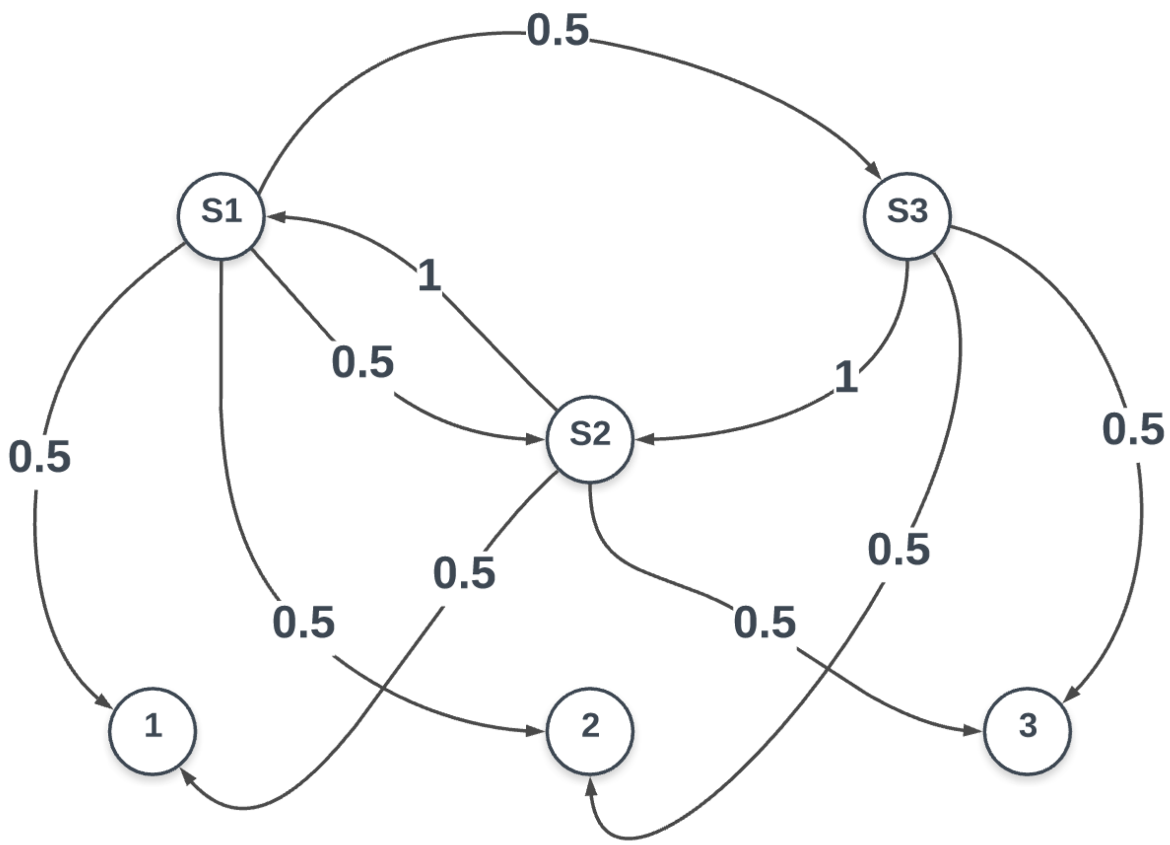
DS-GA 3001.001 Special Topics in Data Science: Probabilistic Time Series Analysis
Homework 3

Due date: Oct 25, by 6pm
YG390

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.
Let the three latent states be $\{S_1, S_2, S_3\}$. Given the HMM with parameters $\{A, B, \pi\}$, the model can be described as:

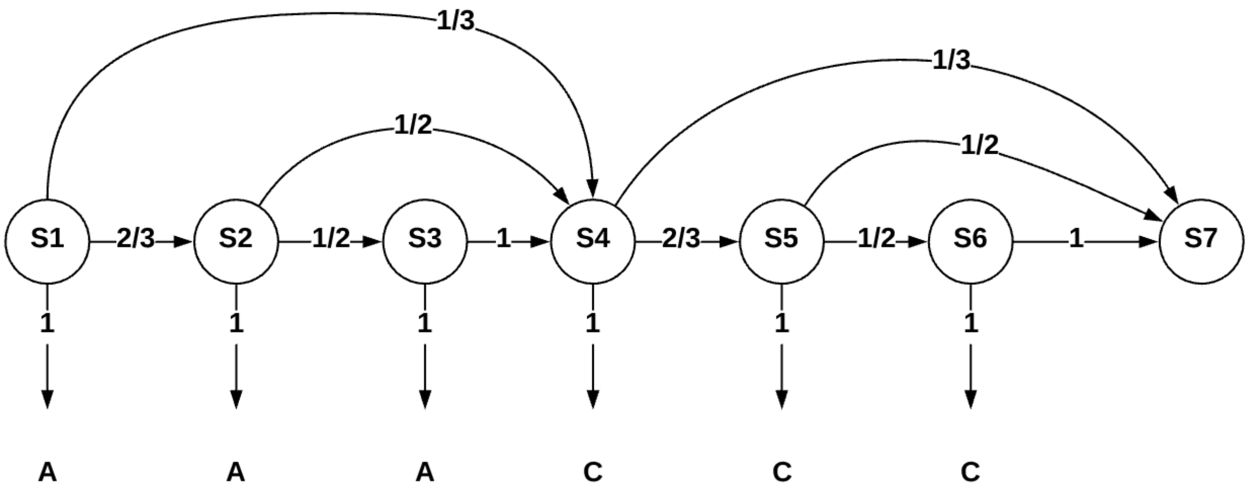


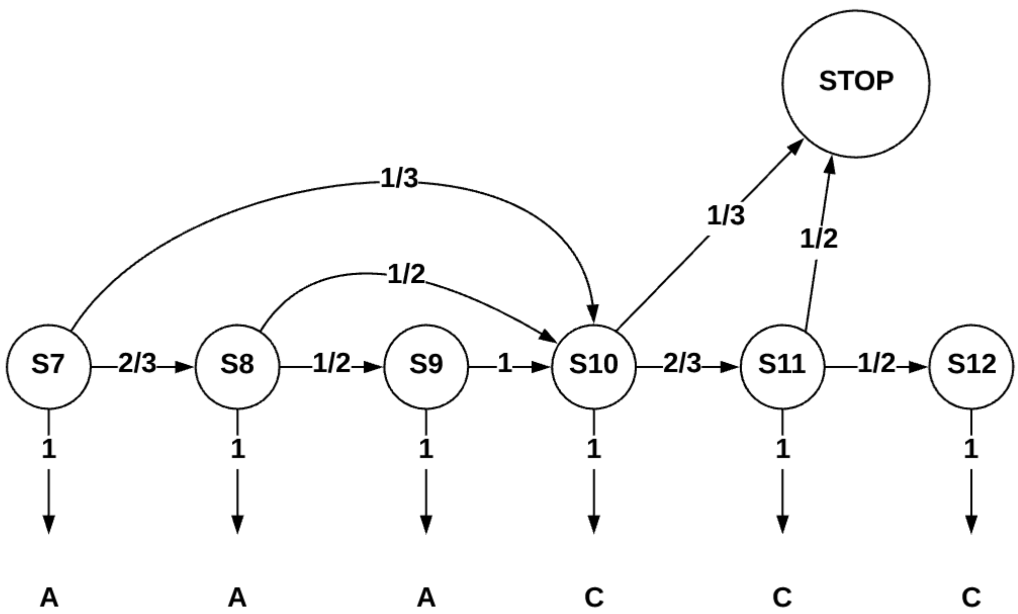
When observing the observations 1,2, and 3 with initial distribution π which implies we start in state S_1 then observing 2 means we have 0.5 probability of being in state S_1 or S_3. However staying in state S_1 is not possible so the only state after seeing the sequence 1,2 is state S_3. Then the last observation 3 means we are in state S_2 or S_3. but from state S_3, the only existing transition is from S_3 to S_2. So the sequence of observations 1,2,3 corresponds to the state sequence $\{S_1, S_3, S_2\}$.

When observing 1,2,1 with initial distribution π, seeing 2 after 1 implies we can be in state S_2 50% of the time or S_3 the rest of the time. The last observation 1 implies that the latent space is either S_1 or S_2. Based on the transition matrix $\boldsymbol{A}$ then the only possible state sequences for the sequence of observations $\{1, 2, 1\}$ are $\{S_1, S_2, S_1\}$ or $\{S_1, S_3, S_2\}$.

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.

The HMM model is defined by three class of parameters $\{\boldsymbol{A}, \boldsymbol{B}, \pi\}$, and is described in the graph below. The latent states are $\{S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8, S_9, S_{10}, S_{11}, S_{12}, STOP\}$ Note that there is a direct connection between the states S_4,S_5, S_6 and S_7 which I broke down into two for convenience of representation. Sequences based on k_i have equal probability: $\frac{1}{3}$, meaning, that for example, for k_1: A, AA, AAA have the same probability $\frac{1}{3}$. The transition matrix has for elements the probabilities indicated on the graph and $\sum_j A_{ij} = 1$. The emission probabilities $\boldsymbol{B}$, are deterministic. All the chains start with A in S_1, the initial distribution is $\pi = [1000000000000]^T$.





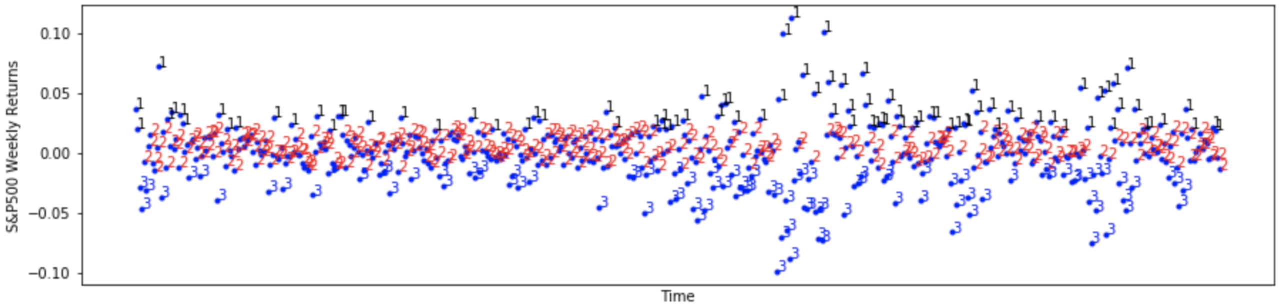
There is a way to avoid the STOP state by linking the last A to either the first C of the last sequence or the last C of the last sequence, and repeating the same for the first C and the second C of the last sequence of Cs. But it was mentioned that a STOP state is an acceptable solution.

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: Use Example 6.17 from tsa4 textbook as guideline for plots and interpretation.

For the parameters initialization:

1. We use a gaussian mixture which gives us estimates for the mean and the covariance of each components namely μ_k and Σ_k from the handout.
2. The initial distribution of the initial state $\mathbf{z}_1$, $\pi = \mathrm{P}(\mathbf{z}_1|\mathbf{x}_1)$ is estimated using an expectation-maximization algorithm with the dataset and fitting the parameters of a gaussian distribution.
3. Based on the KMeans algorithm that we implemented, we obtain a label for every point of the dataset. We then use this labeling to approximate the initial transition matrix, by counting the number of transitions among labels and normalizing the frequencies so each row of the transition matrix sum up to one.

This is the plot of the S&P 500 returns over many years with estimated labels:



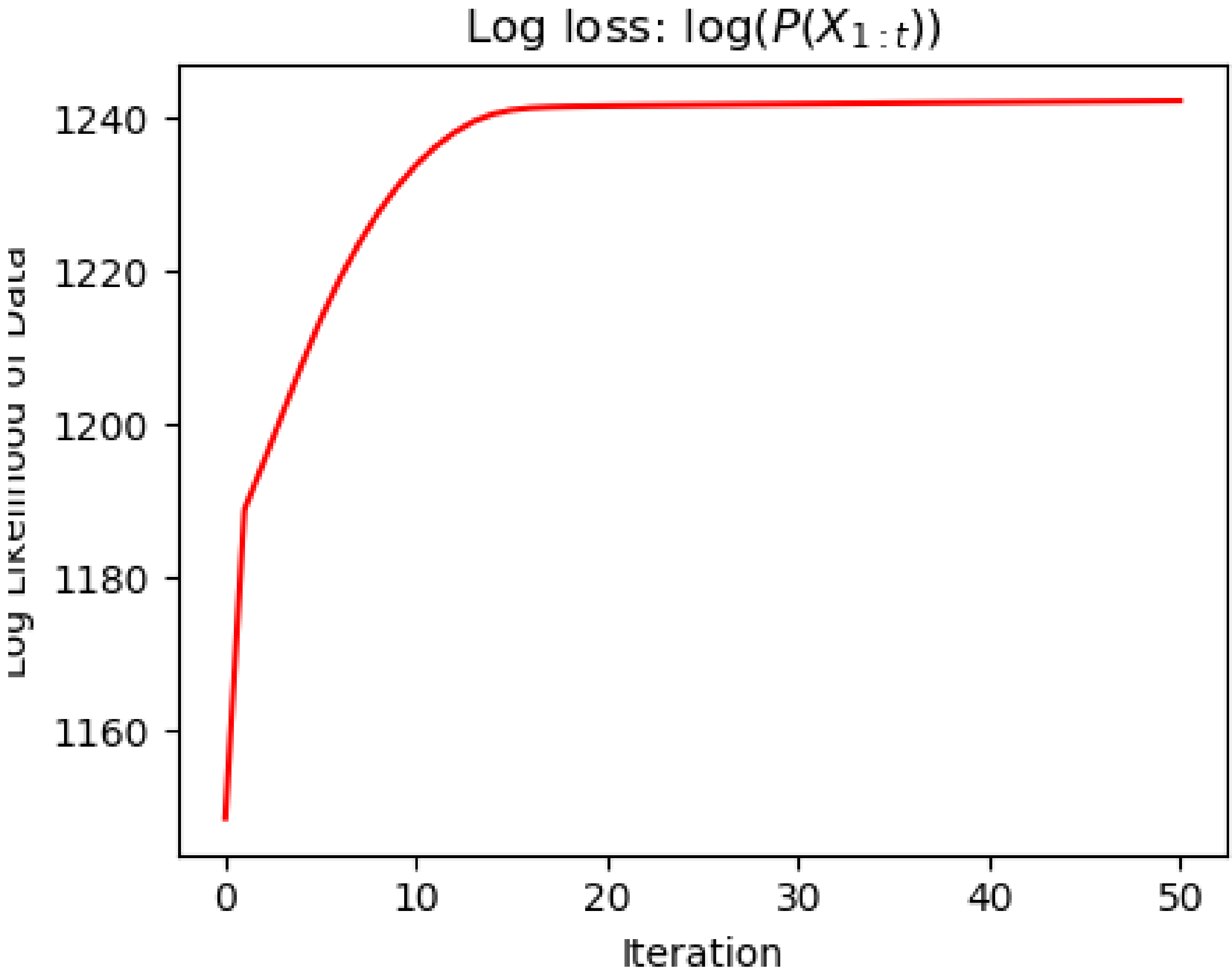
Once the initialization is completed, we alternate between the E steps, for which we run the alpha-beta algorithm with the current estimates, determining:

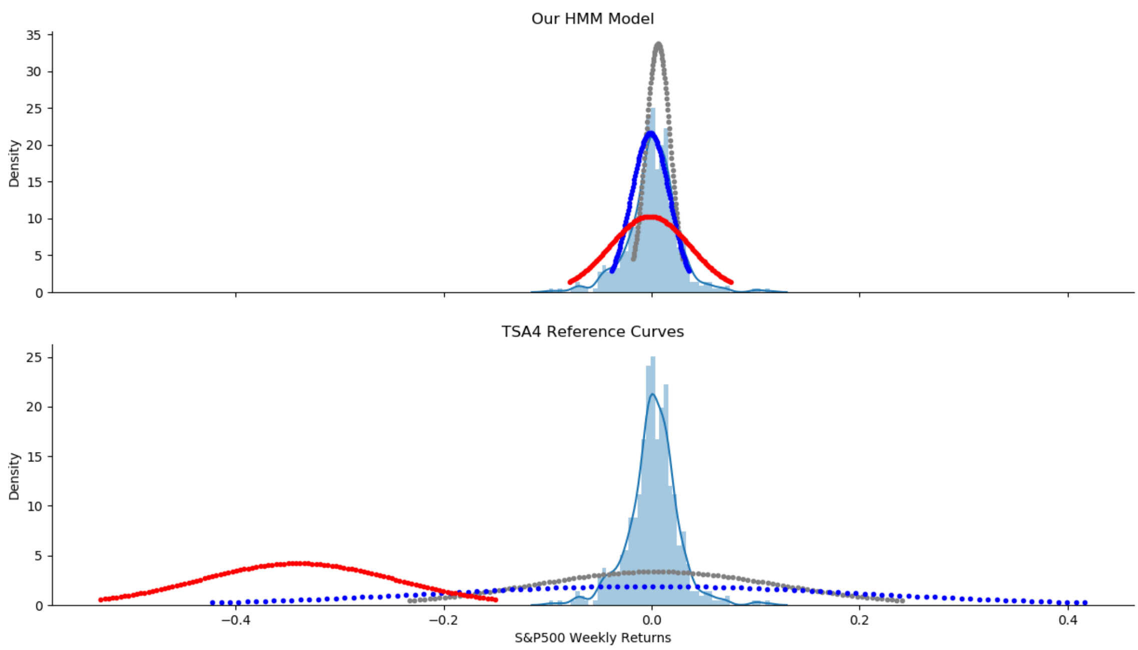
$$
\begin{aligned}
\hat{\alpha}(\mathbf{z}_i) &= \mathrm{P}(\mathbf{z}_i|\mathbf{x}_{1:i}) = \frac{\alpha(\mathbf{z}_i)}{\mathrm{P}(\mathbf{x}_{1:i})} \\
\hat{\beta}(\mathbf{z}_i) &= \frac{\mathrm{P}(\mathbf{x}_{i+1:t}|\mathbf{z}_i)}{\mathrm{P}(\mathbf{x}_{i+1:t}|\mathbf{x}_{1:i})} = \frac{\beta(\mathbf{z}_i)}{\Pi_{j=i+1:t} c_j} \\
\gamma(\mathbf{z}_i) &= \hat{\alpha}(\mathbf{z}_i)\hat{\beta}(\mathbf{z}_i) \\
\xi(\mathbf{z}_{i,j}, \mathbf{z}_{i+1,k}) &= c_{i+1}^{-1}\hat{\alpha}(\mathbf{z}_i)\mathrm{P}(\mathbf{z}_{i+1}|\mathbf{z}_i)\,\mathrm{P}(\mathbf{x}_{i+1}|\mathbf{z}_{i+1})\,\hat{\beta}(\mathbf{z}_{i+1})
\end{aligned}
$$

And the M step, which updates the parameters:

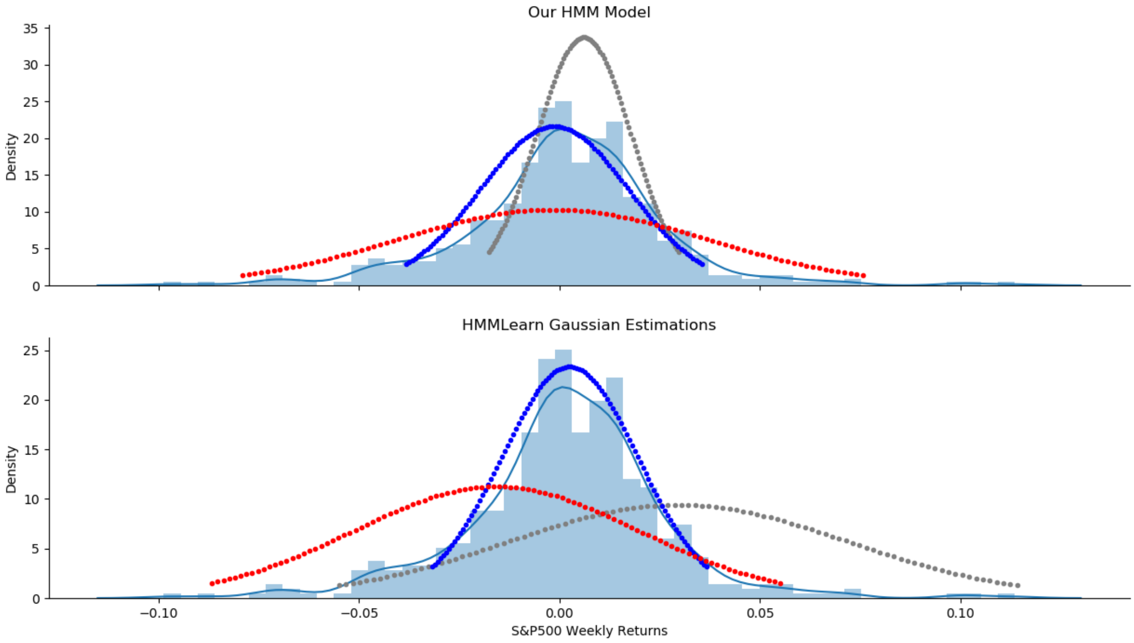
$$
\begin{aligned}
\pi_k &= \frac{\gamma(\mathbf{z}_{1,k})}{\sum_j \gamma(\mathbf{z}_{1,j})} \\
\boldsymbol{A}_{jk} &= \frac{\sum_i \xi(\mathbf{z}_{i,j}, \mathbf{z}_{i+1,k})}{\sum_{i,l} \xi(\mathbf{z}_{i,j}, \mathbf{z}_{i+1,l})} \\
\mu_k &= \frac{1}{\sum_i \gamma(\mathbf{z}_{i,k})} \sum_i \gamma(\mathbf{z}_{i,k})\mathbf{x}_i \\
\Sigma_k &= \frac{1}{\sum_i \gamma(\mathbf{z}_{i,k})} \left(\sum_i \gamma(\mathbf{z}_{i,k})\mathbf{x}_i\mathbf{x}_i^t \right) - \mu_k\mu_k^t
\end{aligned}
$$

The result for $K = 3$, are plotted below with the top plot showing the log loss of the data, the middle plot the inference obtained with our model and at the bottom, the plot using the means and covariances indicated by ts4 textbook. We can see that our model, to some extent, captures better the distribution of the raw data.

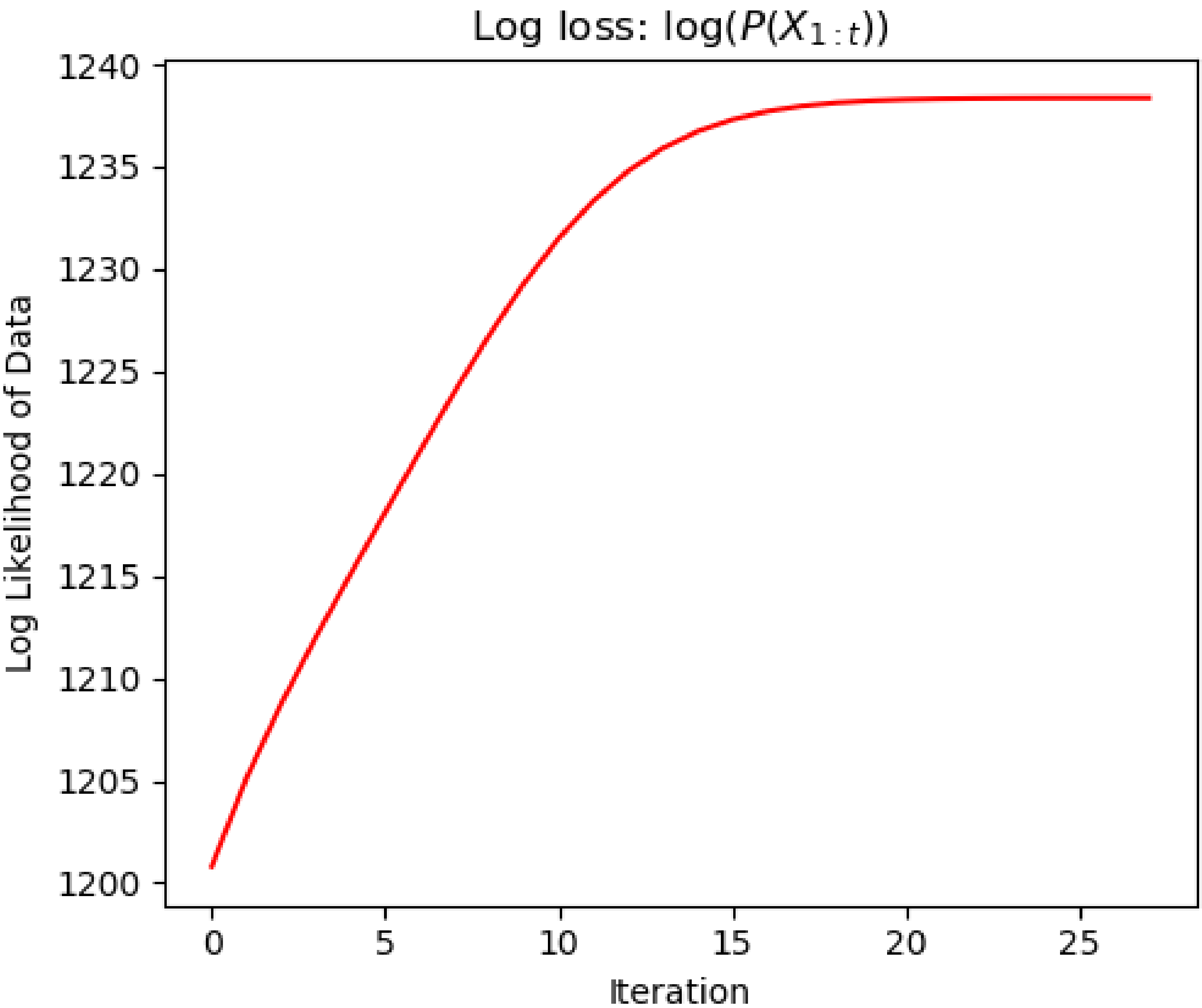


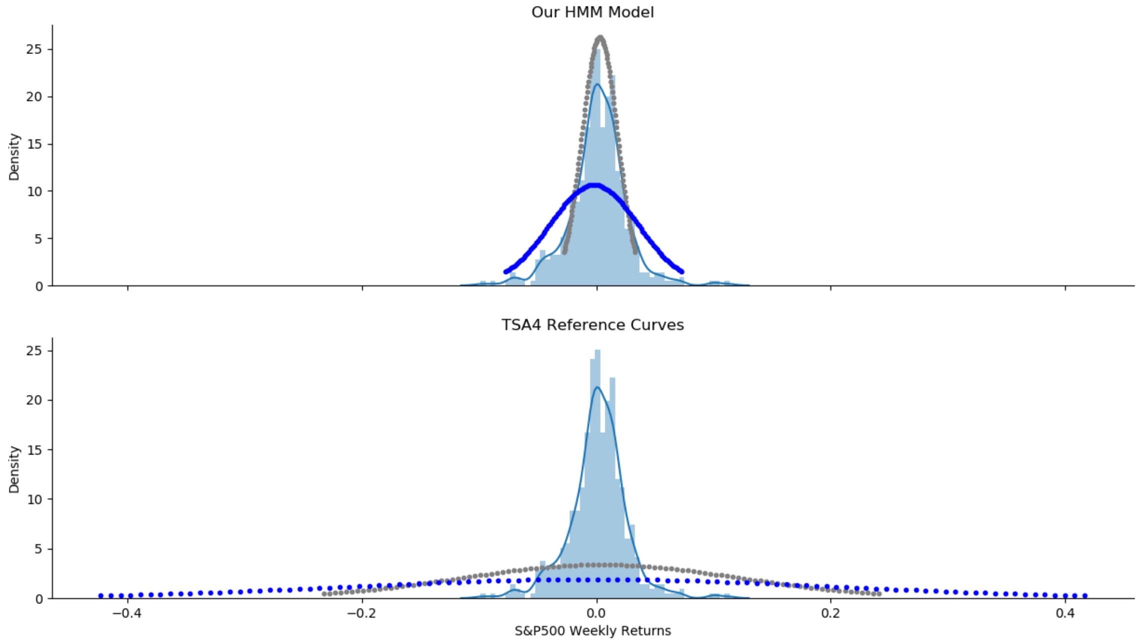


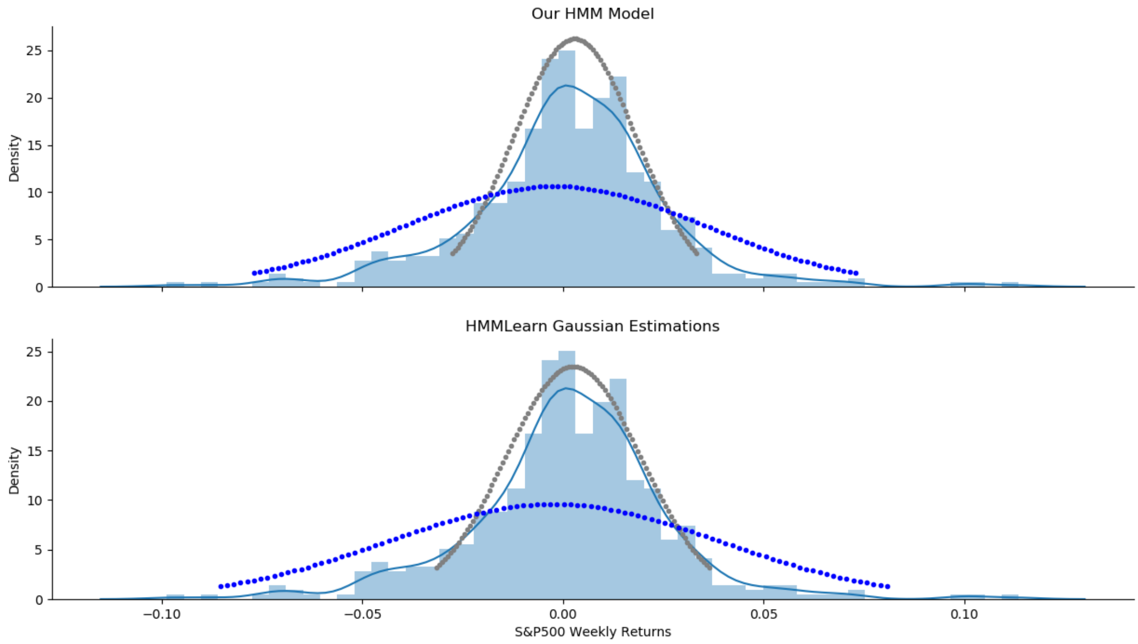
The performance of our model compared to the hmmlearn Gaussian estimator (GaussianHMM), are similar, the GaussianHMM model covering more of the data along the x-axis whereas our model covers the distribution better along the y-axis.



For $K = 2$, the fit is worst but we can see that, unlike our model, the "tsa4" model does not cover the distribution along the y-axis. Compared to the hmmlearn gaussian model (GaussianHMM), our model performance is very similar.




Our HMM Model
TSA4 Reference Curves
Density
S&P500 Weekly Returns


Our HMM Model
HMMLearn Gaussian Estimations
Density
S&P500 Weekly Returns