Homework II

STA 629, Fall 2024

Due: October 4, Friday

1. (70 pts) For this problem, use the training and test sets of the zip code data available from the textbook website (`https://hastie.su.domains/ElemStatLearn/data.html`). You can limit consideration to classifying between the digits "3" and "8". Then compare and contrast the following classifiers in terms of test misclassification error:

   (a) Naive Bayes Classifier

   (b) KNN Classifier

   (c) LDA

   (d) Logistic regression

   (e) Regularized logistic regression. Which type of regularization did you choose? Why?

   (f) Linear SVMs

   (g) Kernel SVMs. Which kernel did you choose? Why?

   If there are tuning parameters associated with the model, you can choose the parameters that give the best training performance.
   Based on the comparison, interpret the results. Which method perform the best? Why?

2. (30 pts) For this problem, use the training and test sets of the zip code available from the textbook website (`https://hastie.su.domains/ElemStatLearn/data.html`). Use all the digits for this problem. Compare the following two methods for multi-class SVMs:

   (a) SVMs: One vs. All.

   (b) SVMs: One vs. One

   Reflection and interpret the results. Which method performs better in terms of test error? Why? Show the confusion matrix for multi-class misclassification. Which classes are most often misclassified by the two methods? Why?

- Bonus question (10 pts): Textbook Problem, ESL, Exercise 4.2 (a)