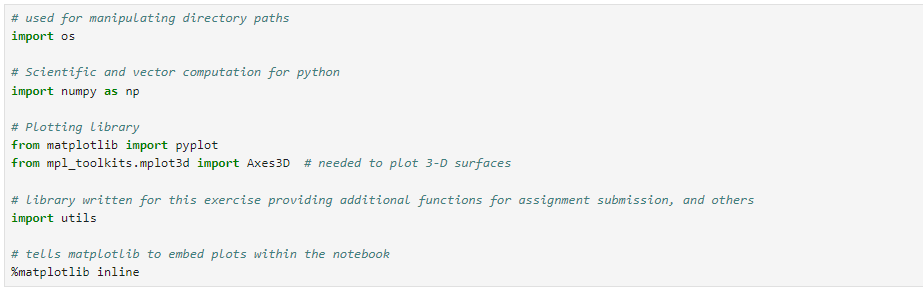
Programming homework 1: Linear Regression

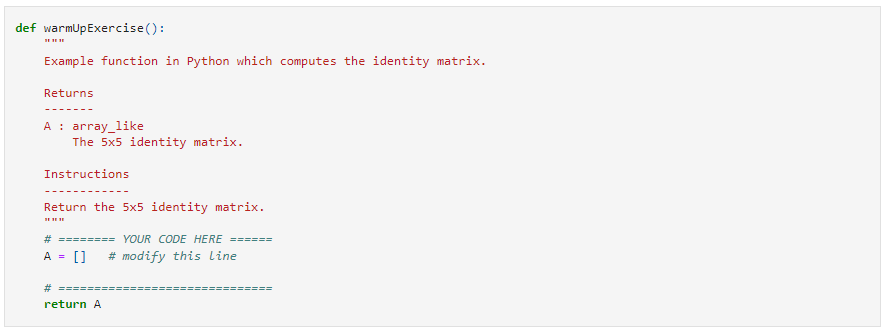
In this homework, you will implement linear regression and get to see how it works on data.

Before we begin with the exercises, we need to import all libraries required for this programming exercise. Throughout the course, we will be using numpy for all arrays and matrix operations, and matplotlib for plotting.



You can find instructions on how to install required libraries in the README file

## **1 Simple python and numpy function (10 points)**

The first part of this assignment gives you practice with python and numpy syntax and the homework submission process. In the next cell, you will find the outline of a python function. Implement the warmUpExercise() and Modify it to return a 5 x 5 identity matrix. 



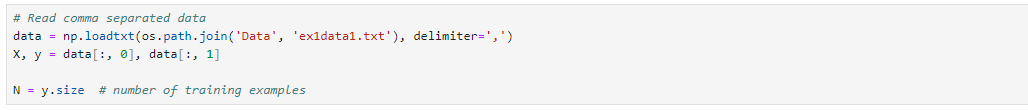
**Submission for part 1: Screenshot your code and the output**

## **2 Linear regression with one variable**

Now you will implement linear regression with one variable to predict profits for a food truck. Suppose you are the CEO of a restaurant franchise and are considering different cities for opening a new outlet. The chain already has trucks in various cities and you have data for profits and populations from the cities. You would like to use this data to help you select which city to expand to next.

The file Data/ex1data1.txt contains the dataset for our linear regression problem. The first column is the population of a city (in 10,000s) and the second column is the profit of a food truck in that city (in $10,000s). A negative value for profit indicates a loss.

The code needed to load this data is already provided. The dataset is loaded from the data file into the variables x and y:

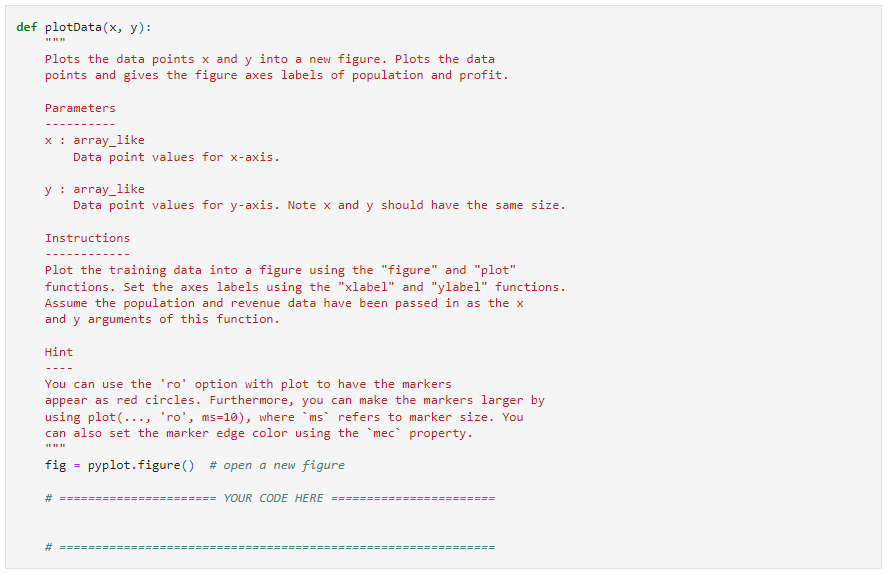


### **2.1 Plotting the Data**

Before starting on any task, it is often useful to understand the data by visualizing it. For this dataset, you can use a scatter plot to visualize the data, since it has only two properties to plot (profit and population). Many other problems that you will encounter in real life are multi-dimensional and cannot be plotted on a 2-d plot. There are many plotting libraries in python (see this [blog post](https://blog.modeanalytics.com/python-data-visualization-libraries/) for a good summary of the most popular ones).

In this course, we will be exclusively using matplotlib to do all our plotting. matplotlib is one of the most popular scientific plotting libraries in python and has extensive tools and functions to make beautiful plots. pyplot is a module within matplotlib which provides a simplified interface to matplotlib's most common plotting tasks, mimicking MATLAB's plotting interface.

In the following part, your first job is to complete the plotData function below. Modify the function and fill in the necessary code to plot the data so that y-axis display the “profit in $Profit in $10,000” and x-axis displays the “Population of City in 10,000s”





To quickly learn more about the matplotlib plot function and what arguments you can provide to it, you can type ?pyplot.plot in a cell within the jupyter notebook. This opens a separate page showing the documentation for the requested function. You can also search online for plotting documentation.

To set the markers to red circles, we used the option 'or' within the plot function**.**

****

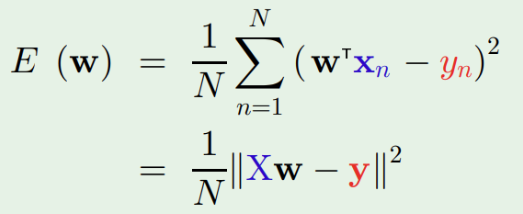
**Submission for part 2.1: Screenshot your code of plotData function and the output figure that the function plotted.**

### **2.2 Computing the optimal weights for the linear regression model**

Back to the slides content of linear regression, we calculate the credit line for a user by using:



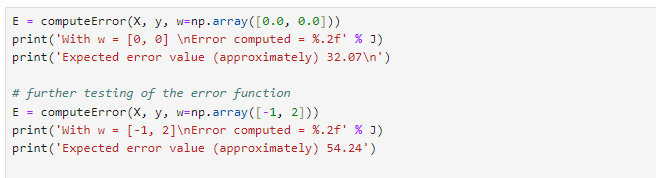
And then we compare the regression model predicted credit line h(x) with the real credit line given by the bank officer y. The difference E of h(x) and y is the mean square error caused by the model, our target is to minimize the E so that the h(x) can fit for all the historical data.



Your next task is to complete the code for the function computeError which computes E(w), As you are doing this, remember that the variables X and y are not scalar values. X is a matrix whose rows represent the examples from the training set and y is a vector whose each element represent the value at a given row of X

.

Once you have completed the function, the next step will run computeError two times using two different initializations of w. You will see the Error printed on the screen.



**Submission for part 2.2: Screenshot your code of computeError function and the output.**