

Assignment 2

Due: Monday, March 20, by 8:00PM

Problem: In this assignment, you will set up and solve a clustering problem via a Bayesian model.

The provided dataset consists of independent and identically acquired datapoints $w_n = (x_n, y_n)$ of two scalar components each. The components of each datapoint $w = (x, y)$ follow the statistics

$$x|s \sim \text{Normal}\left(X_s, \frac{1}{\tau}\right)$$
$$y|s \sim \text{Normal}\left(Y_s, \frac{1}{\tau}\right)$$

where X_s, Y_s are cluster specific parameters.

1. Assuming that the measurements are generated by 3 clusters, formulate a Bayesian model for cluster analysis. Make your own choices for the necessary priors and briefly reason on your selection.
2. Represent graphically your model.
3. Describe a Markov chain Monte Carlo scheme to sample from your model's posterior.
4. Implement your Markov chain Monte Carlo scheme and generate samples to characterize the joint posterior distribution of all cluster specific location parameters. Summarize your results graphically.
5. Use your samples to approximate the posterior probability that the datapoints w_{11} and w_{982} belong to the same cluster.

Associated data: The dataset is provided in `cluster_data.mat`.