Coding Assignment 1

Susan Cherry

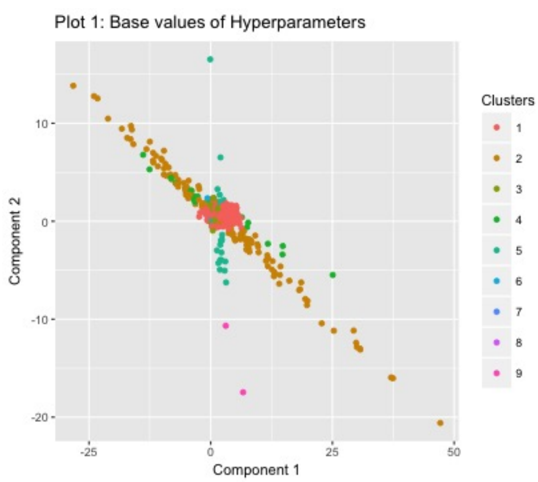
In this assignment, I'm asked to code a generative model of a Dirichlet process mixture of Gaussians using a Normal-Inverse Wishart prior on the Gaussian parameters. I will use the Chinese Restaurant Process with parameter α as a representation of the Dirichlet process.

The generative model is a follows:

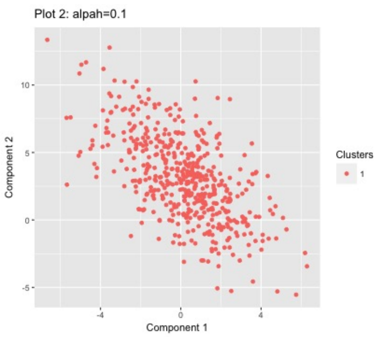
$$z_1, z_2, ...z_n \sim CRP(\alpha)$$
$$(\mu, \Sigma)_{1,2,....\infty} \sim NIW(\mu_0, \kappa_0, \Psi_0, \nu_0)$$
$$x_1, x_2, ..., x_n \sim N((\mu, \Sigma)_{z_i}).$$

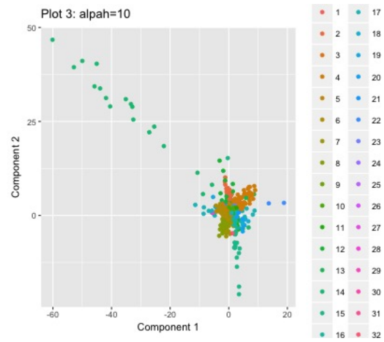
$x_1, x_2, ..., x_n$ are the data points,$z_1, z_2, ...z_n$ are the cluster assignments, and $(\mu, \Sigma)_{1,2,....\infty}$ are the parameters for the clusters. I use this model to generate several datasets, varying both the hyperparameters in the Normal-Inverse Wishart Prior and also the concentration parameter, α, in the Chinese Restaurant Process. I will draw datasets of 500 points. In my datasets, each datapoint has two components but my code can easily be modified to generate data in higher dimensions. The base values for the hyperparameters are as follows: $\alpha = 1, \mu_0 = [1, 1], \kappa_0 = 1, \Psi_0 = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right], \nu_0 = 2$. I will vary the parameters individually to investigate how my datasets change. The hyperparameters that are not specified in each dataset take on the base values.

My results for the base value hyperparameters are below. There are 7 clusters, with Cluster 1 appearing to have by far the most points.

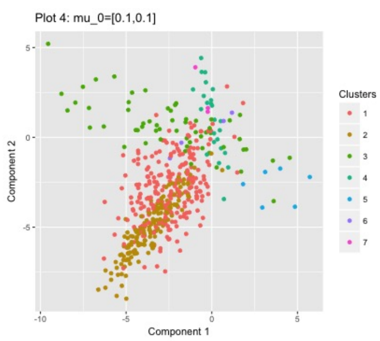


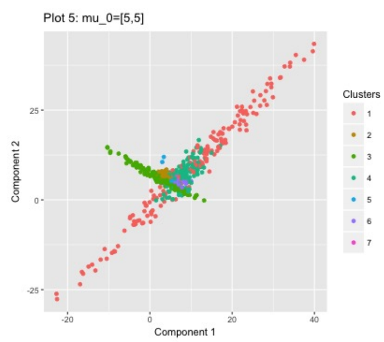
Next, I vary the α parameter. Below to the left is the plot with $\alpha = 0.1$. In this case there is only one cluster. To the right is the plot with $\alpha = 10$. This plot has 42 clusters. These results are not surprising. α is the concentration parameter. As it increases, the probability of beginning a new cluster increases.



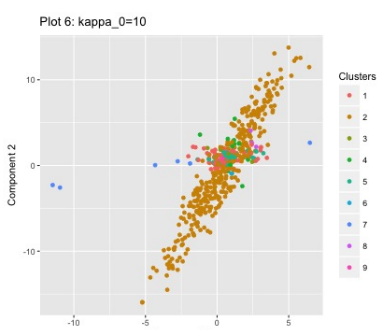


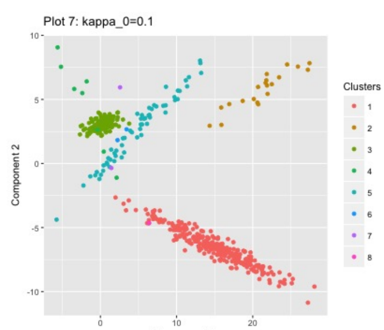
Now I vary the μ_0 parameter. The plot to the left has $\mu_0 = [0.1, 0.1]$. The plot to the right has $\mu_0 = [5, 5]$. Recall that μ_0 is the hyperparameter for the mean of the Normal-Inverse Wishart (from which we draw the individual cluster parameters). While it is difficult to draw conclusions from just these two graphs, note that the mean of the clusters appears to be much lower in Plot 4 (slightly above 0) than in Plot 5 (around 6-7).





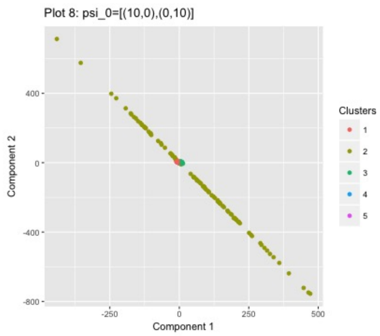
Next I vary the κ_0 parameter. In general, I would expect a lower κ_0 parameter to correspond to clusters with higher variance. Plot 6 shows $\kappa_0 = 10$. We can see that the clusters seem to be fairly spread out. Plot 7 shows $\kappa_0 = 0.1$. The clusters appear to be much more concentrated.

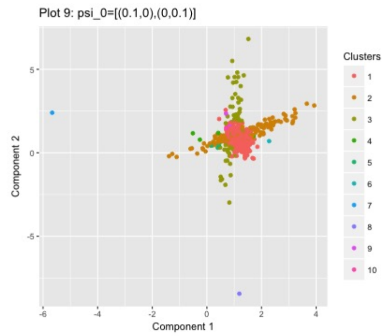




Varying the Ψ_0 parameter gives the results below. Plot 8 shows $\Psi_0 = \left[\begin{smallmatrix} 10 & 0 \\ 0 & 10 \end{smallmatrix}\right]$. Plot 9 shows $\Psi_0 = \left[\begin{smallmatrix} 0.1 & 0 \\ 0 & 0.1 \end{smallmatrix}\right]$. The smaller hyperparameters make the prior more informative:

Plot 9 corresponds to saying that we have a lot of confidence in what the Ψ_0 parameter should be, while Plot 8 allows for more uncertainty through the large Ψ_0 values.





Finally, I vary the ν_0 parameter, which is the degrees of freedom of the Inverse Wishart. Plot 10 has $\nu_0 = 20$. Again, this corresponds to a less informative prior and results in data that is much more spread out (have a high variance). Plot 11, on the other hand, has $\nu_0 = 2$, which corresponds to a much more informative prior and results in clusters that are less spread out (have a lower variance).