Alex Anderson

EECS 658

Question Answers

1. Based on the accuracy of each model the neural network and linear discriminant analysis are the best. Both models had an accuracy score of 97.34% placing them in the lead above every other model by over a percentage point. The neural network model is scoring so high because it could be over-fitting. This means that the neural network could be memorizing the parameters of how to classify an iris species rather than actively learning and predicting. The linear discriminant analysis method is predicting so well because it is maximizing the separability of each class. This makes telling the iris species apart based on petal/sepal features remarkably accurate.
    a. The naïve Bayesian model predicted an accuracy that is very close to NN/LDA but not quite as good. The naive Bayesian model is better suited to data that is all numerical so that may influence the disparity in accuracy.
    b. Linear regression performed almost as well as the best models due to the data having a linear bias.
    c. The polynomial of degree 2 regression had a similar outcome to the linear regression. However, this model falls short likely due to the nature of the data being spread in a more linear way.
    d. The polynomial of degree 3 regression had considerably less accuracy. This suggests the model is not a good fit and the data is dispersed in a non-cubic way.
    e. The kNN model uses classification that deals with discrete numbers while the regression model is continuous, so I believe that this is the reason for the lower accuracy.
    f. The quadratic discriminant analysis did not perform as well. This is likely due to the model assuming our distribution follows a gaussian distribution.
    g. Svc had an okay accuracy score, but it was less than the best because it needed more training data.
    h. This accuracy score was very close to the best. I believe this is due to the model being supervised. The model prelabeled the data which made it easier to predict the right species.
    i. The random forest model did not perform as well because it does not have an organized flow of data like the neural network.
    j. The extra trees classifier relies on randomness and that is why it did not perform as well on accuracy.
2. a. Yes
    b. 20%
    c. 100
    d. 20
    e. 2
    f. 100
    g. 10
    h. Epoch 1 -100