ECE 532
Homework 8

Due Thursday March 10 at the beginning of class

1. Generate 100 samples from a two-component Gaussian mixture distribution with $\alpha_1 = \alpha_2 = 1/2$, component mean vectors both equal to $[0\ 0]^T$ and covariance matrices

$$\Sigma_1 = \begin{bmatrix} 1.25 & 1 \\ 1 & 1 \end{bmatrix} \qquad \Sigma_2 = \begin{bmatrix} 1.25 & -1 \\ -1 & 1 \end{bmatrix}$$

Display the resulting samples in a scatter plot.

2. Consider a multidimensional Gaussian mixture density of the form

$$p(x|\theta) \ = \ \sum_{j=1}^{k} \alpha_j \, \mathcal{N}(\mu_j, \Sigma_j)$$

Derive expressions for the mean and covariance of a random vector $X \sim p(x|\theta)$.

3. In addition to modeling more complicated densities, Gaussian mixture models are very useful for *unsupervised learning* and *clustering*. In unsupervised learning, we seek to find meaningful classes or "clusters" of the *unlabeled* data in the feature space.

a. Implement the EM algorithm in Matlab for fitting Gaussian mixtures to data.

b. Experiment with the algorithm using the Iris dataset. To do this, simply lump the data from all three classes together, and fit a three-component Gaussian mixture model to the data. Hopefully, the EM algorithm will produce three Gaussian components that correspond to the three underlying classes, learning the classes in an unsupervised (i.e., unlabeled) manner.