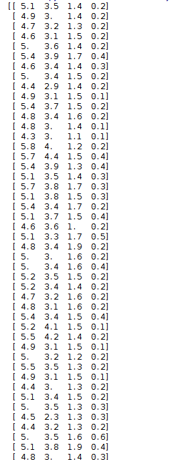
**CS5542 – Big Data Analytics and Apps**

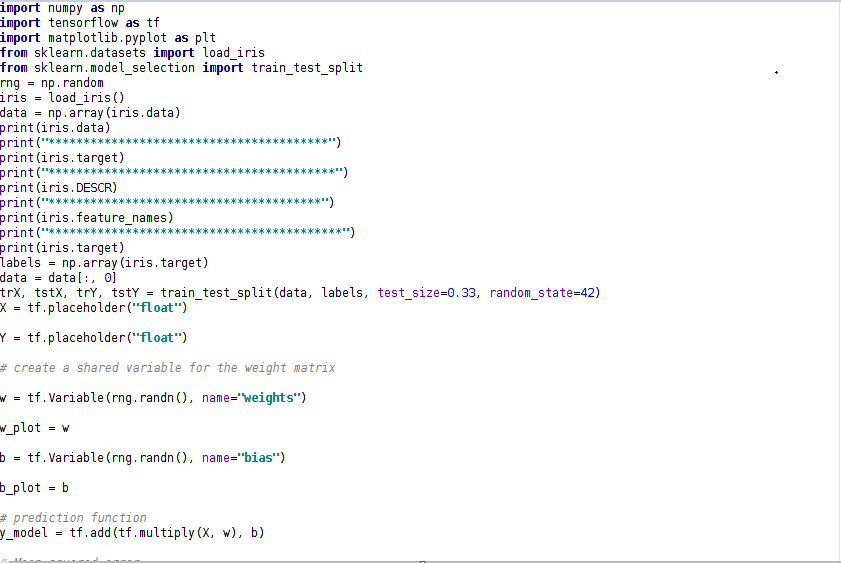
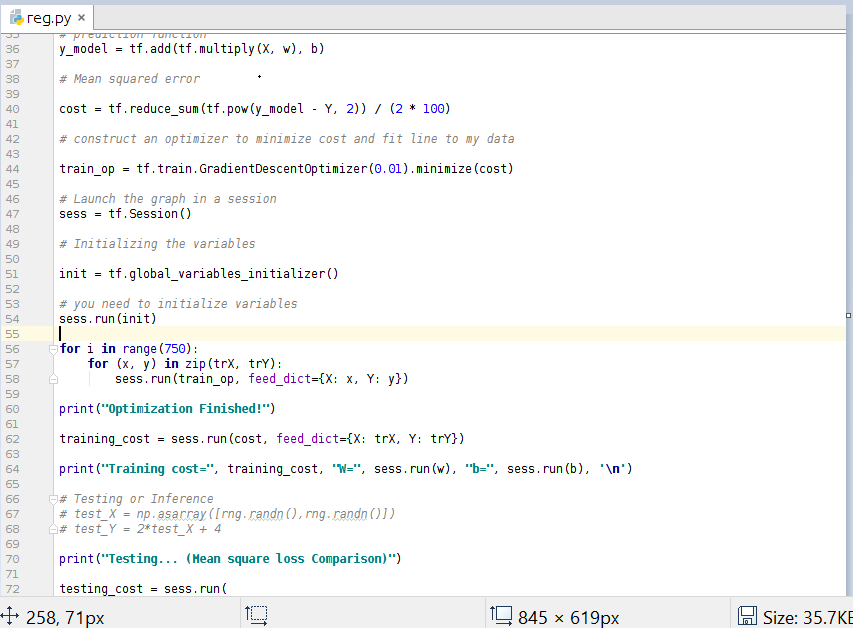
**Lab 7 – Assignment Submission**

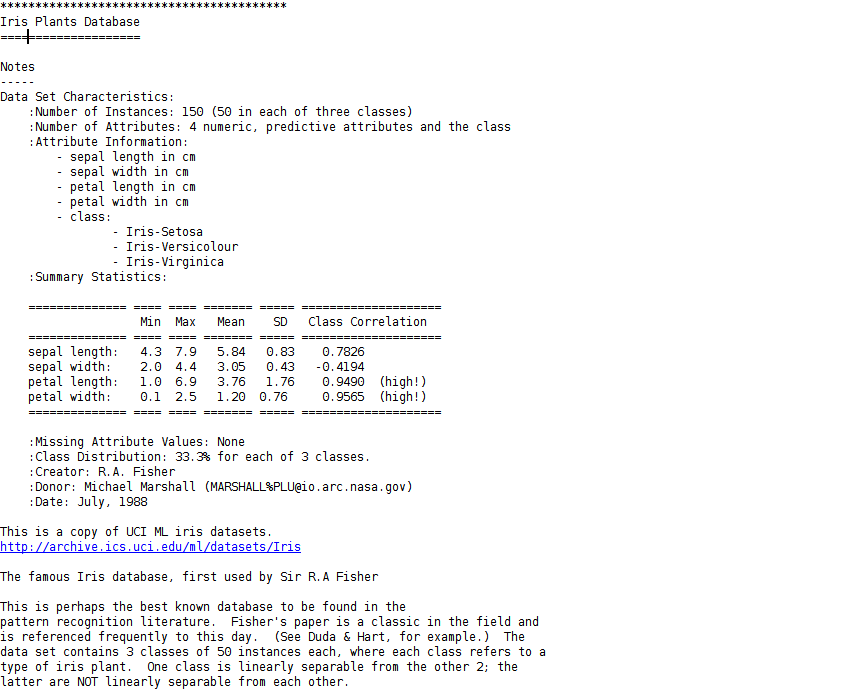
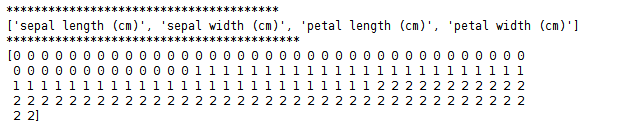
Name: Lava Kumar S

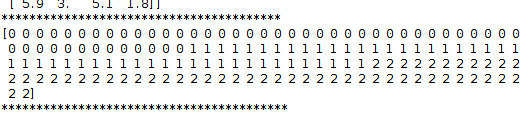
Class ID : 38

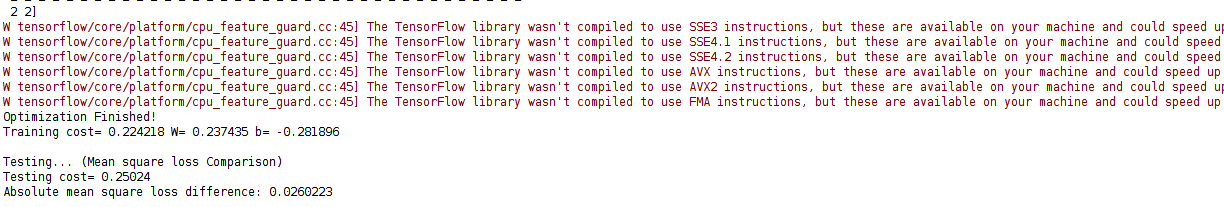
**Task 1**  
**DataSet**  
The Iris Dataset This data sets consists of 3 different types of irises' (Setosa, Versicolour, and Virginica) petal and sepal length, stored in a 150x4 numpy.ndarray. The rows being the samples and the columns being: Sepal Length, Sepal Width, Petal Length and Petal Width. The below plot uses the first two features.  
  


The below program is written for linear regression and plotting a line.

**Program**  
  
  
  


**Dataset Description**  
  
After loading the dataset the description of the datase is obtaoined as below.  
  
  
**Dataset Features**  
  
Similary the dataset is description is obtained below.  
  


**Dataset Label**  
  
In the training data the label is obtained with the dataset.  
  


**Output After training**  
After training the test data is given and the mean squared error,Training cost and testing cost are obtained.  
  


**Task 2**

Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and IPython shell, the jupyter notebook, web application servers, and four graphical user interface toolkits.

**Matplotlib Graph**  
A plot has been done based on the plotted x and y values that is median value and sepal length.  
  
