Sean Cordrey

October 13, 2014

Stat 3654

Roberts

Weekly Paper 7

Chapter 5 is the introductory chapter to pandas. In pandas there are two important structures that are central to our goals, Series and Data Frames. Series act very much like arrays from NumPy with some added functionality. Series can be used as dicts, preserve the indexes when applying Boolean or math related functions and automatically aligns differently indexed data when having functions applied to them. Data Frames act very much like the ones we’ve seen in R. pandas has a number of ways of cleaning up data in a data frames such as adding and deleting rows/columns, renaming rows/columns, reindexing, selecting specific rows/columns, and merging multiple data frames while aligning similar row/column names.

Chapter 6 concerns itself with datasets, importing and exporting them. There are a couple functions that can be used to import data, pd.read\_csv and pd.read\_table, as well as importing only certain portions of data. Chapter 6 also introduces using JSON objects (JavaScript Object Notation). These are used to transfer data to web browsers and other applications through HTTP. It’s possible to encounter massive datasets so memory efficiency is a real concern. Python has a built in method called pickle which stores data in an efficient format that can be easily loaded back into Python with its load function.