Nikki Kothari

ML Independent Study

Spring Quarter

Week 1

* Downloaded R
* Spent a couple hours getting familiar with how R works, tried to some basic data cleaning on auto set in ISL
* Read ISL Preface/Ch1

Thoughts: R was way more complicated than I thought it would be. I’m already struggling with the basic things, but ISL was extremely helpful in giving me things to test out and work along with.

Week 2

* Read ISL Ch 2/3
* Figured out how to use python
  + Downloaded anaconda navigator and use jupyter notebooks
* Cleaned auto data set again in python using pandas and numpy to gain familiarity
* Read through a lot of package documentation on pandas, numpy
* Read through many python how-to pages on google to figure out things like functions, methods, calling a method, how to index/slice data, etc.

After talking to my boss at Nielsen this week, it seems like I’m never going to use R. While ISL was making it easy to learn it, I don’t think it’ll be useful. I decided to make the switch over to Python. I once again felt like I was thrown into the deep end as coding really is not my forte. I did a lot of reading this week on basic things because I could not code anything myself. I thought I had a working understanding of these libraries and while I understand how they work in theory, I clearly had not practiced them enough. I managed to clean the auto data pretty well though which is great! Small steps, but felt like a big deal.

Week 3

* Read through Week 1-2 slides for class
* Worked through conceptual questions for Ch 2/3
* Began working on HW1 from week 1 slides
  + Spent a lot of time trying to figure out how to get the data into python
  + How to use packages in practice

Got using python approved and access to the course slides and syllabus. I spent a ton of time re-reading the chapters in super detail and talked through the conceptual questions with a friend who had the answer key in front of them. Helped me ground the concepts I thought I already knew. HW1 feels a bit impossible in python. ISL has all these tips for R but I’m googling

Week 4

* Continued working on HW1 from week 1 slides
  + I spent ~8 hours this week on this trying to figure out how to code and do things in python that were not in R
* Read ch 5+6 in ISL
* Work through conceptual questions in labs for chapters 5 and 6
* Week 3 slides

I gave up once I got to chapter three coding questions. I spent hours on end trying to figure out code for chapter 2, and it was extremely frustrating for things like comment on the range because python does not have an easy way to get the range without doing some stuff I didn’t understand. I felt like I was copying/pasting code from elsewhere to get the answer rather than building intuition for it. Instead, I decided to work through chapters 5 and 6 in the book and skim some of the lecture slides.

Week 5

* Got homework psets from Data Science course, started working through them
* Psets 1-4
  + I spent over 22 hours on this this week!
* Ch 8 in ISL

The data science course psets are really really helpful. They’re built for someone who has very little experience in python and are jupyter notebooks with questions laid out in step by step formats. I’m going to go through these and then try to jump back to the ISL questions.

Week 6

* Ch 9 in ISL
* Week 4 slides
* Psets 4-7 (finished quarter 1)
  + This was also a 15+ hours week ☹

There is no way I have enough time to go back to the ISL questions as I’m trying to cram two quarters worth of data science psets and a quarter worth of ML learning/reading into these last four weeks. But, python is getting way easier and the topics in ISL are lining up really nicely with the psets so that’s good. I also need to cut back on ML things, as I have four other classes right now. I’m going to try to finish up these psets and the reading at the very least and see where that gets me.

Week 7

* Q2 psets 1-2
* Read through a few powerpoint slides

I finally have a handle on things. I can code basics in python, and I started messing around with scikitlearn which is the machine learning package. I’ve also finally started moving into more advanced coding things that like regression that I couldn’t do before understanding how to code to start with (lists, dictionaries, etc). I’m going to try to do some solid machine learning coding next week now that I’m comfortable with how this works and what I can do. These two psets took me a decent amount of time (~10 hrs) which I believe is approximately how much time they should take. I’m glad I’m finally keeping pace.

Week 8

* Q2 psets 3+4+5

I finally got into some machine learning. I used sci kit learn to segment data by clustering, messing with the K values and seeing the effects. I also plotted elbow curves to figure out the ideal K. I finally feel like I’m getting something really ML focused out of this coding. Also, wow! I would never have believed I would get here this quarter. I struggled so much to even be able to create a data frame and now I can execute actual statistical techniques in python. I am a little worried about my ability to code by myself though. These psets have been really helpful as a guided explanation to the concepts and clean data with answers that I am supposed to reach. But, I’m not sure I can get these data sets to the point where I can reproduce these techniques. I want to spend the last few weeks working on my data cleaning skills to make sure that I can get datasets to the point of being able to work with them since its unrealistic to expect this kind of clean data in the future.

PLANNED

Week 8

* Q2 psets 3-4
* Paper readings

Week 9

* Q2 psets 5-6
* Get through all slides

Week 10

* A final project of sorts, picking a random dataset, cleaning it, and performing basic data analysis on it.