Santa Clara University
COEN 240- Assignment 2

Syeda Gousia Sultana- W1587235

Rank: 15
Accuracy: 84.52%
Learning rate: 0.5
Number of iterations: 18000

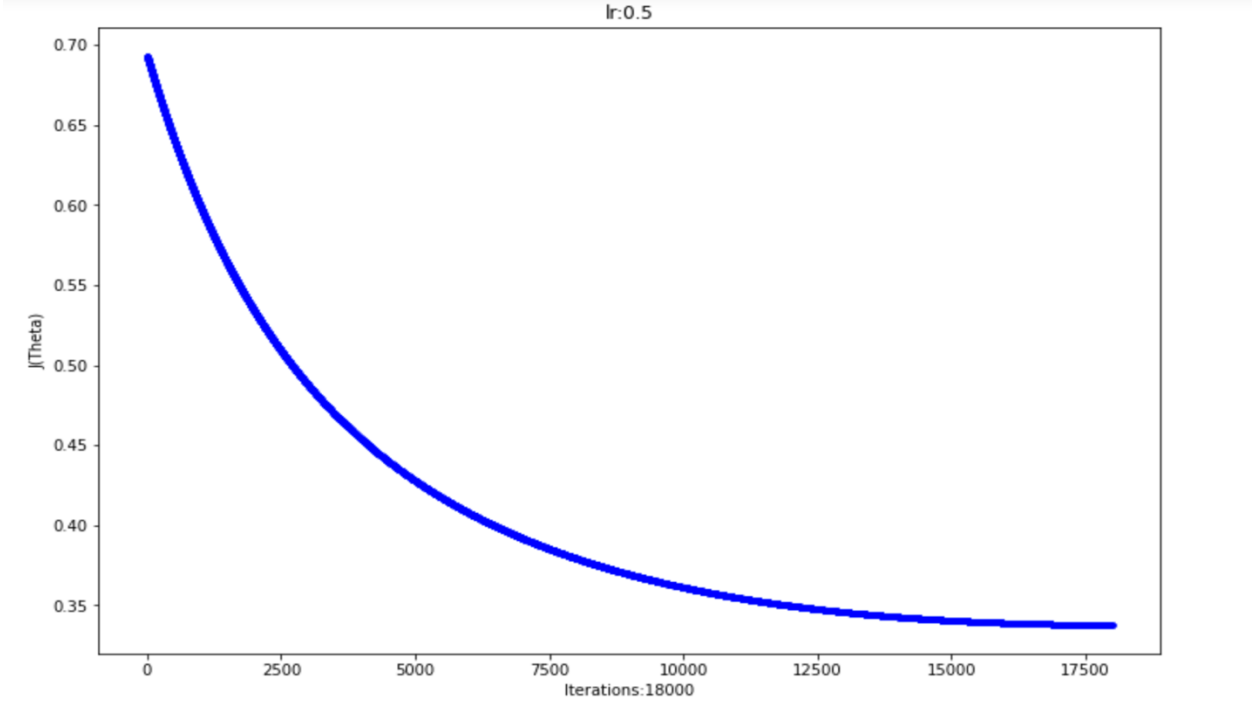
Approach:

I have built a model to fit the training data by keeping track of all the parameters to know at which point is it on the gradient. This was done with the help of cost function, whose values gave a prediction of how far we are from the target. Implemented gradient descent for optimizing the cost function. The prediction function in the program predicted the output for the given test.mat file.

Choosing the approach and associated parameters:

Logistic regression is easier to implement when the data is preprocessed and normalized. The given problem is a classification problem where the target variables or labels are dependent on the input data and is categorical. The associated parameters were obtained through a lot of trial and error.

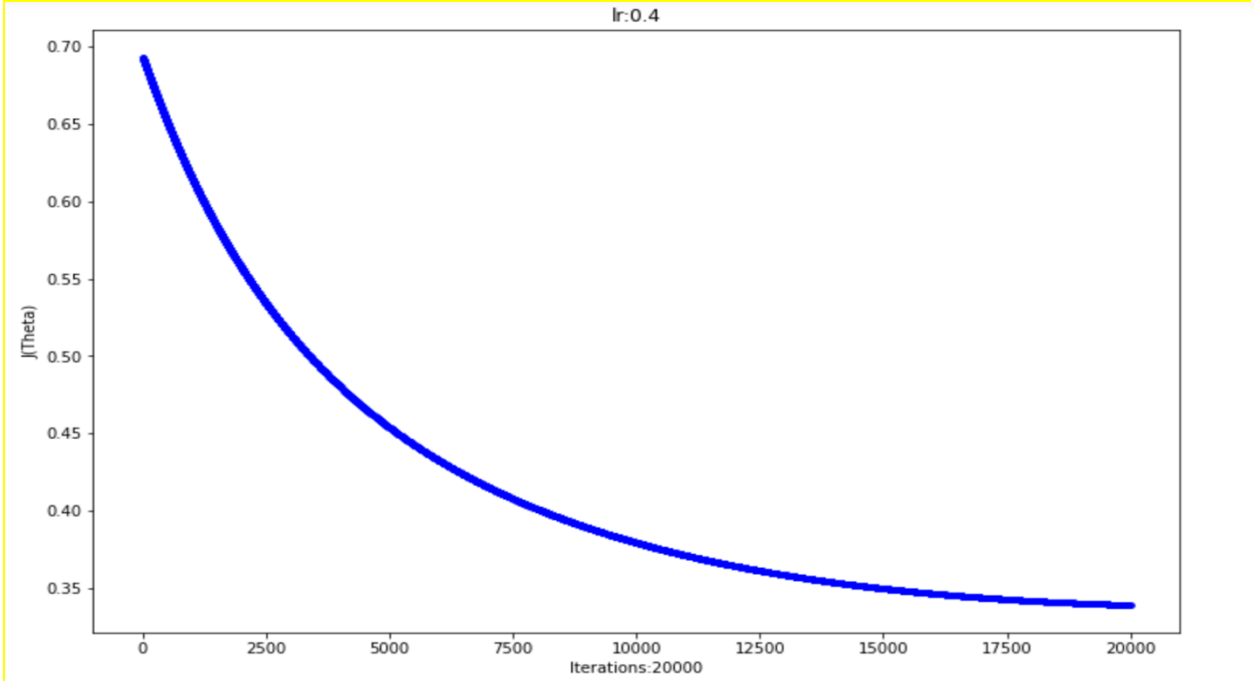
```
Final Cost :  0.3377609193297449
Final Parameters : 40.364893284598985,24.48716079054316,22.54755198047614
```



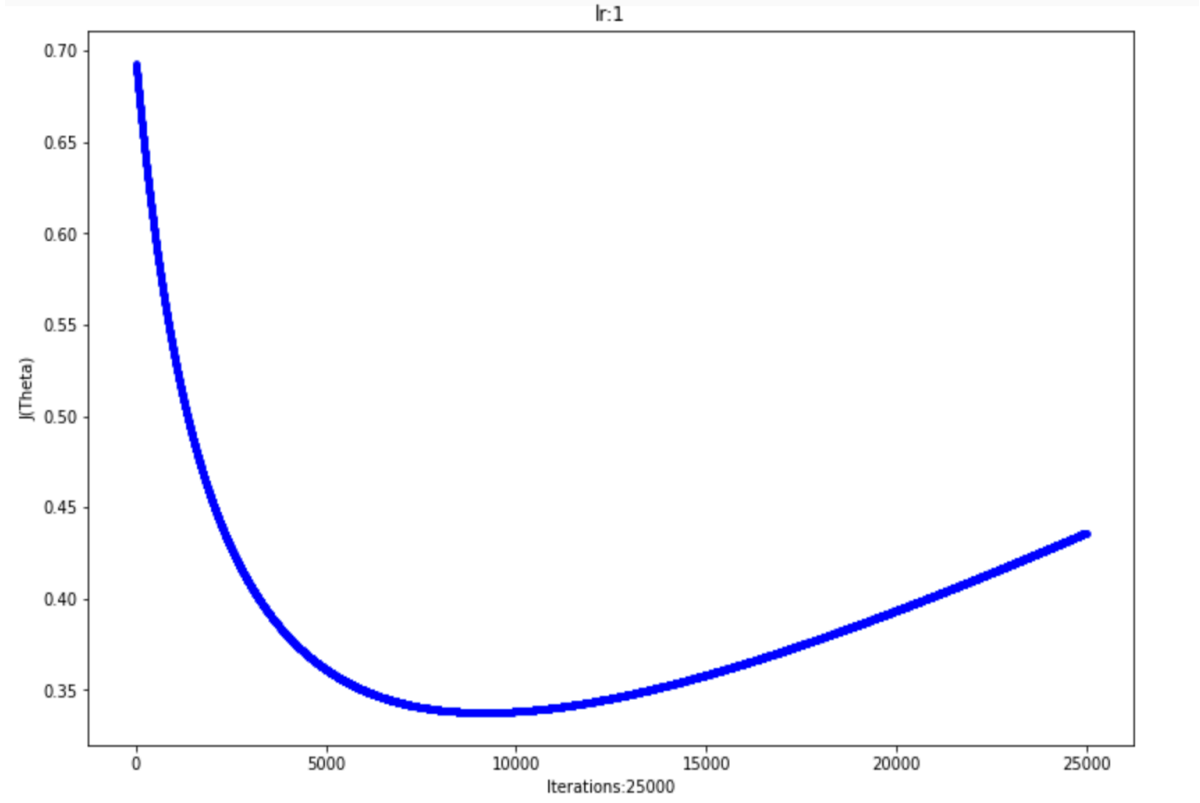
Santa Clara University
COEN 240- Assignment 2

Syeda Gousia Sultana- W1587235

```
Final Cost :  0.3390733275166536
Final Parameters : 37.282933855629196,22.471641612783824,20.569195563404445
```



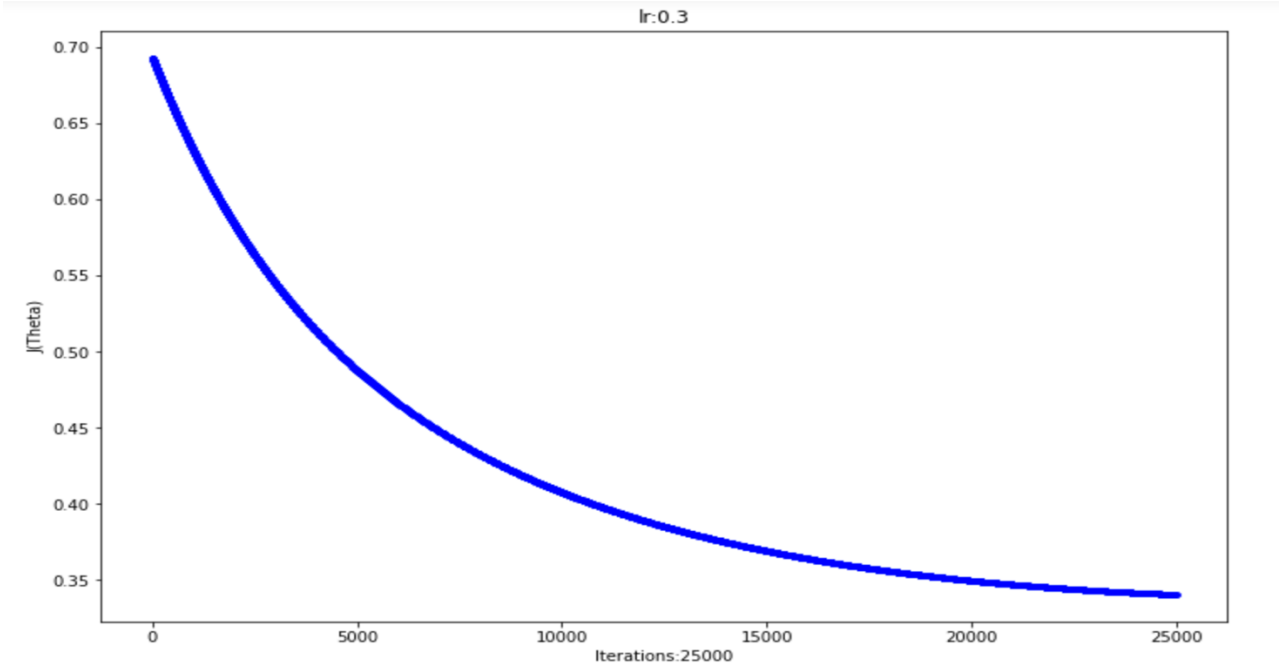

```
Final Cost :  0.43613983472856227
Final Parameters : 85.67359140071895,54.41455139904188,51.68242424104669
```



Santa Clara University
COEN 240- Assignment 2

Syeda Gousia Sultana- W1587235

```
Final Cost :  0.3405941789221778
Final Parameters : 35.709143244211724,21.44588511236427,19.562223214975393
```



The learning rate which controls the steps it takes plays a very important role. If the learning rate is high, it is not converging and producing nan errors. When it is low, it is taking a lot of time to converge. The number of iterations is also equally important, if it is low it stops iterating which leads to ending the program before error is minimized which leads to poor accuracy. If there are a lot of iterations it takes a lot of time and the cost starts increasing after it has reached a particular point. When I used the maximum likelihood cost function there are negative values produced for the cost function and the accuracy remained the same even after all the iterations.

Bias/Variance analysis:

Managing to reduce both bias and variance has lead to a more accurate model. High bias and variance both are bad for a model to generalize. When the model had high bias it lead to underfitting, where as high variance tends to overfitting which gave a high accuracy but it failed to give good accuracy when the new set which is obtained by splitting the train set into train and test set.