

Homework 1 Introduction to Classifiers

I have included the ISLP lab that goes with Chapter 4 in this repository for your reference. This lab makes use of functions that are in the ISLP module, which you would have to install if you want to run all the functions in the lab.

If you decide you would like to have ISLP for running the labs in either your terminal (mac, linux) or Anaconda PowerShell (windows)

```
conda create -- clone base --name islp
conda activate islp
pip install islp
```

When you are done with the lab

```
conda deactivate
```

When doing the homework **please do not use the ISLP module functions** and **use built in functions from scikit-learn**. You may have to do some snooping around to find them.

Chapter 4 Conceptual Problems

1,2,4,6. You can answer these in a pdf. If you are too lazy to type out equations, and can write them neatly, I potentially could be convinced to accept an image. But, if I can't read what you did, I might ask for a typed version.

Chapter 4 Applied Problems

14. Provide a jupyter notebook to solve this problem.

In addition to examining accuracy as stated in the homework also

- examine the confusion matrix
- take a look at other performance metrics (where possible) like precision, recall, roc curve.

Interpretability is also a consideration which suggests you likely want to use a **StandardScaler**

I've included the auto.csv data set provided with the text in this repository.

14 b) provide some exploratory data analysis and explain to justify your choices here.

14 c) investigate and use some of the parameter choices in test training split using **train_test_split** in particular the stratified and shuffled options.

14 e) Do not do QDA. Instead investigate using log transforms to improve model performance.