

STAT 387

Introduction to Statistical Learning, Winter 2023

Alana, Leif, Sang

Instructions:

- Due date: March 21
 - Submit a typed report.
 - Do a good job
-

Preamble :

Consider the wine quality dataset from <https://archive.ics.uci.edu/ml/datasets/Wine+Quality>. We will focus only on the data concerning white wines (and not red wines). Dichotomize the quality variable as good, which takes the value 1 if quality ≥ 7 and the value 0, otherwise. We will take good as response and all the 11 physiochemical characteristics of the wines in the data as predictors.

Problem Statements:

Use 10-fold cross-validation for estimating the test error rates below and compute the estimates using caret package with seed set to 1234 before each computation.

- (a) Fit a KNN with K chosen optimally using test error rate. Report error rate, sensitivity, specificity, and AUC for the optimal KNN based on the training data. Also, report its estimated test error rate.
- (b) Repeat (a) using logistic regression.
- (c) Repeat (a) using LDA.
- (d) Repeat (a) using QDA.
- (e) Compare the results in (a)-(d). Which classifier would you recommend? Justify your answer.