Donald E. Brown
Bill Basener DS 6014

Exercise 8.5
EM Algorithm

Instructions: You may discuss this assignment with other students in the class, but you must submit your own answers to the questions below. Include an honor pledge with your submission. Submit online and in PDF. This exercise is ungraded.

1. Download and run the Python code EM_choosing_parameters.py, which tests parameters on some random data consisting of two Gaussian distributions. From the resulting figure, explain what parameters (covariance type and number of components) should be chosen for this data.

2. Download the run the Python code EM_unsupervised_image_clustering.py, along with the image Haiti_Image.tif. Use this code and image to answer the remaining questions.

3. Each time you re-run the code, you might get different results. Why is this happening?

4. Try modifying the tolerance using the tol parameter. What is this parameter control and how does the clustering result change with respect to this parameter? [HINT: You can consult the Python documentation https://scikit-learn.org/stable/modules/generated/sklearn.mixture.GaussianMixture.html.]

5. Try changing the n_components parameter along with the tol parameter to find the lowest possible BIC value. Try changing the n components parameter, along with the tol parameter to find the unsupervised classification result that looks best visually. Document your best result by each criteria and include the resulting plots. Do these results match? Why or why not?