

ML_Assignment-2 Report

Team members :-

Vaka Yagna Karthik(S20170010170)

Rayaprolu Sasikanth(S20170010125)

Question-1 :-

Dataset:-

We have chosen 1st 200 patterns form the dYacht Hydrodynamics Data Set for training.The remaing patterns are used for testing.

Algorithm:-

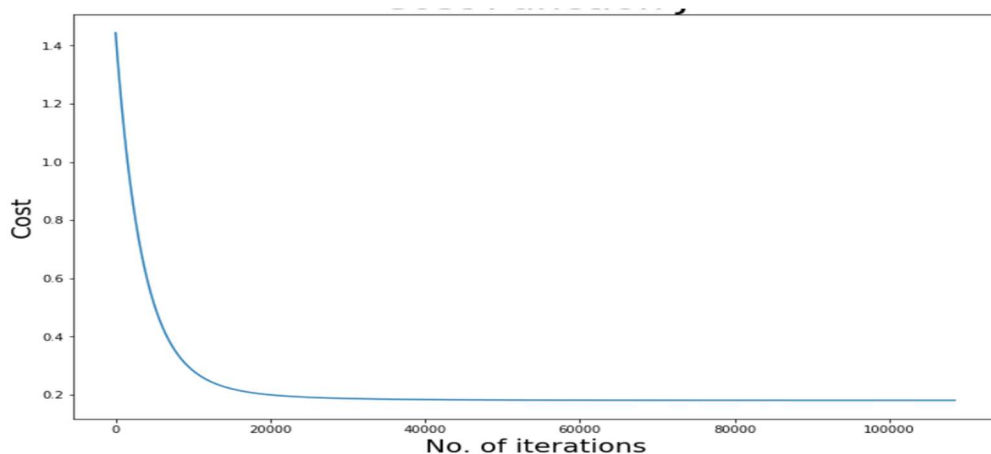
- 1.All the points(i.e the feature sets) are stored as list of lists(which is 'list' in our code).
- 2.All the corresponding output values(residuary resistance per unit weight of displacement, adimensional) for each point is stored in another list at the corresponding indices.
- 3.Now 1 is augmented in the 0th index of every point(to make the dot product easy).
- 4.Shuffling of all the points is done so as to maintain the randomness of the dataset.
- 5.Now, a random weight vector is assumed with all the values in it as '0'('list6' in our code).
- 6.Then the dot product of these two vectors('list'.'list6') is done and all the outputs are stored in

another list namely predicted vector('list7' in our code).

7. Now the error is calculated as the difference of predicted value and original value, these errors are stored in another list ('list8' in our code).

8. The mean square of these errors is calculated, which is the cost function and the weight list is updated.

9. Now these steps are repeated for 'n' (10000 to 100000 in our case) until the optimum accuracy is achieved (i.e. cost function is minimized).



Error :-

Training error – 28 to 55 (due to both shuffling and non-shuffling and due to change in iterations)

Testing error – 31 to 50% (due to both shuffling and non-shuffling and due to change in iterations)

Question-2 :-

Dataset :-

We have stored the iris data set that belongs to Versicolor and Virginica classes, shuffled the data set and then used the 1st 70 patterns for training and remaining 30 patterns for testing.

Algorithm :-

1. We stored all the points (feature sets including the class label) in a list.
2. Then '1' is augmented at the starting index of every point.
3. Now, in these points the class labels 'Iris-versicolor', 'Iris-virginica' are replaced with '1' and '-1' respectively and this entire dataset is stored in 'list3' in our code.
4. Now the dataset is shuffled and the 1st 70 are used for training (list4 in code) and remaining ('list41' in code).
5. Now a Weight vector is assumed with all the initial values as '0'.
6. Dot product between all the feature set and the weight list is done ($\text{list4}[i][j] * \text{list6}[i]$).
7. Now if the i^{th} output is ≥ 0 it is assigned to class '1' else to class '-1' and stored in another list ('list7' in code).
8. Now the difference of list7 and list5 (i.e. original - predicted)

is stored as error in another list('list8' in the code)

9.The mean square of these errors is calculated(cost function) and the weight list is updated.

10.This Algo is repeated until optimal accuracy and minimum error is achieved.(up to 10000 iterations)

11.Then using the weight vector corresponding to minimum error is used to classify the test set and error on the test set is calculated.

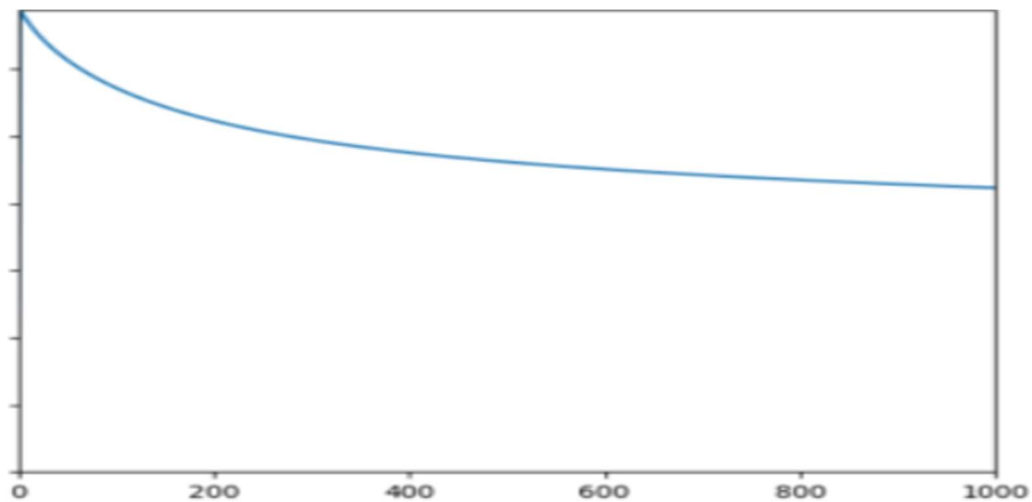


Fig: Cost vs Iterations

```
yagnakarthisk@yagnakarthisk-virtual-machine: ~/Desktop/python_ML
File Edit View Search Terminal Help
yagnakarthisk@yagnakarthisk-virtual-machine:~/Desktop/python_ML$ python3 Group23_Q1.py
total training error function -- 45.4262005777233
total testing error function ....
30.40176160726196
yagnakarthisk@yagnakarthisk-virtual-machine:~/Desktop/python_ML$ python3 Group23_Q2.py
minimum training error function value is 0.34385714285714286
total testing error function ....
3.3333333333333335
yagnakarthisk@yagnakarthisk-virtual-machine:~/Desktop/python_ML$
```

Errors:-

Training error – 0 to 5(due to both shuffling and non-shuffling and due to change in iterations)

Testing error – 0 to 10%(due to both shuffling and non-shuffling and due to change in iterations)

Thank you.....