David Kinney DSC 550

Professor Deitenbeck Winter 2020

2.4 Reflection: Week 2 Reflection

Conceptually, I enjoyed this week. I have virtually no experience with this sort of text processing; it’s always been numbers and categorical data. I may have mentioned last week that I’ve spent most of my professional career as a coder of one form or another. Most recently, before I started working at the Bank three years ago, I was a full stack web/c# developer. This week was predominately about coding (although there was a fair amount of theory hidden amongst the Discussions).

I’m also finding that I like the way our textbook, “Machine Learning with Python Cookbook” is structured: problem, solution, discussion. There’s a real danger with just providing a solution. If I can paraphrase an old well-worn saw, that’s just giving me a fish, not teaching me how to catch my own. The Discussion sections provide that valuable information.

The exercise was hard! Or maybe I just got lulled into a false sense of security with how easy and short the Week 1 exercise was. At any rate, it took me quite a bit of time to get through it.

One thing I’m struggling with are pandas’s Series and DataFrames. They are fantastic to work with in their intrinsic forms. Where I struggle is when I have to “un-panda” one of the structures. For example, the pandas data structures were indispensable to clean this week’s exercise data—converting to lower case, removing punctuation and stop words. However, when when it came to requiring just string data, converting the pandas structures back to strings to operate on became a real head scratcher.

All that being said, I forced myself to work through it without referring to the cheat sheets. I can’t honestly say it ended up in a light bulb moment, but when the dust settles, I have my Jupyter notebook to refer back to. I’m confident that running through it again a few dozen times, the light bulb will eventually ignite.