189Z Homework 3: Ethan Greenberg

HMM Stock Model

In the first paper, researchers used a HMM to model the value of the S&P 500. Using a variety of jargony model performance metrics and training algorithms, researchers trained different models with two to six hidden states on 10 years of monthly S&P 500 value data. After evaluation, the researchers proceeded with a model containing four hidden states. They then applied this model to the real world and used it to trade stock. The HMM yielded higher returns that traditional buy and hold methods of trading.

Please don't take this the wrong way, but I really didn't find this source to be helpful in understanding HMM. I think it was way too jargony and didn't explain how HMMs work so much as it talked about specifics of the stock model implementation. Also, as someone who has a very limited/poor math background, I found this article very hard to read.

HMM Gene Finding

This paper was definitely better than the stock model paper. Here, researchers used HMMs to search for genes in eukaryotic genomes. This article did a good job of explaining (in relatively easy to understand terms) how/what tools were used to identify parameters for the HMMs. The process used involved two primary steps 1) identify similar segments 2) identify introns and exons.

The code included at the bottom was super helpful. I didn't go through it all, but it helped to put some of the math in a context that was more understandable to my math-illiterate self.

Source Summary

AWS Marketplace: COVID-19 - World Confirmed Cases, Deaths, and Testing. (n.d.). Retrieved April 23, 2020, from https://aws.amazon.com/marketplace/pp/COVID-19-World-Confirmed-Cases-Deaths-and-Testing/prodview-3b32sjummof5s

This source will be really important to our project. It is a dataset maintained by Our World in Data and includes data about testing worldwide. Our goal is to develop a better prediction for how many people in the US had currently or previously been exposed to Covid19 just before widespread social distancing measures were implemented. This dataset is updated daily and includes information about cases and tests.