Panth Patel
Machine Learning: Data to Models
Project Proposal
April 23rd, 2019

      I am planning on using the FRED database to obtain data regarding a number of economic indicators (Total Nonfarm Employment, Gross Domestic Income, GDP, Employment Rate, Average Hourly Earnings, Average Expenditure, Effective Federal Funds Rate, Consumer Price Index, and so forth) as well as the database's series on U.S. Recession Probabilities to learn a model that represents how the former indicators are related to Recession Probabilities. There are a few questions that I am interested in with this data. First, I would like to identify the five indicators that are most closely linked to the probability of a recession occurring and how swings in those indicators affect the recession probability. The latter may be answered based on a conditional probability assessment of how the probability changes for different (ex. high/low) values of the indicators. Subsequently, I would like to use the model I learn to determine what the probability of a recession is currently given those five (or all) indicators, and then compare that probability with the Fed's U.S. Recession Probability. There are a number of other questions that I would also be interested in exploring such as fluctuations in seasonally adjusted GDP given immigration rates and the role of lower corporate taxes on increased investment and consumer spending for some cursory analysis of trickledown economics.

      The plan for learning the structure and parameters for this data is to first use the PC algorithm and GES algorithm to learn the structure of the graph. If there is sufficient time, I would like to also create my own graph based on my understanding of the relationships between the various indicators to compare that against the subsequent results from the algorithmic methods. The Causal Discovery Toolbox has already implemented for learning structures from data so I am planning on using that to run the PC and GES algorithms. For parameter learning, I am interested in using Maximum Likelihood Estimate (MLE). The pgmpy package has a number of estimators for parameter learning, including MLE, so I intend to use that to obtain the needed parameters. To perform inference, the pgmpy package also has classes that can be used for various inference methods. For this project, I am planning on using Variable Elimination and Belief Propagation to query the topics of interest that were mentioned before.

      In terms of challenges that I forsee, the main current challenge has been obtaining the data of interest. When creating the graphs on FRED, several of the economic indicators needed are available but then do not show up when downloaded from the website. As a result, I will either have to find other sources on FRED that contain this data or may have to look directly at the sources that FRED uses and see if the data may be obtained from there. Additionally, integrating the two packages (Causal Discovery Toolbox and pgmpy) may be difficult, although both appear to work with the networkx graphs that the first will output. In the case that these packages do not perform as expected, however, I have also found other packages including libpgm, PyOpenPNL, and PyMC that appear to perform various parts of what is needed for this project.