different results.
d) What is your recommendation for the company?