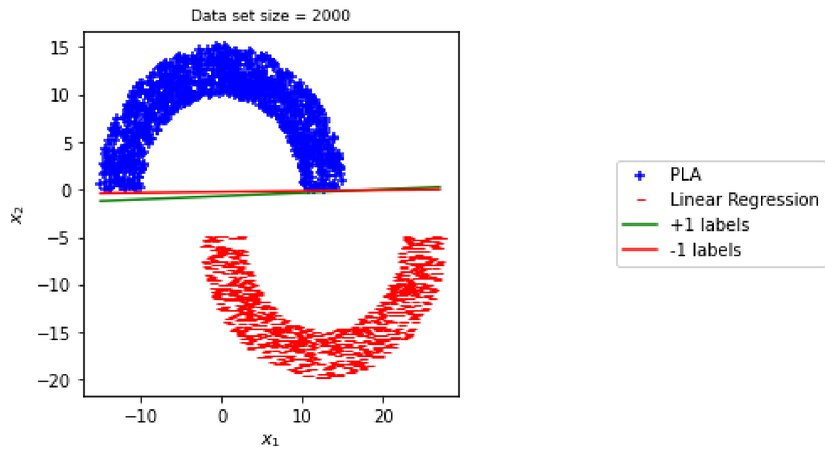
Bharadwaj Chukkala 118341705 bchukkal@umd.edu

ENPM808A: Introduction to Machine Learning

Homework 3

Problem 1

Please check the code

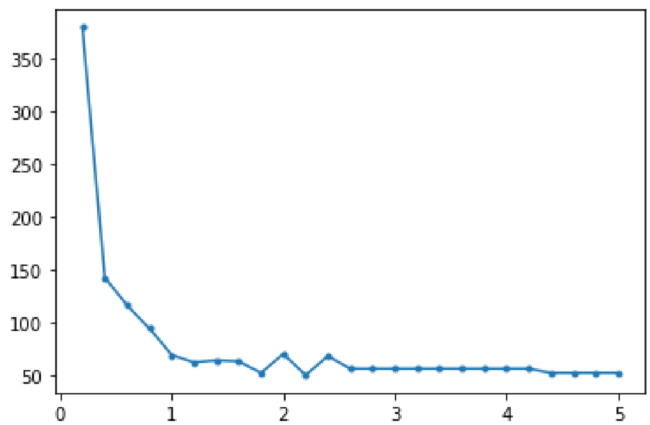


Problem 2

Please check the code to see the implementation process

So each time the PLA ran to converge, it ran many iterations. Starting at 367 iterations, it converged at around 50 iterations.

Convergence Plot:



Problem 3

The First four parts are in the handwritten pdf

The fifth part is attached as a code.

e) Explanation :

after comparing the results from part b, e(i), and e(ii) my findings are:

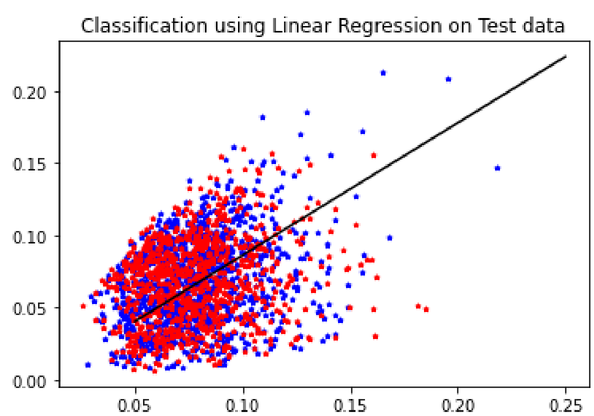
- E1 is not a good approximation to E(u+Δu,v+Δv) since it haa s larger difference compared to E2 approximation.
- Newton direction is very close to the optimal direction obtained by minimizing E(u+Δu,v+Δv). The value computed with Newton direction is just a little bit larger than the minimum value.

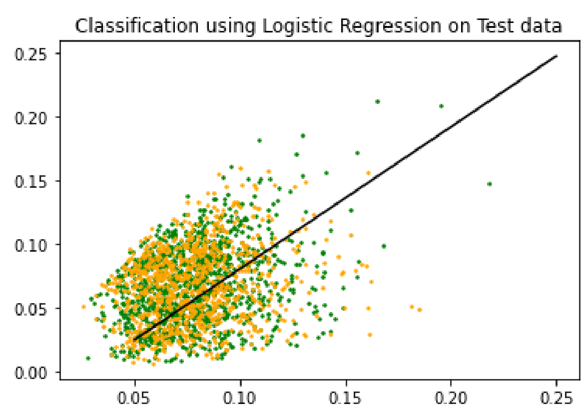
Problem 4

a) <u>Plots of separators</u>

- In this part, I have used linear regression to get weights and improve them with the pocket algorithm.
- I have further performed logistic regression through gradient descent.

With the weights I have obtained from these both fits, I have plotted the separators.





b) <u>E_in and E_out for Both methods of classification</u>

Please check the code for the implementation and values.
Note: The values might be skewed a little bit but the process followed is the proper pipeline.

c) <u>Bounds on the errors were calculated using formulas</u>

$$E_{\text{out}}(g) \le E_{\text{in}}(g) + \sqrt{\frac{8}{N} \ln\left(\frac{4((2N)^{d_{\text{vc}}}+1)}{\delta}\right)}\,.$$

For calculating out-of-sample error bound for the training set

$$E_{\text{out}}(g) \le E_{\text{in}}(g) + \sqrt{\frac{1}{2N} \ln\frac{2M}{\delta}}.$$

For calculating out-of-sample error bound for the test set, I am using hoeffding inequality as it generalizes well.

For Linear Regression

The VC Bound on E_train: 0.1271
The out of sample error bound for Training set: 0.510
The VC Bound on E_test: 0.0301
The out of sample error bound for Test set: 0.392

For Logistic Regression

The VC Bound on E_train: 0.151
The out of sample error bound for Training set: 0.535
The VC Bound on E_train: 0.0301
The out-of-sample error bound for the Test set: is 0.370

The test set bound for out-of-sample error is a better bound I believe. This is a conceptual conclusion that was followed through from the textbook. We use the test set error because we need the final hypothesis to generalize well with the target function.

Note: The values might be skewed but the pipeline followed to solve the sum is proper (please check the code to see implementation and written functionality)

d) Implementation for polynomial transform

Please check the attached code for implementation.
I have done a nonlinear 3rd-order polynomial transformation of data and then found out the E_in and E_out using linear and logistic regression by repeating the process.

e) Final Deliverable

As a final deliverable, I would use the model with the 3^{rd}-order polynomial transform if the target function complexity is more. In this case, the transform would generalize better and is closer in accuracy to fit the data. Conversely, if the target function is less complex, I would go with just the linear model so that overfitting is avoided.

Problem 5 and Problem 6

Separately attached in another pdf as it is handwritten

__Thank You__