

ML Assignment - 2

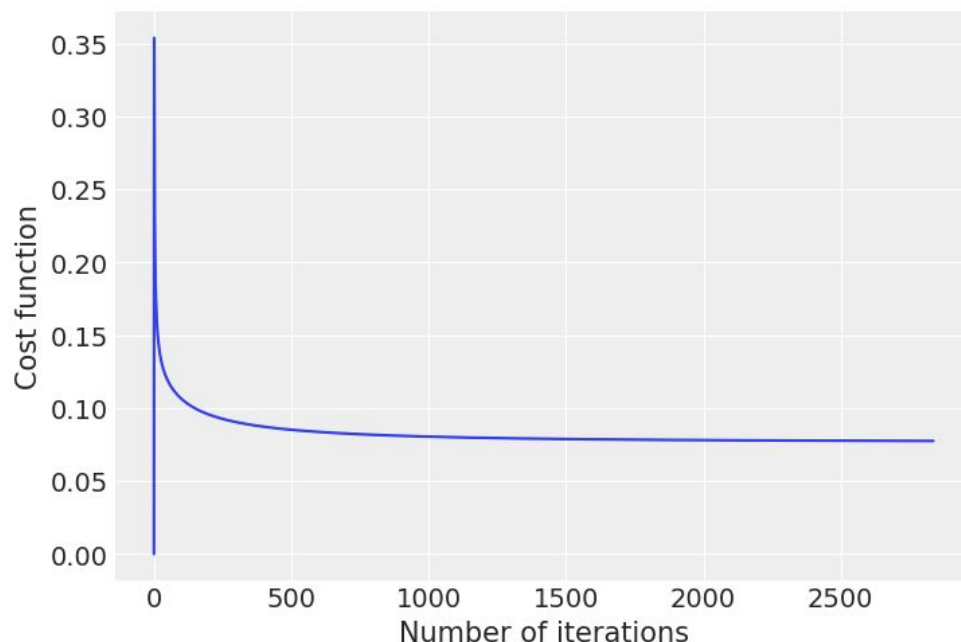
Logistic Regression for IRIS dataset

Abstract: We are given a task to classify given data with two labels. We are also provided with an IRIS dataset. This report includes both the parts of classification by Logistic Regression with stopping criteria and by number of epochs. We got good results while running Logistic Regression with stopping criteria.

Part 1: Run Logistic Regression with stopping criteria

Code: The idea is run as many number of epochs until stopping criteria is met. The code runs 28,00,000 epochs. So, storing cost value at each epoch would result in a memory problem. We just opted to store cost value after 1000 epochs. This helped in plotting results better.

```
In [11]: plot_cost(cost_list)
```



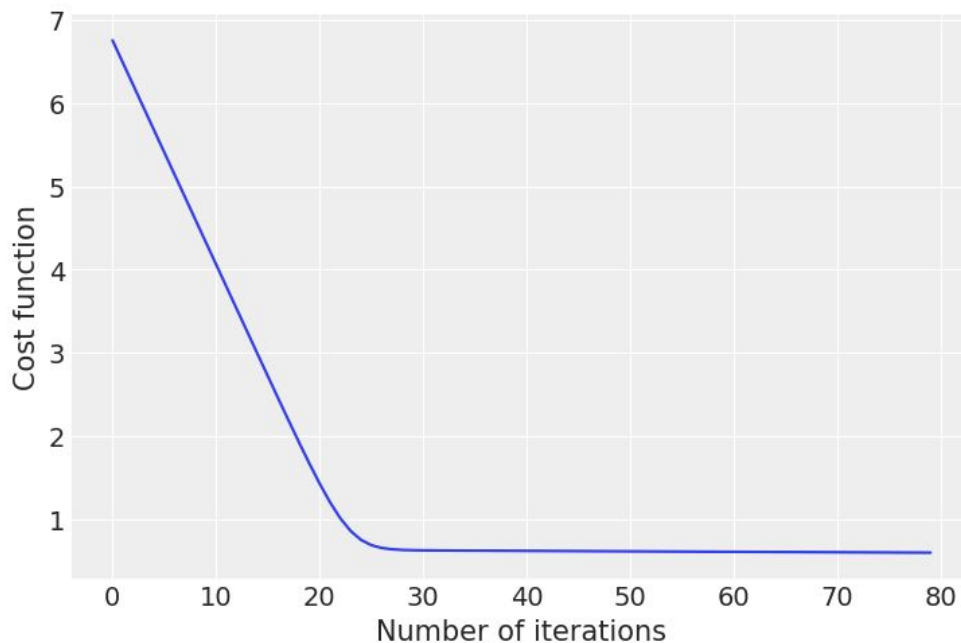
```
In [12]: predict(X_train, y_train, weights, 'train', f'with epsilon={epsilon}')
predict(X_test, y_test, weights, 'test', f'with epsilon={epsilon}')
Accuracy for train set with epsilon=0.01: 97.183 %
Accuracy for test set with epsilon=0.01: 100.000 %
```

Result: We got 97% and 100% accuracies for train and test set respectively. Because this was a two class problem, we got good results since this was a convex optimization problem.

Part 2: Run Logistic Regression with number of epochs

Code: As the name suggests, we run Logistic Regression update rule batch wise by a given number of epochs. The weights vector is initialized to all ones, assuming all features play an identical and important role in classification for the label.

```
In [15]: plot_cost(cost_list)
```



```
In [16]: predict(X_train, y_train, weights, 'train', f'with {epochs} epochs')
         predict(X_test, y_test, weights, 'test', f'with {epochs} epochs')

Accuracy for train set with 80 epochs: 67.606 %
Accuracy for test set with 80 epochs: 62.069 %
```

Observations and Result: Since, we are running less number of epochs, the accuracy is not very high. Also, the accuracy depends on how the weights are initialized. In this case, the weights vector is initialized with all ones and we got 62 % percent accuracies. If we normalize data and add a small value while computing log operations, the accuracy can be increased. The accuracy can still be increased by running for more epochs.

Thank you

Submitted by - Group 9

Sayam Kumar - S20180010158

Atharv Ramesh Dhabekar - S20180010050

Sai Kshitij - S20180010117