

DATS 7860: Statistical and Machine Learning for Big Data

Homework 2

Textbook (required): *An Introduction to Statistical Learning with Applications in R*, 2nd Edition (ISBN: 978-1071614174), by Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani (2021), Springer (**Referred as ISL in this course**).

The PDF version of the book, R codes for labs and datasets are publicly available from the book's website: <https://www.statlearning.com/>

A PDF version of ISL was also uploaded to D2L “textbook” folder.

Problems:

1. Problem 2 from Chapter 4 (Page 189).

Optional task: for those of you familiar with matrix algebra, please prove the case of LDA with more than one predictor, i.e. $f_k(x)$ is a multivariate normal distribution $N(\mu_k, \Sigma)$, the discriminant function will be Equation (4.24), which is the multivariate form of the case of a single predictor.

2. Problem 3 from Chapter 4 (Page 189).

Optional task: for those of you familiar with matrix algebra, please prove the case of QDA with more than one predictor, i.e. $f_k(x)$ is a multivariate normal distribution $N(\mu_k, \Sigma_k)$, the discriminant function will be Equation (4.28).

3. Problem 13 from Chapter 4 (Page 192).

Some specific notes for this problem:

- Please submit your complete R codes for this problem.
- The Weekly dataset is available from the R package ISLR2. You can use the following R codes to install the package and load the dataset:

```
install.packages('ISLR2')
```

```
require(ISLR2)
```

```
data(Weekly)
```

- For Part (i), the following are the minimum numbers of models you should experiment with each method: for logistic regression, please try at least three different models; For LDA, QDA and naïve Bayes, please try at least three combinations of predictors for each; for KNN, please try at least three different values of k for the same set of predictor(s).

4. Problem 14 from Chapter 4 (Page 193)

Some specific notes for this problem:

- Please submit your complete R codes for this problem.
- The Auto dataset is available from the R package ISLR2. You can use the following R codes to load the dataset:

```
require(ISLR2)
```

```
data(Auto)
```