**Homework # 5**

1. Attached you will find the file `TimeSeries.csv`. The file contains a 1000 by 20 matrix. Each row represents a sample of a random vector $X \in \mathbb{R}^{20}$, but $X$ represents time series data, so that $X_1, X_2, \ldots, X_{20}$ represent measurements at times $1, 2, \ldots, 20$, respectively. Often, we have a collection of time series samples and would like to separate the samples into similar groups, i.e. cluster. Here we'll do this by using a multivariate normal mixture model.

   (a) To visualize the data, produce a line plot of each sample. In R, you can execute

   ```
   plot(m[1,], type="l", ylim=c(-12,12))
   for (i in 2:1000) {
     lines(m[i,])
   }
   ```

   where `m` is the matrix in the csv file. You'll see that the time series are not easy to distinguish. The file `make_timeseries.R` contains the code used to make the data. The data is based on 4 underlying time series found in the file `BaseSeries.csv` which contains a $4 \times 20$ matrix. Look through the files and explain how the data was generated.

   (b) Now assume the following model for $X$

   $$X = \begin{cases} \mathcal{N}(\mu^{(1)}, \Sigma^{(1)}) & \text{with probability } p_1 \\ \mathcal{N}(\mu^{(2)}, \Sigma^{(2)}) & \text{with probability } p_2 \\ \vdots & \\ \mathcal{N}(\mu^{(K)}, \Sigma^{(K)}) & \text{with probability } p_K \end{cases} \tag{1}$$

   Each of the $\mu^{(i)} \in \mathbb{R}^{20}$ and each $\Sigma^{(i)}$ is a $20 \times 20$ covariance matrix. $K$ is the number of mixtures, which we must choose. (In this case, since you know the solution, you can set $K = 4$.)

   (c) To fit this model using EM, you need to know how to derive the MLE of a multivariate normal. Let $Z$ be an $n$-dimensional multivariate normal with mean $\mu$ and covariance matrix $\Sigma$. Let $\hat{Z}^{(i)}$ be iid samples from $Z$ for $i = 1, 2, \ldots, N$.

i. Write down the log-likelihood and use it to show that the MLE estimate $\hat{\mu}$ of $\mu$ is given by the sample mean of the $\hat{Z}^{(i)}$

ii. The MLE estimate $\hat{\Sigma}$ of $\Sigma$ satisfies

$$\hat{\Sigma}_{kj} = \frac{1}{N} \sum_{i=1}^{N} (Z_k^{(i)} - \hat{\mu}_k)(Z_j^{(i)} - \hat{\mu}_j) \tag{2}$$

To convince yourself that this is true you can do one of the following

- Prove it yourself.
- Work your way through the attached proof from Chris Murphy's Machine Learning book and then write up his proof yourself.
- Suppose that $Z$ is 2-dimensional. Let $\mu = (1, 2)$. Let

$$\Sigma = \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix} \tag{3}$$

Produce 100 samples of $Z$ (using your sampler from last week or the sampler in the solutions) and show that the MLE formulas above give you the highest likelihood relative to some other estimates that you compare against.

(d) Take a **hard** EM approach to estimating the parameters of the model. To do this, pick a value for $\theta$ (the vector collecting all the parameters of $X$). Using this $\theta$, assign each sample in the data to the mixture it is most likely to come from. Use the samples assigned to each mixture to update the mixtures parameters using the MLE in (c). The updated parameters give you a new $\theta$. Repeat this process. When you stop your iteration, plot the $\mu^{(i)}$ and determine if you have recovered the underlying time series used to generate the data.

(e) Take a **soft** EM approach. To do this, pick a value for $\theta$ (the vector collecting all the parameters of $X$). Using this $\theta$, determine for each sample in the data, the probability that it comes from each respective mixture. Take a weighted approach of assigning the samples to each mixture and update the mixture parameters to give yourself a new $\theta$. Repeat this process and plot the $\mu^{(i)}$ you derive. (This is the EM algorithm that maximizes likelihood). Compare your result to what you found in (e).