DS-GA 3001.001 Special Topics in Data Science: Modeling Time Series
Homework 3

Due date: March 30th, by midnight

Problem 1. (20pt) Consider an HMM with $K = 3$ latents and multivariate gaussian observations, $d = 2$.

- Sample T=10, 50, 100 data points from the model and visualize the corresponding observations for largest T.

- Using the HMM lab code as a starting point implement marginal (alpha-beta) inference in this model. Comment on the strategies you used for testing the code: how did you make sure that the implementation is correct? [Hint: think of some extreme versions of the parameters where the latent state is easy to work out analytically.]

- Run alpha-beta and Viterbi inference on your sampled data and compare the predictions of the two. Find examples where the MAP estimate is different from the peak of the marginals.

- How does this change as you increase the number of latent states, say from 3 to 5 to 10?

Problem 2. (optional; extra credit: 10pt and 1 day deadline delay for last homework)
Implement EM for an HMM model with K states and gaussian observations (full derivations in handout). Use this code to fit the weekly S&P 500 returns data (sp500w.csv) for $K = 2$ vs. $K = 3$ and compare the two results. [Hint: Use Example 6.17 from tsa4 textbook as guideline for plots and interpretation.]