Homework III

STA 629, Fall 2024

Due: October 25, Friday

1. (50 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) KNN Classifier

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

   (c) Linear SVMs

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

   Please use cross-validation to choose the tuning parameters. Based on the comparison, interpret the results. Which method performs the best? Why? (This task replicates Problem 1 from HW2, emphasizing the use of cross-validation for tuning parameter selection rather than relying on training performance).

2. (50 pts) Using the same dataset employed in Problem 1, write a function to perform 5-fold cross-validation for the tuning parameter associated with regularized logistic regression. Compare and contrast results for 1) minimizing the CV error, and 2) employing the One Standard Error Rule associated with the following loss functions:

   (a) Misclassification error.

   (b) Binomial deviance loss.

   (c) Hinge loss.

   How do the selected tuning parameters perform in terms of test error?

- Bonus question (20 pts): Textbook Problem, ESL, Exercise 7.1.