Dirichlet Process Mixture of Gaussians

August 25, 2016

Variation of the DP hyperparameter

The α parameter is the concentration parameter for the stick-breaking process. In the stick-breaking process, increasing α means the draws become more spread out. In our case, increasing α means more clusters are generated by the Dirichlet process. Figure 1 shows the results of varying α from 0 to 2. Each plot has the number of clusters listed in the title. As α increases, so does the number of clusters.

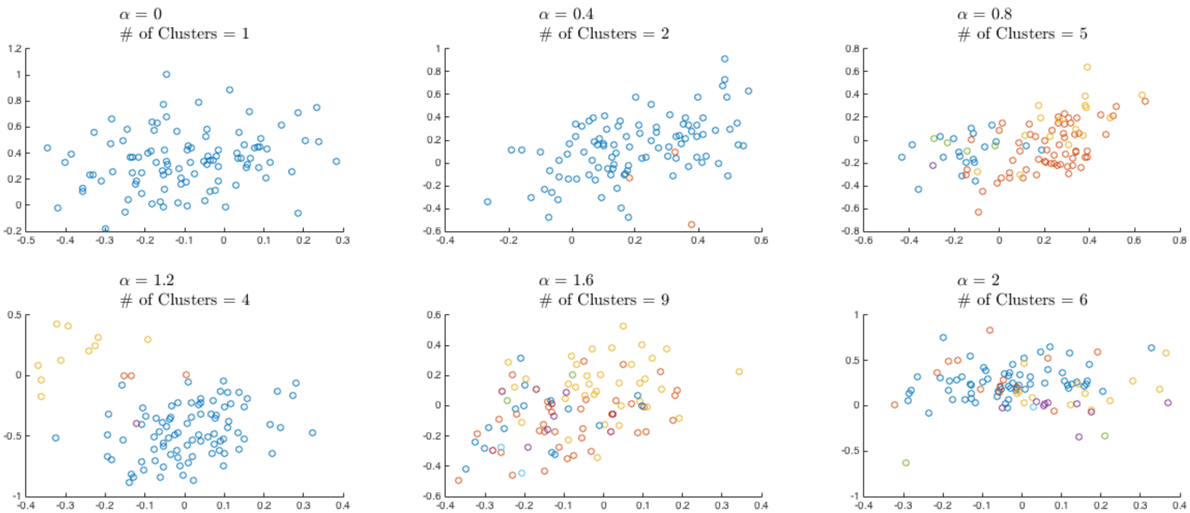

Figure 1: Results of varying α from 0 to 2.

Variation of the Normal-Inverse Wishart Parameters

The Normal-Inverse Wishart distribution is parameterized by four variables:

$$f(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\boldsymbol{\mu}_0, \lambda, \boldsymbol{\Psi}, \nu) = \mathcal{N}(\boldsymbol{\mu}|\boldsymbol{\mu}_0, \tfrac{1}{\lambda}\boldsymbol{\Sigma})\mathcal{W}^{-1}(\boldsymbol{\Sigma}|\boldsymbol{\Psi}, \nu)$$

where $\mathcal{W}^{-1}$ is an inverse Wishart and $\mathcal{N}$ is a multivariate normal. Here we show how each parameter changes the generated data from the Dirichlet process.

$\boldsymbol{\mu}_0$:

Changing $\boldsymbol{\mu}_0$ changes where the generated points are centered. This can be seen from the pdf of the distribution, as $\boldsymbol{\mu}_0$ is a parameter of the multivariate normal. To illustrate this, I generated 6 different plots with varying $\boldsymbol{\mu}_0$ and with all other variables fixed. The center of the clusters correspond to the value of $\boldsymbol{\mu}_0$.

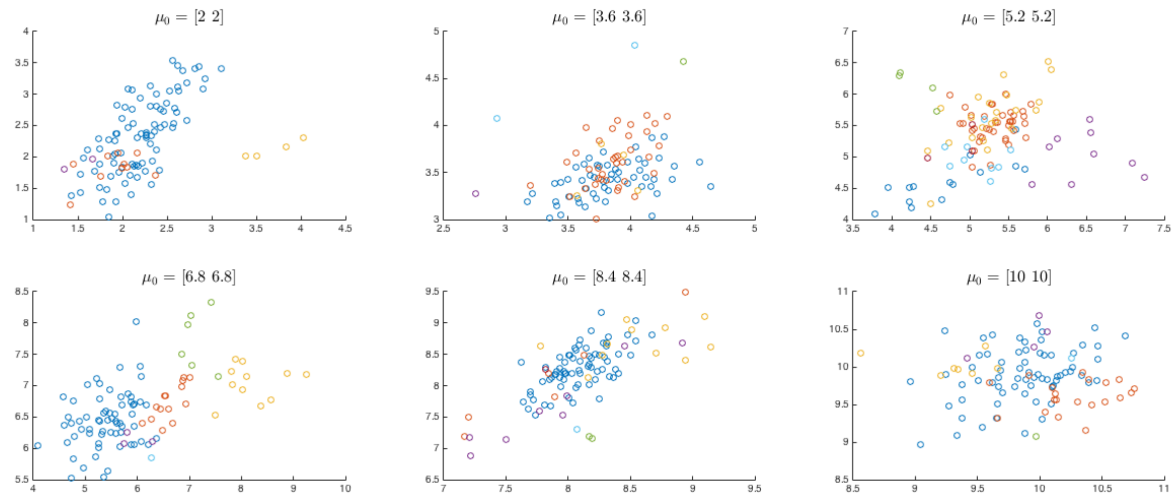

Figure 2: Results of varying $\boldsymbol{\mu}_0$ from [2,2] to [10,10].

$\boldsymbol{\lambda}$:

Changing λ changes how separable the generated clusters are. Larger values of λ mean the clusters are grouped together tightly. This makes sense, since λ is a scaling parameter for the covariance matrix of the multivariate normal.

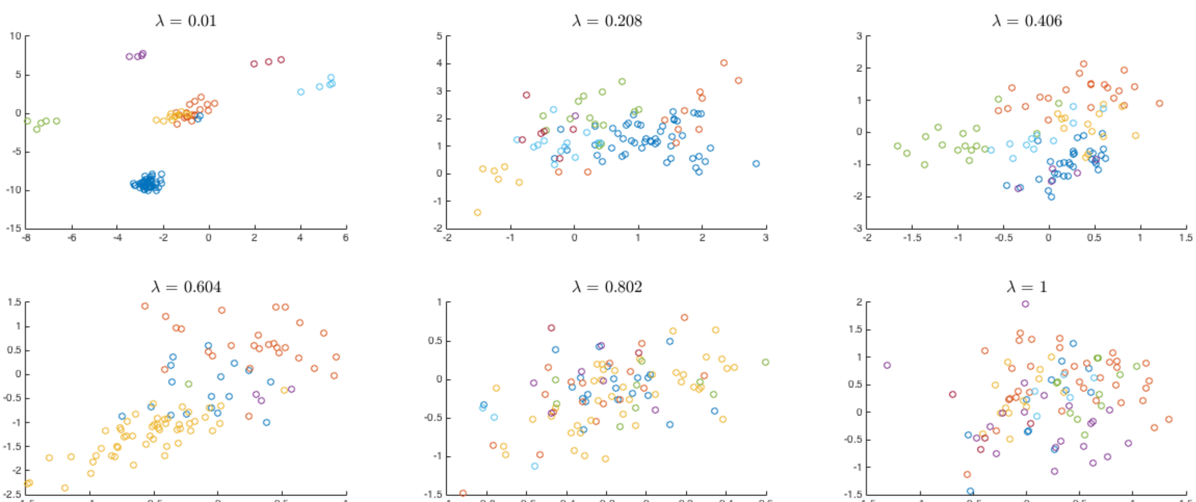

Figure 3: Results of varying λ from 0.01 to 1.

$\mathbf{\Psi}$:

Changing the $\mathbf{\Psi}$ parameter changes how the data is dispersed. Because $\mathbf{\Psi}$ must be a positive semidefinite matrix, it's a little harder to sweep over all possible entries for $\mathbf{\Psi}$. To remedy this, I parameterize $\mathbf{\Psi}$ as a mixture of an identity matrix and a matrix of all ones:

$$\mathbf{\Psi} = \gamma\mathbf{I} + (1-\gamma)\mathbf{J}$$

where $\mathbf{I}$ is an appropriately sized identity matrix and $\mathbf{J}$ is an appropriately sized matrix of all ones. With this formulation, I can sweep over two extremes of positive semidefinite matrices by sweeping over a scalar parameter γ.

The results in Figure 4 show that changing $\mathbf{\Psi}$ changes whether the data lies on a diagonal, or if the data points are distributed more roundly.

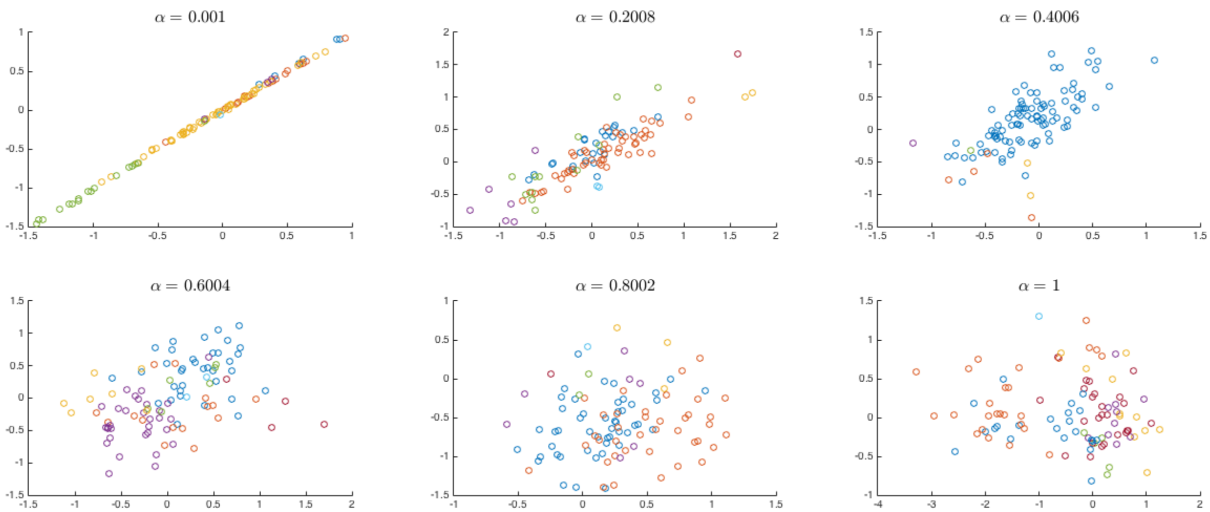

Figure 4: Results of varying γ from 0.001 to 1.

ν :

Changing ν changes the domain and range of the data. This term is the "degree of freedom" parameter for the inverse Wishart distribution. As ν increases, the range and magnitude of the data points decreases.

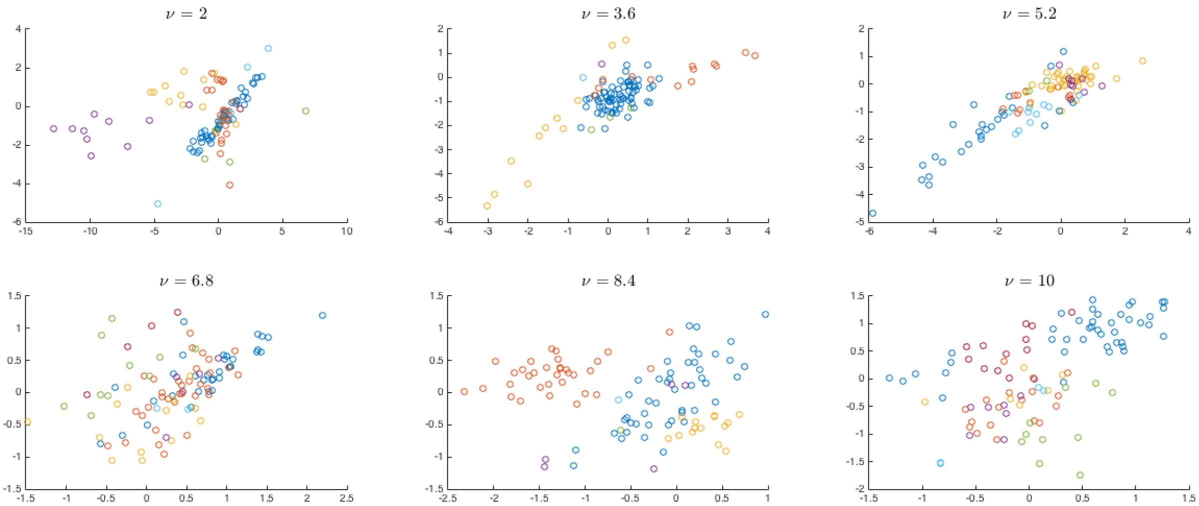

Figure 5: Results of varying ν from 2 to 10.