CS274A: Probabilistic Learning

Homework 6: EM and Guassian Mixtures

Aydin Mohseni

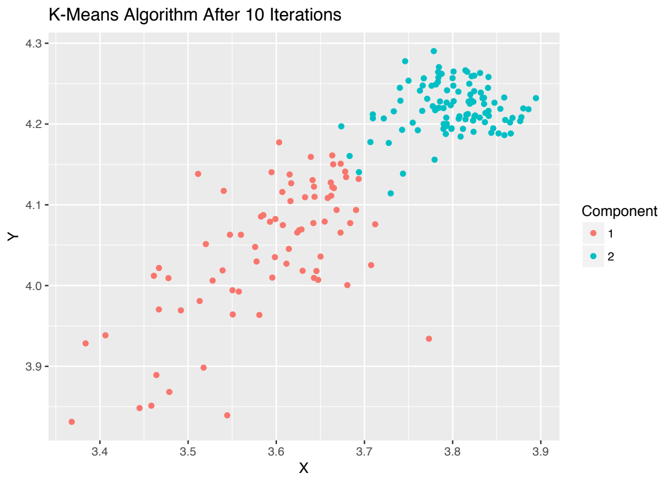
17 March 2017

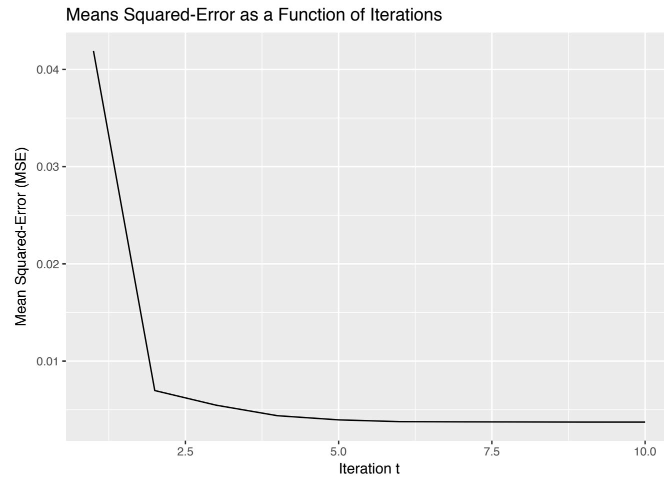
K-Means Algorithm

Scatterplots of Final Classifications

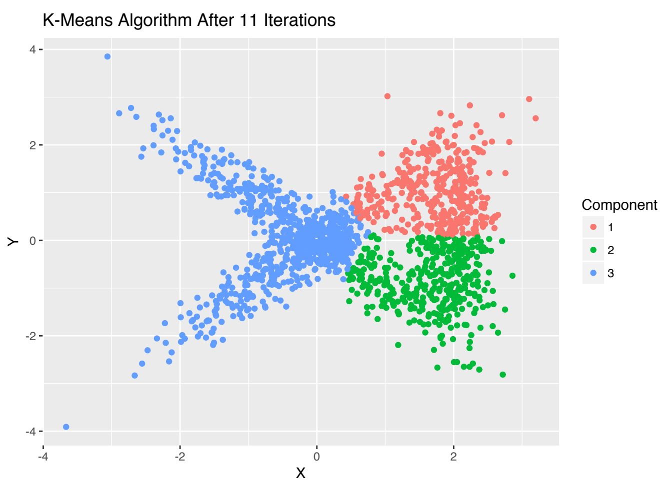
Plot of Mean Squared-Error Over Time

Data Set 3

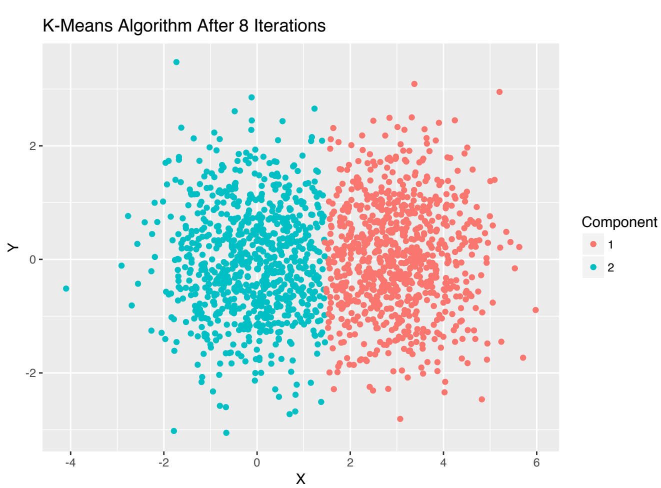


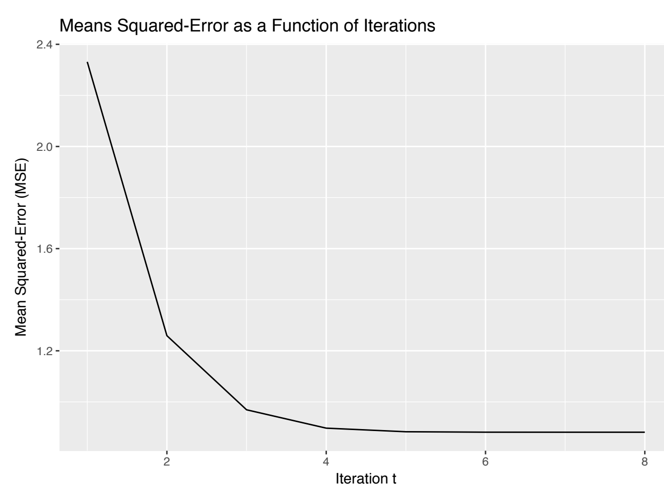


Data Set 2



Data Set 1

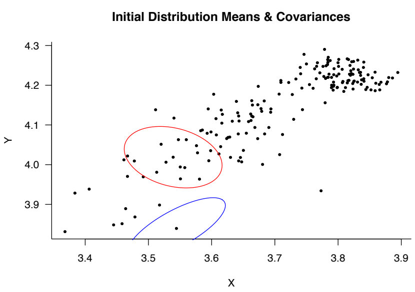


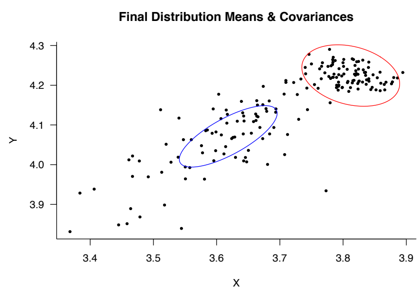


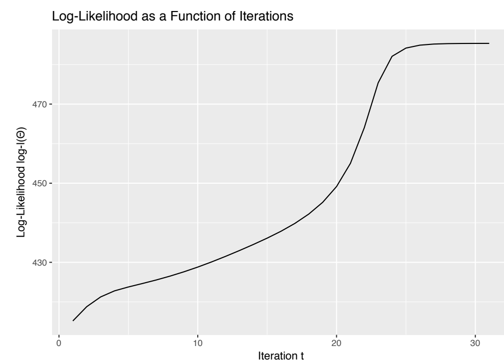
E-M Algorithm

Initial Dist'n Parameters | Final Dist'n Parameters | Plot of Log-Likelihood Over Time

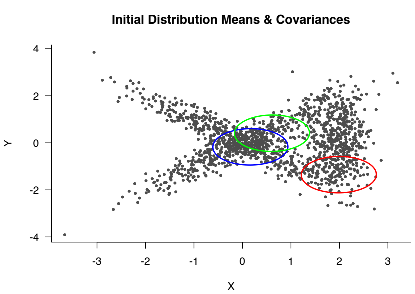
Data Set 3

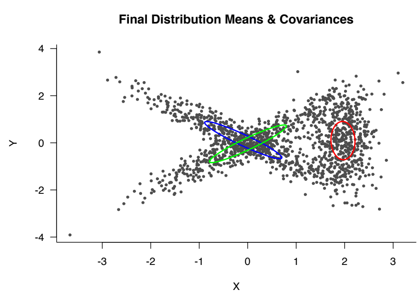


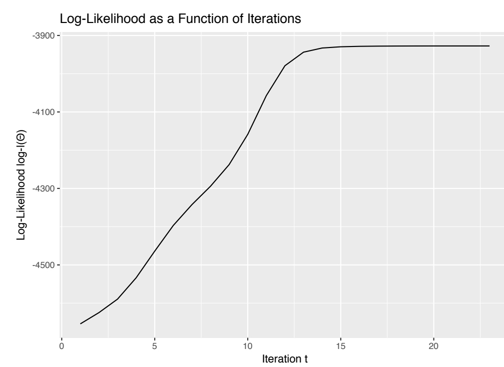




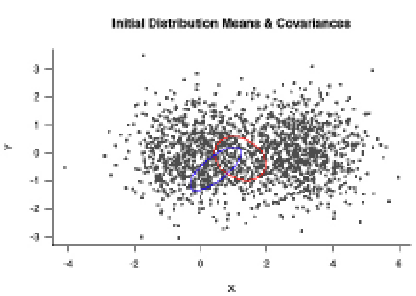
Data Set 2

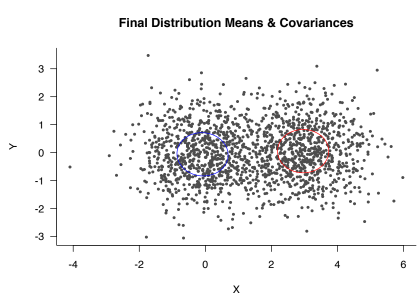


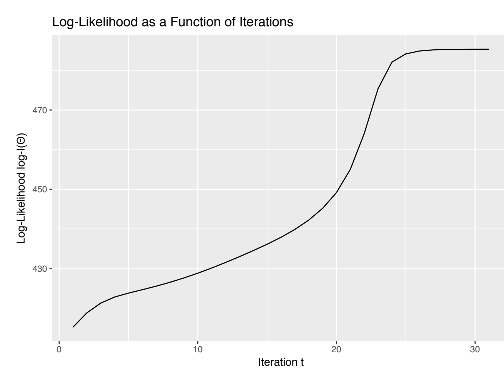




Data Set 1







Comments on the Differences Between K-Means & E-M

for Data Set 3

Both K-means and E-M algorithm solutions ‘break up’ the epidemiological data in roughly the same place. Both identify two clusters corresponding to patients with high red blood-cell volume and mean hemoglobin concentrations (in the upper right) that are unlikely to have iron deficient anemia. How they differ is in the treatment of border-line cases. K-means must make a sharp partition, whereas E-M can encode graded uncertainty about correct classification in the two groups.

for Data Set 2

Here, K-means displays it limitations. Limited by the measure of Euclidian distance as its loss rule, it fails to detect the three underlying Gaussian distribution. The E-M solution, on the other hand, accounts for covariances, and so consistently converges on roughly the true means and covariances of the thrree component (Gaussian) distributions. As a tradeoff, the E-M algorithm (in my implementation) took more iterations to converge to a solution.

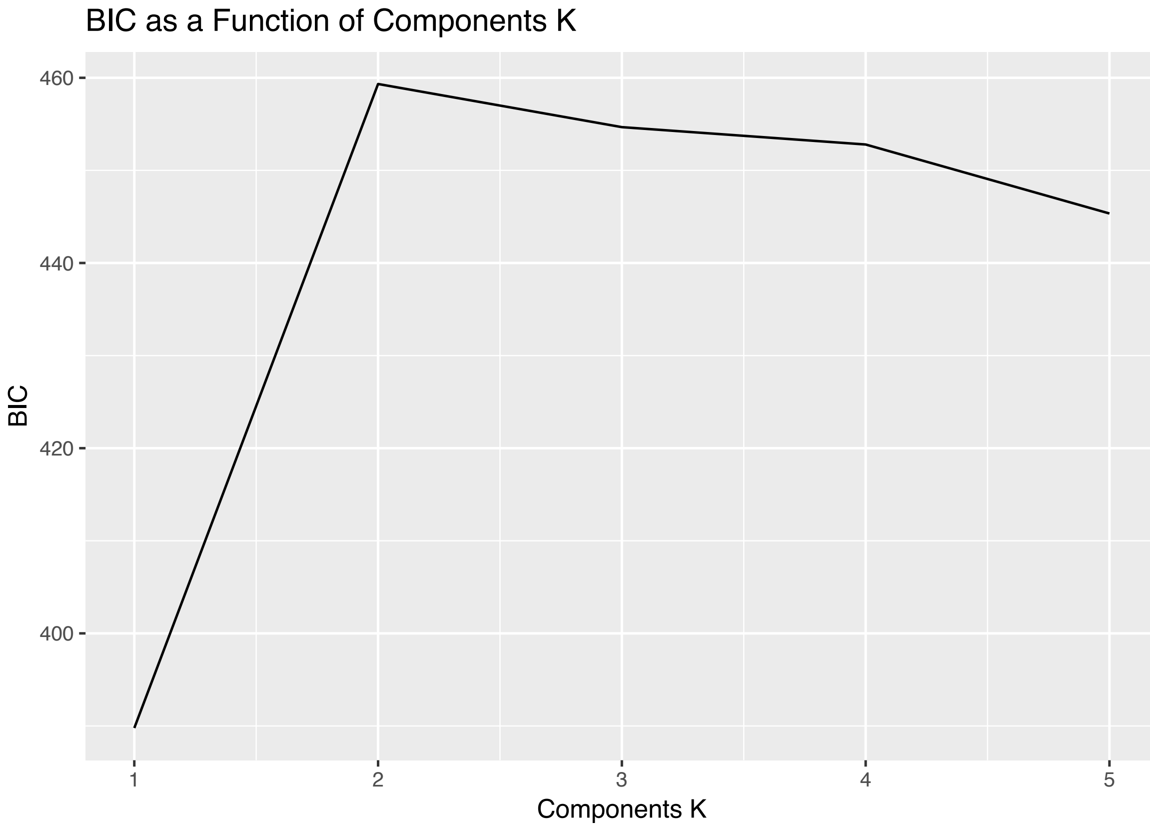
for Data Set 1

Here, both solutions are modest ones. The signifiant overlap of the underlying component (Gaussian) distributions makes it so that the membership of many points in the penumbra of the two distributions is irreducibly uncertain. The K-means solution divides the data space nearly in half, while the E-M solution posits (roughly) accurate Gaussians at the two epicenters of the data.

Automatically Selecting K for Gaussian Mixture Clustering

BIC as a Function of Number of Components K

Data Set 3



Comments

The BIC criterion for data set 3 nicely picks out 2 (Gaussian) components as the optimal number. The BIC makes it so that, as K increases, dimishings gains from increased likelihood are overwhelmed by the penalty for the increased number of parameters. Naturally, this has the potential to deter selection of overly complex models, and so avoid overfitting the data.