Sean Cordrey

October 17, 2014

Stat 3654

Roberts

Weekly Paper 8

Chapter 7 shows us some wrangling functionality using pandas. A large section is focused on merging data frames and a few ways it can be done with each method producing a slightly different data frame. First, data frames can be merged using keys as the method of joining, useful for consolidating data that is marked by same keys which relay some type of information. pandas also can merge on index or concatenate along an axis. Concatenating occurs along the two axis refers to as axis 0 and axis 1. All of these share similar parameters that determine which keys to merge on and an inner or outer parameter that decides whether to exclude or include non-shared keys. There are also transformation methods in pandas which remove duplicates and using mappings to add additional information to the data frame specific to each observation. It’s also simple to turn a continuous variable to discrete by making cutoffs and using the pd.cut method. Lastly in this chapter, there is an easy way to make training and testing sets with permutation and random sampling which takes just a few methods to complete.

Chapter 8 deals with visualizing predominately through figures containing graphs using matplotlib. Typical graphs are easily made in matplotlib and there a number of ways of customizing the way the graphs look, including shapes, colors, line styles, legends, and transparency. The figures can be saved straight into separate files themselves with plt.savefig.