Yuliya Astapova, Gabriel Baierle, Gustavo Poscidonio

Programming Homework 5: Unsupervised Learning, Clustering, Density Estimation, Mixture of Independent Gaussians, K-means; Application: Clustering for Robot Localization

Clustering by Gaussian Naive-Bayes Models

Estimates for π_c

Cluster 1	Cluster 2	Cluster 3
0.5669	0.3384	0.0947

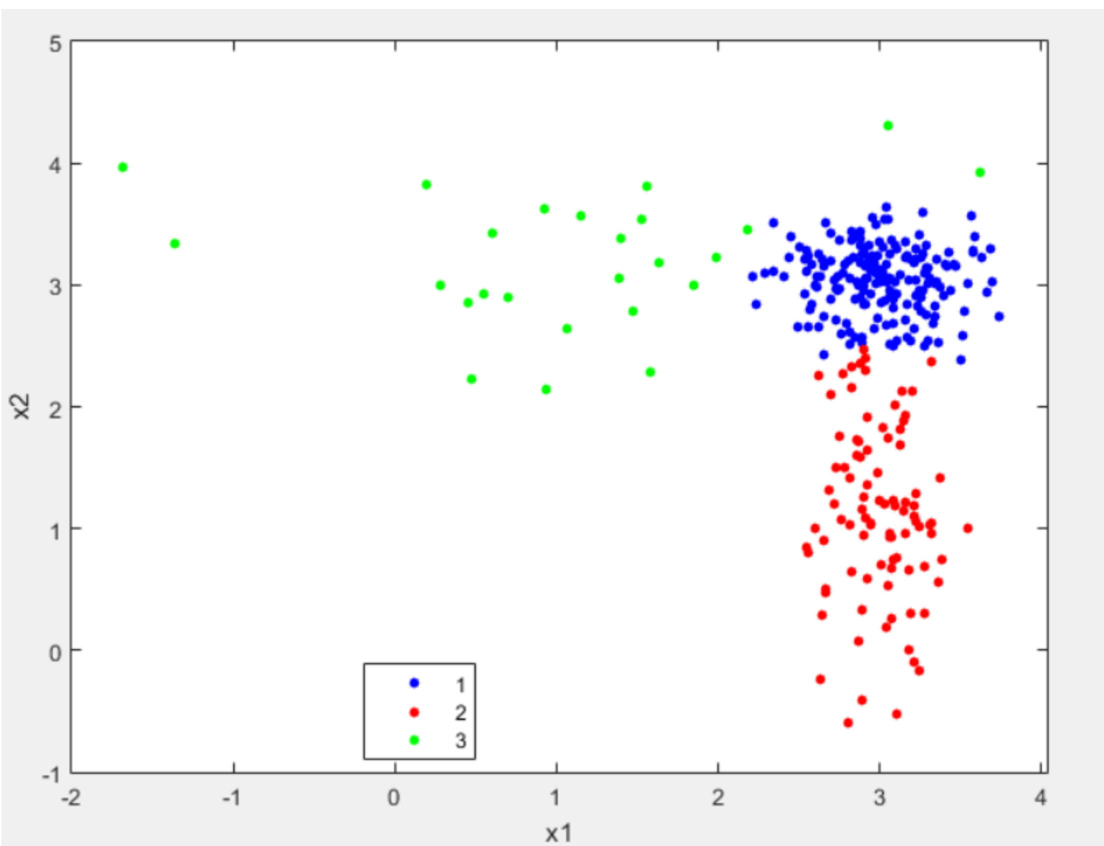
Estimates for $\mu_{i|c}$

	Cluster 1	Cluster 2	Cluster 3
x_1	3.0079	2.9972	1.2977
x_2	3.0503	1.2824	3.1541

Estimates for $\sigma^2_{i|c}$

	Cluster 1	Cluster 2	Cluster 3
x_1	0.1000	0.0473	1.3766
x_2	0.0733	0.7215	0.2566

Plot of Gaussian Clustered Data



Clustering by Generalized K-Means

Estimates for π_c

Cluster 1	Cluster 2	Cluster 3
0.6067	0.3100	0.0833

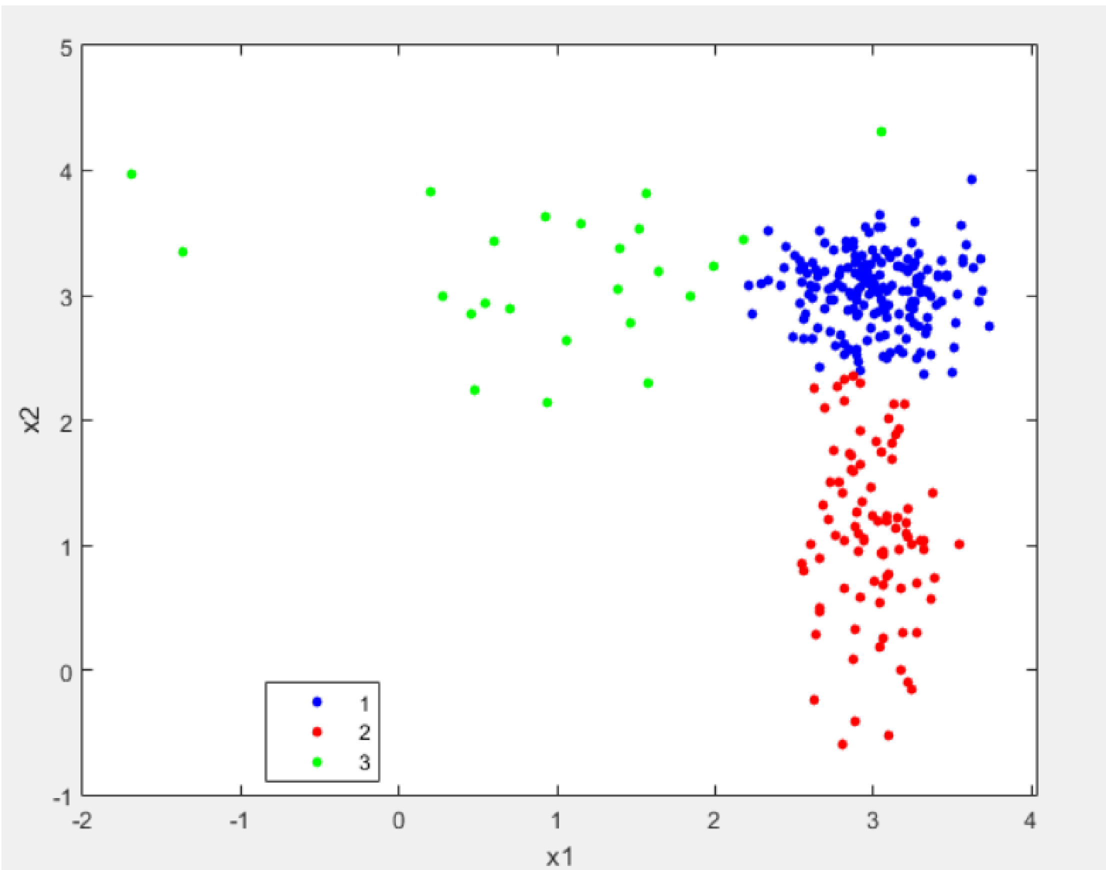
Estimates for $\mu_{i|c}$

	Cluster 1	Cluster 2	Cluster 3
x_1	3.0048	2.9926	1.1013
x_2	3.0386	1.1302	3.2172

Estimates for $\sigma^2_{i|c}$

	Cluster 1	Cluster 2	Cluster 3
x_1	0.0985	0.0464	1.2327
x_2	0.0732	0.4972	0.2980

Plot of K-Means Clustered Data



Analysis

The result of the two methods was very similar, but a few things were slightly different. Three points were changed from cluster 2 to cluster 1 and one point was changed from cluster 3 to cluster 1. In the Gaussian method, $\sigma^2_{2|2}$ was 0.7215, but it sharply decreased to 0.4972 in the K-Means method. The values for π_c and $\mu_{i|c}$ only had very slight differences.

Running the Code

The code to run the Gaussian Naive-Bayes method of clustering is found in em.m, and the code to run the K-Means method is found in kmeans.m. fcondjoin.m is the function to find the conditional joint probability of two x values based on provided $\sigma^2_{i|c}$ and $\mu_{i|c}$, and is used by both em.m and kmeans.m. This code was written, tested, and can be run in Matlab. kmeans.m should be run after em.m because it makes use of the variables saved in the workspace after em.m finishes running.