David Kinney DSC 550

Professor Deitenbeck Winter 2019

7.4 Reflection: Week 7 Reflection

I really felt like a lot of things started coming together for me this week. Most of that has to do with working through the Titanic project, as well as building out my own case study. It has been incredibly helpful that you’ve been providing us with a coding framework for the Titanic project; it’s a great safety net. However, as I work through my own case study I’m forced to explore and apply other methods to analyze my dataset and then reduce its dimensions via feature extraction and selection.

In the beginning I really struggled with the “Machine Learning with Python Cookbook” as a *text*book. I had a really hard time trying to just read it like… well, like a book. But I learned over the ensuing weeks to read the Problem, Solution and Discussion sections, and just realize that there is code there. Also, I accepted that I was not going to absorb it all. But what it did allow me to do was recall potential solutions. For example, as I was approaching the feature selection portion of my case study, in the back of my mind I *recalled* reading various methods for dimensionality reduction—applying thresholds, eliminating redundant variables that are highly correlated and so forth. In addition, my dataset contains a wide variety of scale in the variables; for instance, household income (in thousands of US dollars) and employment rate (a percentage). I again recalled reading earlier in the semester about *scaling, standardizing and normalization.* So, I was able to go back to those topics, peruse the Cookbook, and find methods that, when applied, brought all the variables onto a level playing field.

Lastly, I’m starting to feel that I am building a solid toolbox and mindset when it comes to looking at a data mining project holistically. And as a Data Scientist, that’s a pretty great feeling…