Indian Buffet Process

August 24, 2016

Variation of the α hyperparameter

The number of clusters the first data point belongs to is drawn directly from a Poisson distribution parameterized by α: Poisson(α). The number of clusters added to the nth data point is also drawn directly from a Poisson distribution parameterized by α: Poisson($\frac{\alpha}{n}$). In this way, α directly controls the total number of clusters and how readily clusters are added to the generated data. Since the Poisson distribution is parameterized by it's expected value, increasing the value of α increases the number of clusters added to each data point. Our experiments show that the total number of clusters ranges from 4 clusters at $\alpha = 1$ to roughly 60 clusters at $\alpha = 11$.
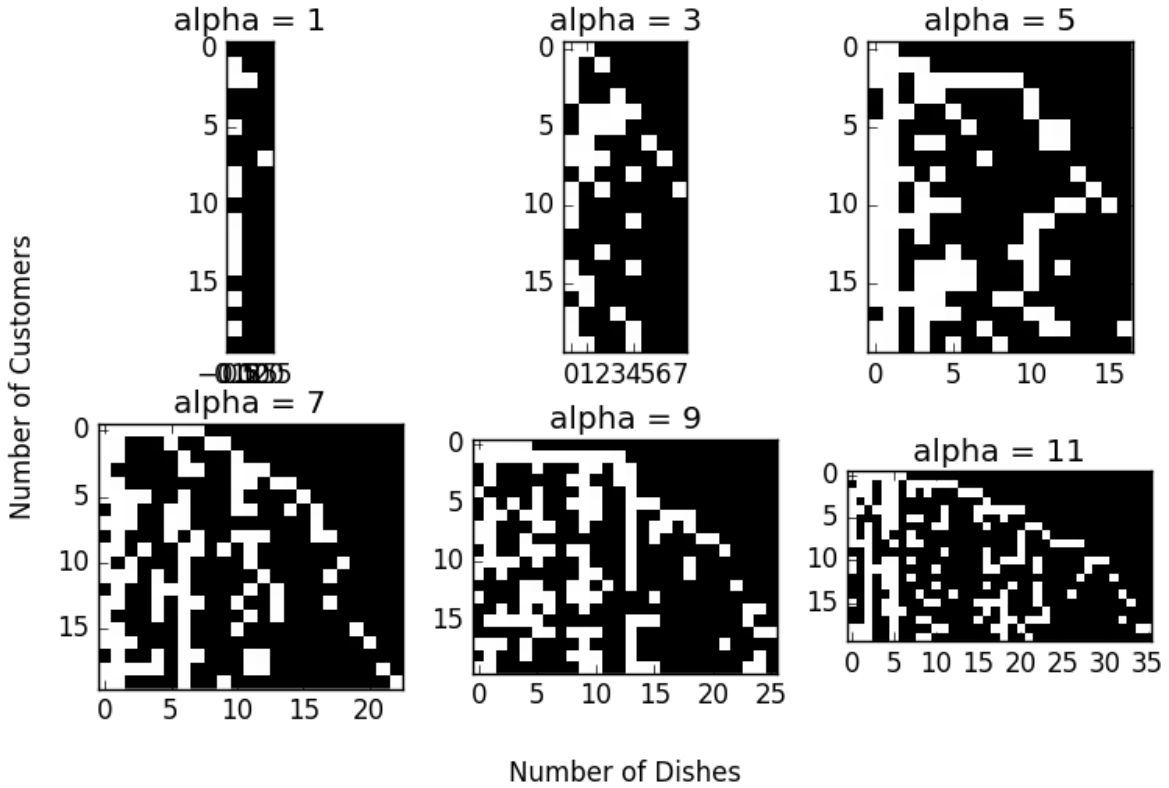

Figure 1: Results of varying α from 1 to 11.

Variation of the Number of Data Points

Varying the number of data points in the IBP slightly increases the number of clusters. This is because every extra data point (customer) gives the matrix another chance to add more clusters (dishes) from the distribution Poisson($\frac{\alpha}{n}$). However, the increase in the number of clusters by adding data points is very slow compared to the increase by varying α. This is because the expected value of new clusters added for the nth customer approaches zero as n gets large. We see that at $N = 10$ we have roughly 40 clusters, and at $N = 110$ we have increased to roughly 60 clusters.
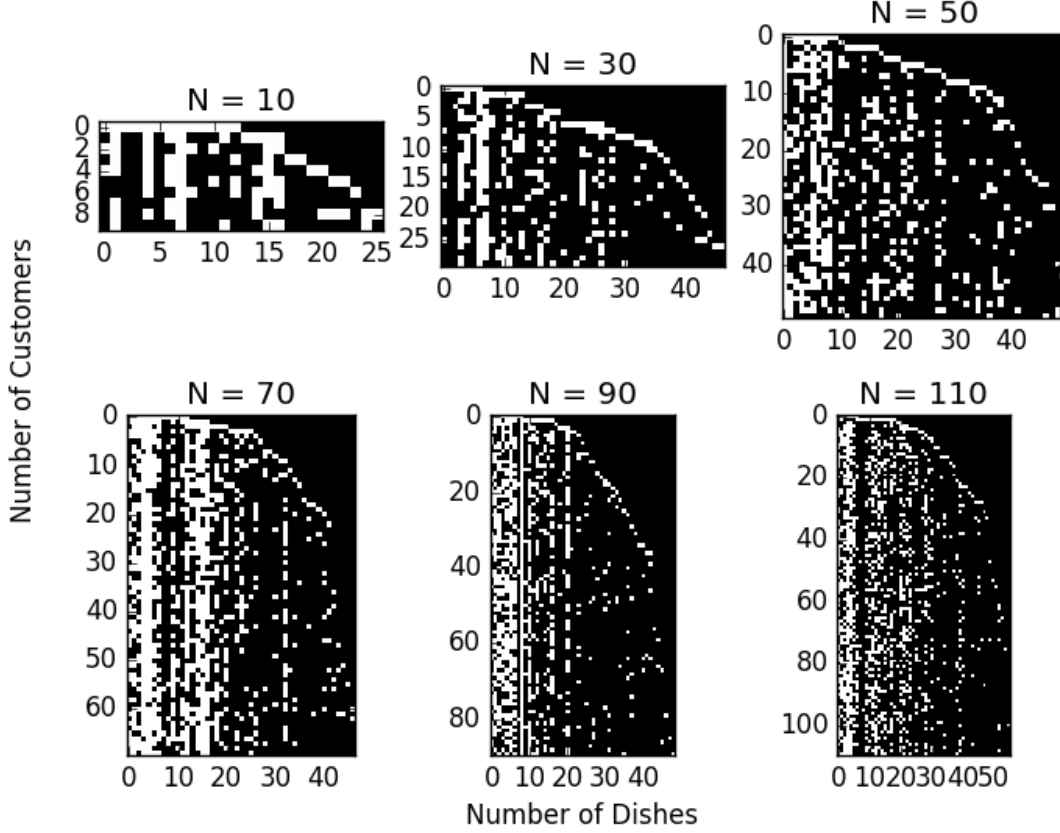

Figure 2: Results of varying the number of data points from 10 to 110.