Panth Patel

Data to Models

Python Packages:

Networkx – modeling the graphs

pgmpy – This is the package I will most likely be using to model the graphs and fit/predict. This guide is very helpful in guiding how to import the data and the later learning/inference: <https://conference.scipy.org/proceedings/scipy2015/pdfs/ankur_ankan.pdf>

PyMC: This is PyMC’s documentation and a tutorial on how to use PyMC <https://pymc-devs.github.io/pymc/tutorial.html>

Pystan

Rmm

Hgmm – for text based

Pystruct- structure learning, mostly CRFs

Look at casauldiscoverytoolbox and see which packages it uses.

pgmlearner

Structured learning for creating graph and then do inference on that (pgmpy)

Structure learning: is there an edge or not

Parameter learning: what is the weight of the edge

-sampling

Pc algorithm: vary alpha parameter

GES algorithm can also be used.

Data Options:

* FRED: Economic data from the Federal Reserve Bank of St louis
  + API to download data
  + CSV of data from graph, missing large sections of data
* Gutenberg
  + Nltk
  + Lda vs hidden topic model
* Fmri
  + Clean-ish data
  + May be easier to model data into relationships between
* Rijksmuseum Art Images
  + Not really HMM but MRF/C
  + Cannot download data
  + Fisher vectors of data using: <https://github.com/tmensink/fvkit/blob/master/README.md>
  + <https://gist.github.com/danoneata/9927923>
  + <https://jacobgil.github.io/machinelearning/fisher-vectors-python>
  + Akt: DCT
  + Pre-trained CNN for tensor
  + <http://vision.stanford.edu/projects/objectbank/index.html>
    - This website also has additional information for the image task
* Lda for text corpus
  + Hidden topic markov models
* Additional python packages:
  + Search github for packages

Questions:

1. Choosing Bayesian vs Markov Model
   1. Some directed edges, some undirected/don’t know relationship?
   2. Could I use both as my two models?
2. How to go about creating graphs/CPTs for Bayesian?
   1. Just play around with values?

Come up with questions that you want to learn about the dataset.

FRED:

* What 5 indicators are most strongly related to the probability of recession.
* What is the probability of a recession currently, given those five indicators.

Rijksmuseum:

PC

* Estimates CPDAG from data
* Libpgm
* Causual inference toolbox
* Pcalg

HMM learning (hmmlearning package), other option

Uses EM to learn parameters, vary num states to represent successfully

Options:

* HMM
* MRF
* DBNs
  + Speech recog
  + Robotics