# Problem Set 2

Jun Wang
Jwang3316@gatech.edu

1. Probability: A is 50%, B and C is 25%. So, I set 1 bit for A and set 2 bits for each B and C. The bit string for A is 0, for B is 10 and for C is 11. The entropy of this coding is – 0.5*log2(0.5) – 0.25*log2(0.25) – 0.25*log2(0.25) = 1.5

2. K means can be seen as an EM algorithm that require the Gaussian densities to be spherical for all features, and in the maximization step a mean of data points was used to update the centroid parameter. The centroid location is the only parameter to be estimated, no covariance or slope for any features.

3.
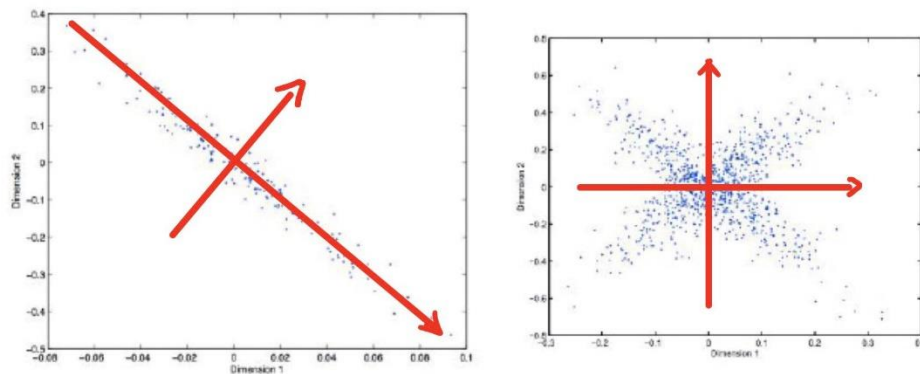
4. a.
   I think the most likely method for this dataset is Hierarchical clustering. Because the distances between points in two clusters have a very sparse distribution. Hierarchical clustering can identify the two classes perfectly since there will always be a very close point pair within the two classes.
   b.
   EM and K means. Because it looks like we only need to estimate two centroid locations for both methods. Given the complexity of Gaussian distribution and posterior probability, I think K means will converge faster than EM. All the hierarchical clustering methods will have problem distinguishing the point that are very close but comes from two classes.
   c.
   GMM. Because the two clusters are overlapping and only GMM can work well on this.

7.
MDP
Iteration 0:

0	0	0
0	0	0
0	0	G

Iteration 1

0	0	0
0	0	10
0	10	G

Iteration2

0	0	8
0	8	10
8	10	G

Iteration3

0	6.4	8
6.4	8	10
8	10	G