ID : 201911027
Manan Gajjar

Lab Report
Lab Assignment 1

Part 1 : Linear Regression (univariant)

- Dataset : Total 400 columns; 340 for Training; 60 for evaluation

	GRE Score	Chance of Admit
0	337	0.92
1	324	0.76
2	316	0.72
3	322	0.80
4	314	0.65

- Data Visualization

- Linear regression was performed using GRE score as the only feature to predict the chance of admission for a student.

- Gradient Descent was implemented from scratch to calculate the slope and bias terms i.e. Θ_0 ,Θ_1 which were initially taken as 0 , 0

- Parameters Taken
  - Learning Rate or Alpha : 0.1
  - Number of Iterations or Epochs : 1000

- Equation used to fit data :

$$Y = \Theta_0 + \Theta_1 * X$$

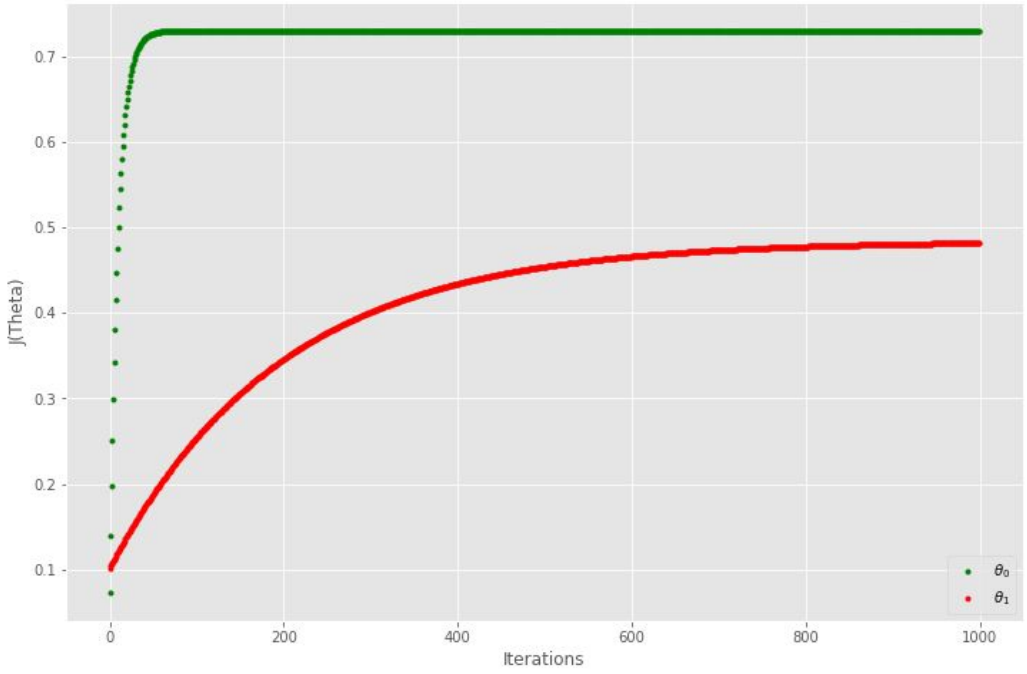
- Means Square Error was taken as Cost function which had to be minimized

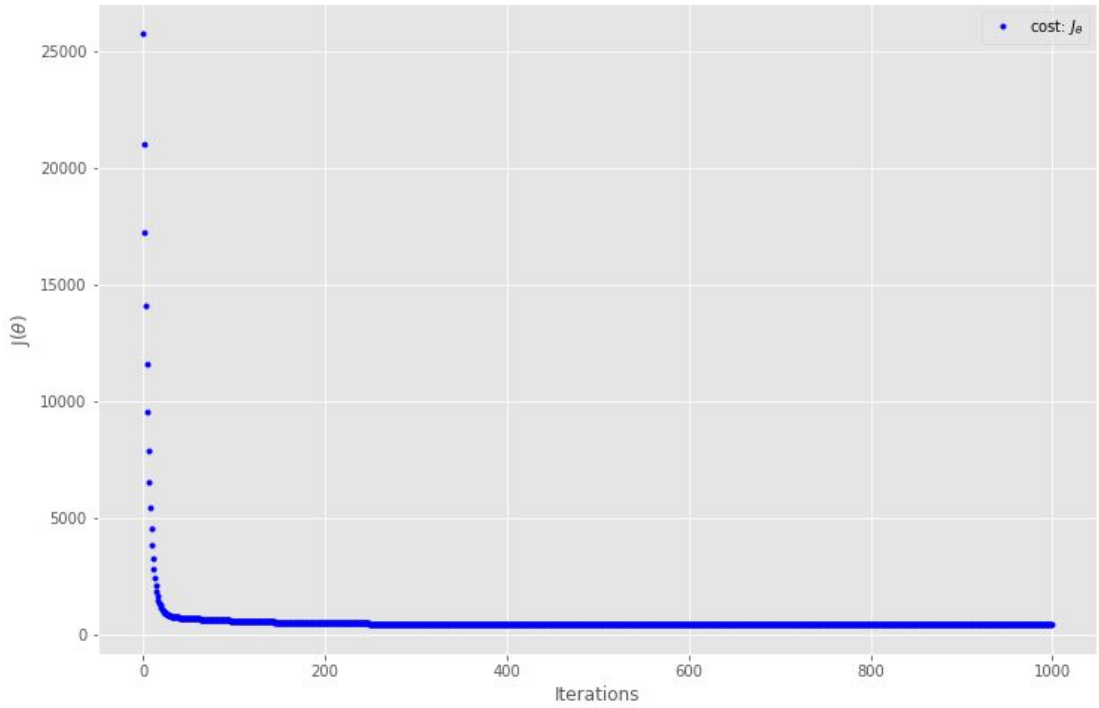
- Optimal Values after 1000 Epochs were :
  - M or Slope or Theta1 : 0.482
  - C or Bias or Theta0 : 0.730
  - Least Error : 418.472

- Prediction for New Data :

```
GRE Score :  340
Chance Of Admision :  [0.67976141]  Error :  [-0.07023859]
GRE Score :  390
Chance Of Admision :  [0.6990253]  Error :  [0.0590253]
```

- Convergence Graph :





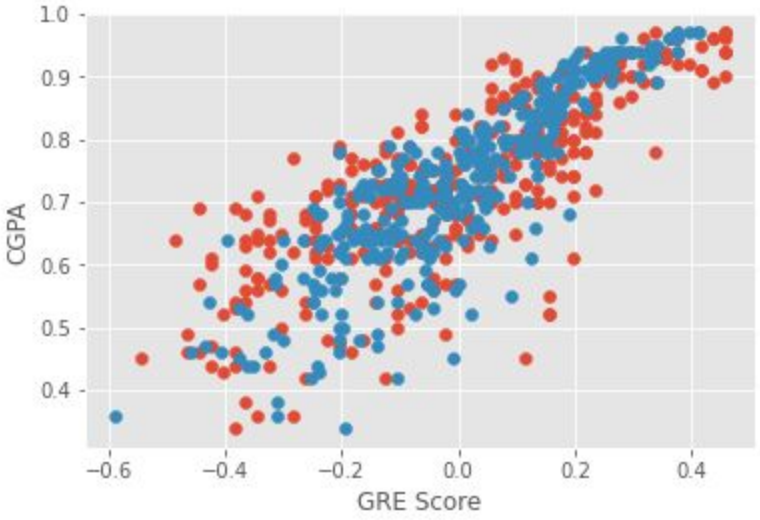
Part 2 : Linear Regression (multivariant)

- Dataset : Total 400 columns; 340 for Training; 60 for evaluation

	GRE Score	CGPA	Chance of Admit
0	337	9.65	0.92
1	324	8.87	0.76
2	316	8.00	0.72
3	322	8.67	0.80
4	314	8.21	0.65

- 2 variables : GRE score and CGPA
- Feature scaling was performed to converge faster than the normal.

- Data Visualization



- Multilinear regression was performed using GRE score and CGPA as the features to predict the chance of admission for a student.

- Equation used to fit data :

$$Y = \Theta_0 + \Theta_1 * X1 + \Theta_2 * X2$$

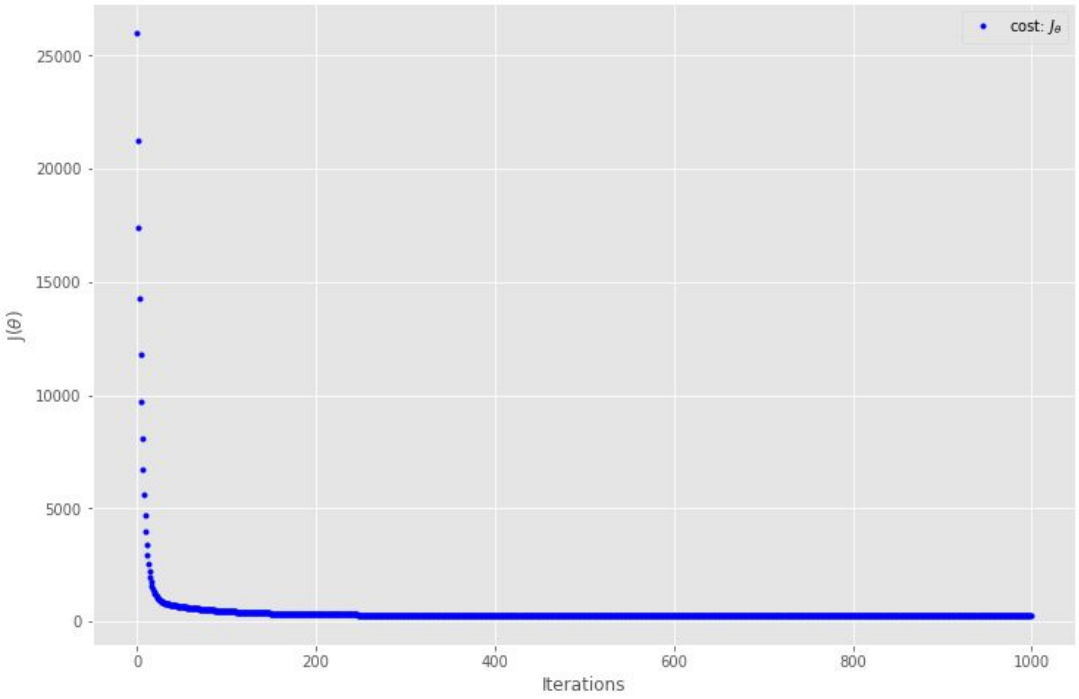
- Means Square Error was used as a Cost function.

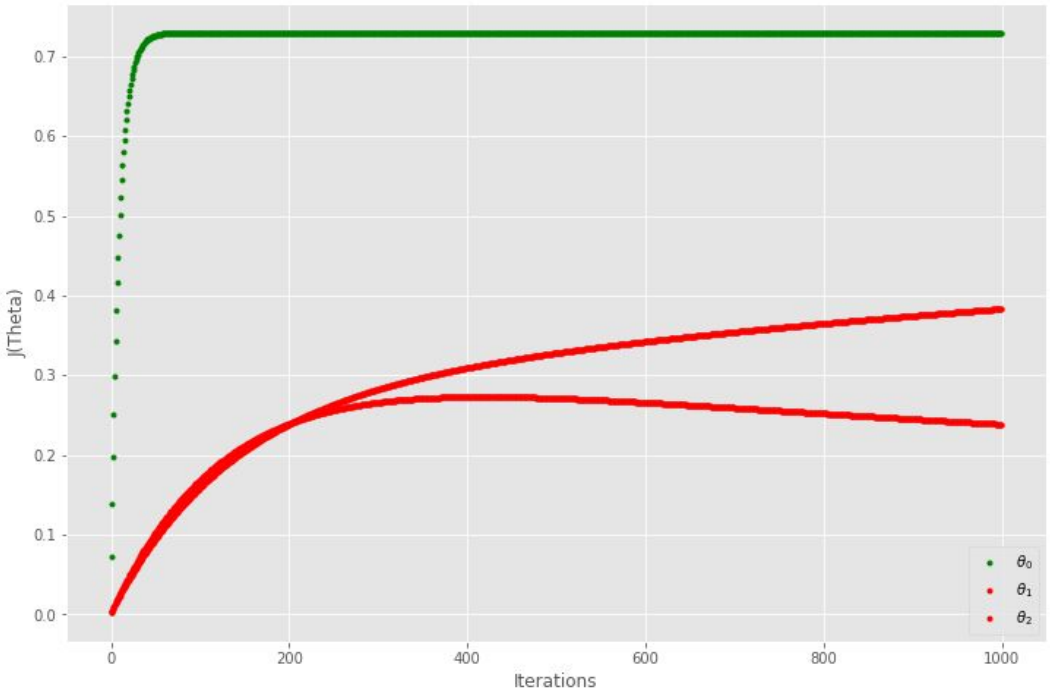
- Gradient Descent was implemented from scratch to calculate the slopes and bias terms i.e. Θ_0 , Θ_1 and Θ_2 which were initially taken as 0 .
- Parameters Taken
  - Learning Rate or Alpha : 0.1
  - Number of Iterations or Epochs : 1000

- Optimal Values after 500 Epochs were :
  - Θ_2 : 0.383
  - Θ_1 : 0.238

- Θ_0 or Bias : 0.730
- Least Error : 270.959

- Convergence Graph :





- Model Performance :

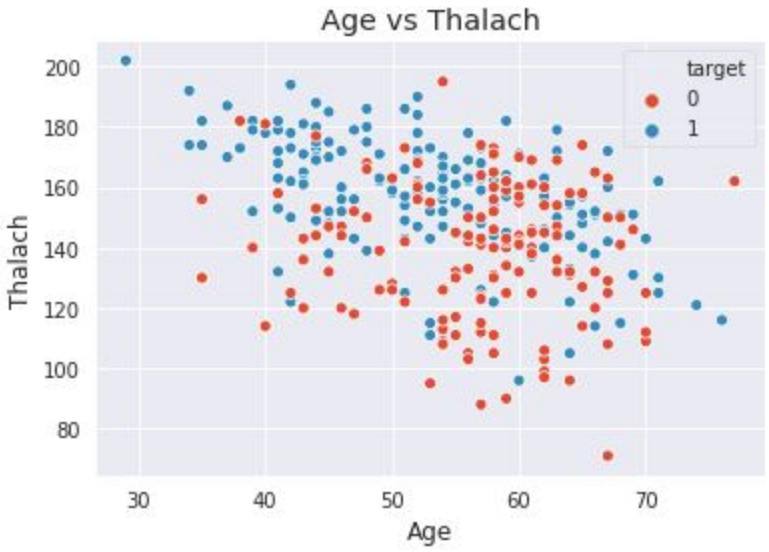
```
GRE Score :  340
Chance Of Admision :  [0.68387467]  Error :  [-0.06612533]
GRE Score :  390
Chance Of Admision :  [0.66639327]  Error :  [0.02639327]
```

Part 3 : Logistic Regression

- Dataset : Total 303 rows

	age	thalach	target
0	63	150	1
1	37	187	1
2	41	172	1
3	56	178	1
4	57	163	1

- Data Visualization



- Feature/Data Normalization was done to converge faster using mean and standard deviation.

- Data after normalization :

```
([[ 0.9521966 ,  0.01544279],
  [-1.91531289,  1.63347147],
  [-1.47415758,  0.97751389],
  [ 0.18017482,  1.23989692],
  [ 0.29046364,  0.58393935],
  [ 0.29046364, -0.07201822],
  [ 0.18017482,  0.1466343 ],
  [-1.1432911 ,  1.0212444 ],
  [-0.26098049,  0.54020884],
  [ 0.29046364,  1.0649749 ]])
```

- Logistic regression was implemented to predict heart attack based on the features “age” and “thalach”.

- Equation used to fit data :

$$Y = \Theta_0 + \Theta_1 * X1 + \Theta_2 * X2$$

$$Sigmoid(Z) = 1 / (1 + e^{y})$$

- Sigmoid function ensures that value is normalized in the range (0, 1)

- Log Loss Function is used as cost function :

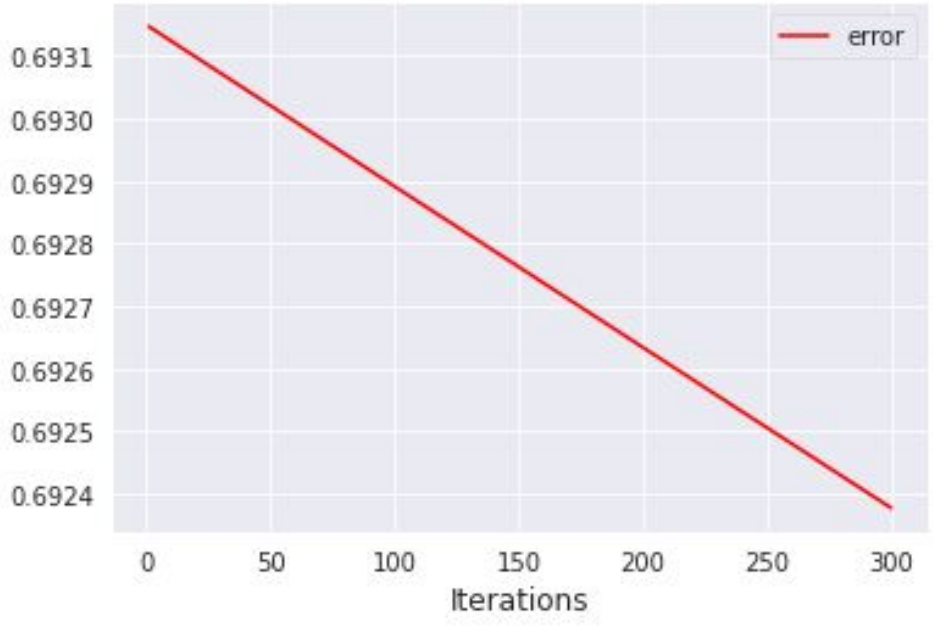
$$-(y \log(p) + (1 - y) \log(1 - p))$$

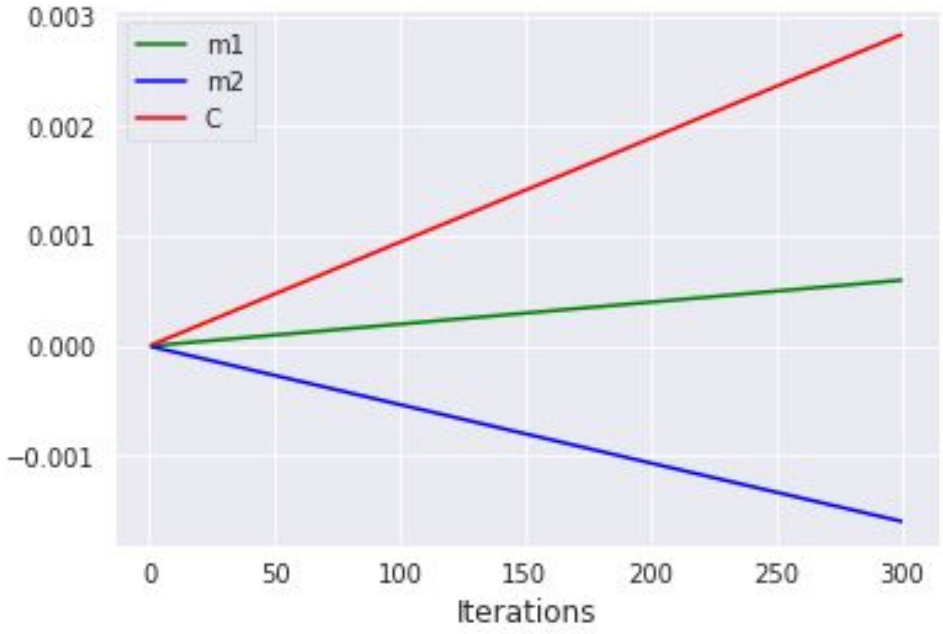
- Gradient Descent was implemented from scratch to calculate the slopes and bias terms i.e. Θ_0 , Θ_1 and Θ_2 which were initially taken as 0.

- Parameters Taken
  - Learning Rate or Alpha : 0.01
  - Number ofIterations or Epochs : 300

- Optimal Values after 300 Epochs were :
  - Θ_2 : 0.00283
  - Θ_1 : -0.00159
  - Θ_0 or Bias : 0.00059
  - Least Error : 0.6923

- Convergence Graph :





- Model Performance
  - Training Accuracy : 0.6745
  - Testing Accuracy : 0.7143

- In-built function's performance
  - Training Accuracy : 0.6934
  - Testing Accuracy : 0.7033

Thank You