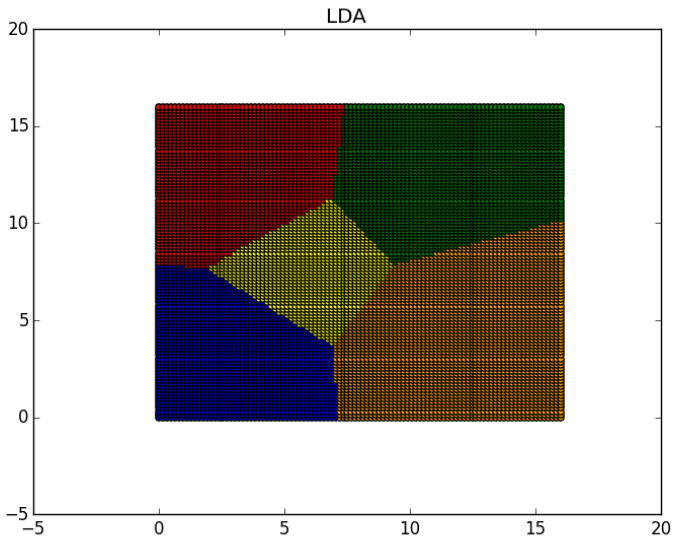
PA-2 Classification and Regression

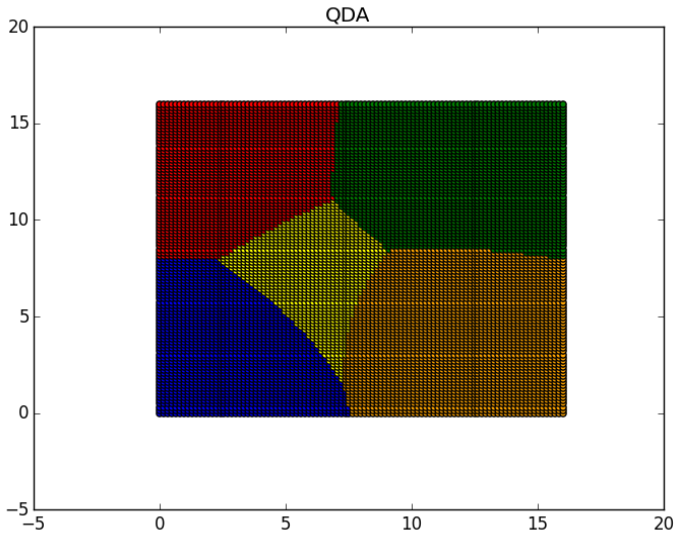
CSE 574 – INTRODUCTION TO MACHINE LEARNING

Susana Dsa, Nikhil Lopes, Avinash Pali (Team 8)
UNIVERSITY AT BUFFALO | CSE 574 INTRODUCTION TO MACHINE LEARNING

1. Accuracy observed for LDA = 97%
   Accuracy observed for QDA = 97%
   Discriminating boundary plots for linear and quadratic discriminators
   a. LDA

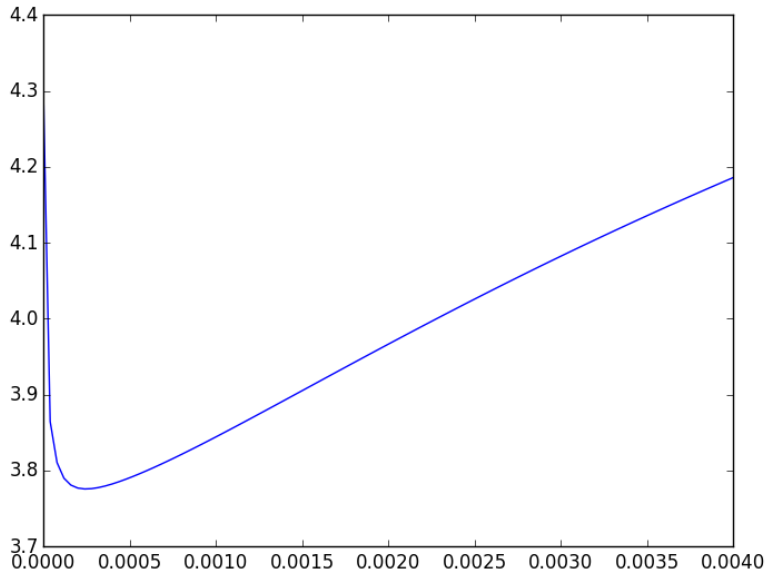


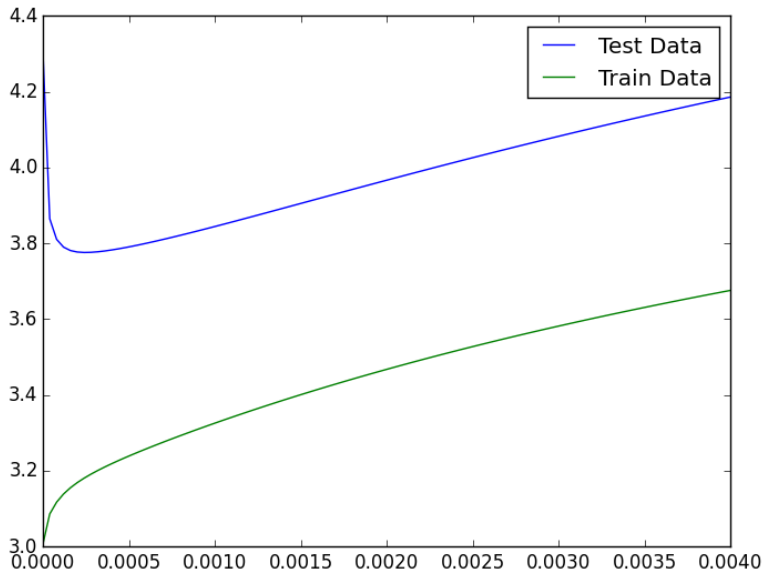
   b. QDA



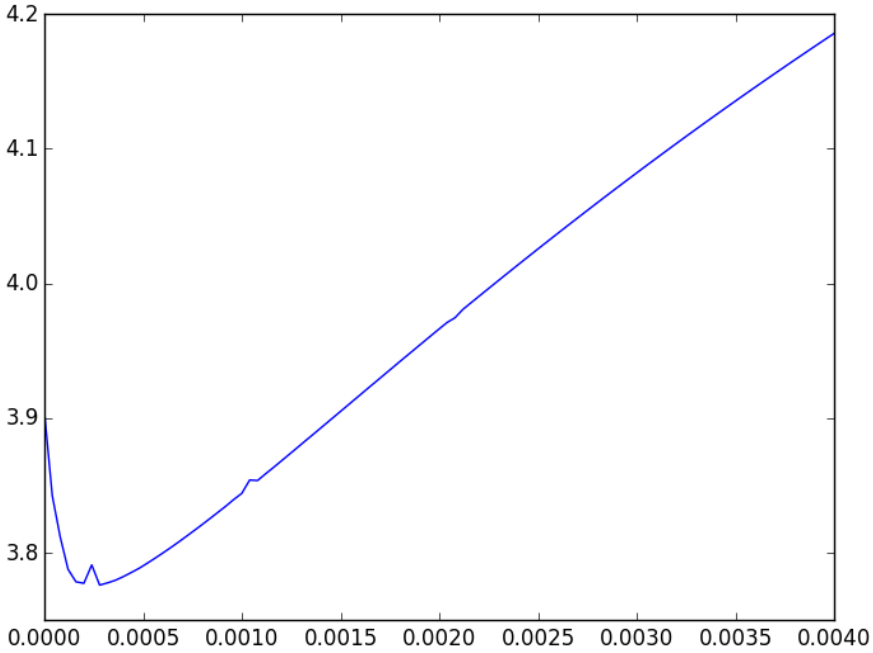
   Conclusion: - The plot boundaries LDA and QDA differ because In LDA, for different k the covariance matrix is identical, and hence classifier becomes linear. The only difference from quadratic discriminant analysis is that we do not assume that the covariance matrix is identical for different classes. For QDA, the decision boundary is determined by a quadratic function and the boundaries appear curved.

2. RMSE without intercept on testing data = 23.1057743385
   RMSE with intercept on testing data = 4.30571723515
   RMSE without intercept on training data = 8.88388057487
   RMSE with intercept on training data = 3.0063021236
   Conclusion: - The method using intercept performs better. Also the method performs much better on training data, maybe because there is less data.
3. The optimal value of lambda is 0.0002, and we can see that RMSE is much lower for train data.

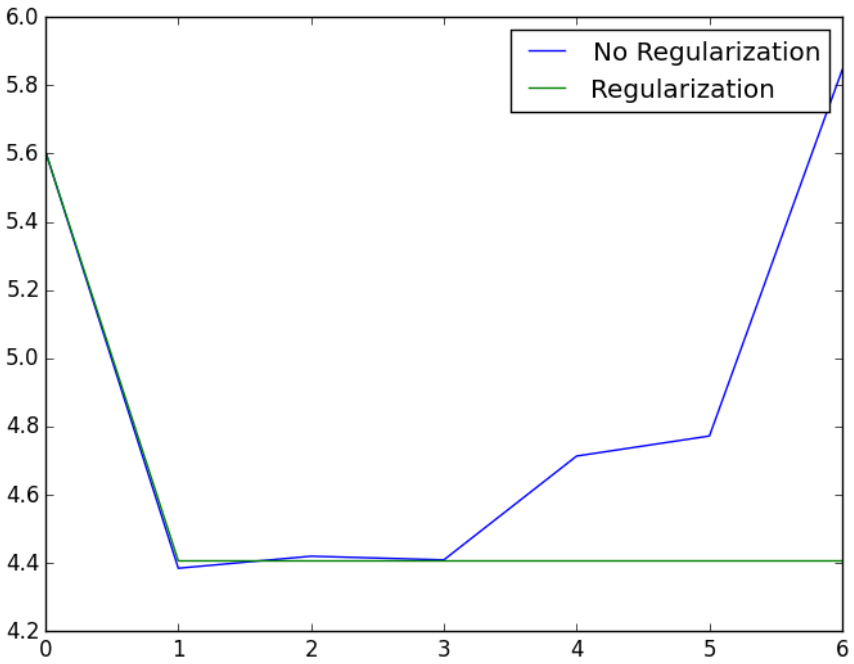




4. Comparing with result of problem 3 one can observe that the results are almost similar but the prior method gives better results as lambda increases.



5. The optimal value observed is p=1, and the regularization method performs better.



6. The RMSE is the square root of the variance of the residuals, indicating the absolute fit of the model to the data – how close the observed data points are to the model's predicted values. Lower values of RMSE indicate better fit. RMSE is a good measure of how accurately the model predicts the response, and is the most important criterion for fit if the main purpose of the model is prediction.

   The best measure of model fit depends on the specific objectives, and more than one are often useful. One may choose the model which gives least RMSE value for particular type of data.