Exercises Statistical Learning - Week 2

Julian D. Karch

1 Applied LDA

Complete Exercise 14 from the book (p.194), but skip parts f) and g). Additionally, answer the following question: Which of the obtained prediction rules (the one from LDA, QDA, or kNN) would you select to make predictions? Note that R users can load the `Auto` dataset from the `ISLR` package, while Python users can access it as a .csv file on Brightspace. Some help:

Solution for a):

```
library(ISLR)
library(dplyr)
summary(Auto)
Auto <- Auto %>% mutate(mpg01 = as.factor(mpg > median(mpg)))
```

You can skip b), just use cylinders, weight, displacement, and horsepower as predictors.

Code for c):

```
set.seed(123)
train_idx <- sample(rep(c(TRUE, FALSE), each = nrow(Auto) / 2),
                    size = nrow(Auto))
train_set <- Auto[train_idx, ]
test_set <- Auto[!train_idx, ]
```

2 Performance Metrics

Compute the confusion matrix (TP, TN, FP, FN), accuracy, sensitivity, specificity, positive predictive value, and negative predictive value for the prediction function you selected in the previous exercises. Are any of those values particularly low?

3 LDA: Computation

Complete Exercise 7 from the book (p. 191).

4 Conceptual: LDA vs. QDA

Complete Exercise 5 from the book (p. 191). Note that question b) is relatively hard.