1. Consider the following dataset.

x	y
12	240
23	1135
34	2568
45	4521
56	7865
67	9236
78	11932
89	14589
123	19856
134	23145

   (a) Fit a linear regression model, two polynomial regression models (one with degree 2 and the other with degree 3) on this data set by using python in-built functions.

   (b) Compare the performance of the three models by comparing their SSE and coefficient of determination R^2.

2. Solve question 1 by writing your own functions for

   (a) the regression coefficients using the formula $\alpha = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{XY})$.

   (b) SSE using the formula $SSE = \sum\limits_{i=1}^{m}(y_i - \hat{y}_i)^2$ and R^2 using the formula $R^2 = 1 - \frac{SSE}{SST}$.

3. Download the Pima dataset from Kaagle. This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases. The objective of the dataset is to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset. Several constraints were placed on the selection of these instances from a larger database. In particular, all patients here are females at least 21 years old of Pima Indian heritage.

   (a) Form a classifier that classifies a patient into one of the two classes (WITH DIABETES, WITHOUT DIABETES) by designing a logistic regression model using this data.

(b) Study the variations on the model by varying (Training Data,Testing Data) ratio to be $(0\%, 20\%)$, $(70\%, 30\%)$ and $(60\%, 40\%)$.

4. Download the Position-Salaries (which is shown in the class) dataset from kaggle. Fit a linear regression model and 3 polynomial regression models with degree 2, 5 and 11 on this data set. Compute the bias and variance of all the models by using the mixtend module in python. Give a report on the issues of underfitting and overfitting in each of these models.

5. Fit a linear regression model on the dataset given in question 1 by finding the regression coefficients using the gradient descent algorithm. You are supposed to implement a function for the gradient descent algorithm that takes the dataset as input and returns the coefficients of the regression model. The inital values of the parameters may be hardcoded within the program. Compare the parameters obtained with those given in question 1.