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

Due date: November 6th, by 6pm

Problem 1. (15p)
Consider the HMM with K=3 latent states and discrete observations $\{1, 2, 3\}$, with parameters specified by: initial distribution $\pi = [1,\ 0,\ 0]$, transition matrix $\mathbf{A} = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$, where $A_{ij} = \mathrm{P}(z_{t+1} = j | z_t = i)$ and likelihood $\mathrm{P}(x_t|z_t)$ described by matrix entries B_{xz}: $\mathbf{B} = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 0.5 & 0.5 \end{bmatrix}$.
Write down all possible state sequences consistent with observations a) 1, 2, 3 and b) 1, 3, 1.

Problem 2. (15p)
Construct an HMM that generates the observation sequence $A^{k_1}C^{k_2}A^{k_3}C^{k_4}$ where A^{k_1} denotes k_1 repeats of symbol A and the number of repeats k_i are drawn from the set $\{1, 2, 3\}$ with equal probability.

Problem 3. (20p)
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 (data/sp500w.csv) for K = 2 vs. K = 3 and compare the two results.
Hint: You can reuse some of the inference code you've worked out for the lab. Use Example 6.17 from tsa4.pdf (yellow textbook) as guideline for plots and interpretation.