

**Statistical Pattern Recognition (CE-725)
Department of Computer Engineering
Quiz #9 (Clustering & GMM) - Spring 2012**

1. (45 points) Consider the dataset $\{0,4,5,20,25,39,43,44\}$. Suppose we want the two top level clusters from this dataset. What will single link, complete link and average link output as the two clusters?

Sol:

Single link: $\{0,4,5\}$ $\{20,25,39,43,44\}$

Complete link: $\{0,4,5\}$ $\{20,25,39,43,44\}$

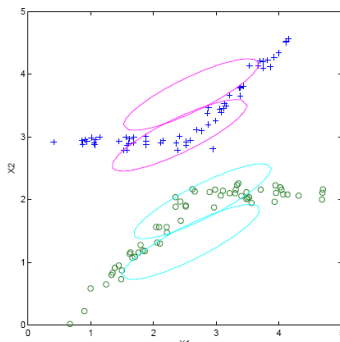
Average Link: $\{0,4,5\}$ $\{20,25,39,43,44\}$ or $\{0,4,5,20,25\}$ $\{39,43,44\}$

2. (25 points) If single link and complete link give the same two clusters, does it follow that average link will output the same cluster? Explain.

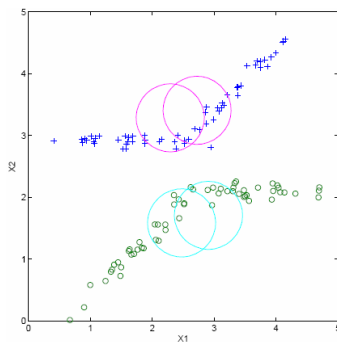
Sol:

No. See the solution of above question.

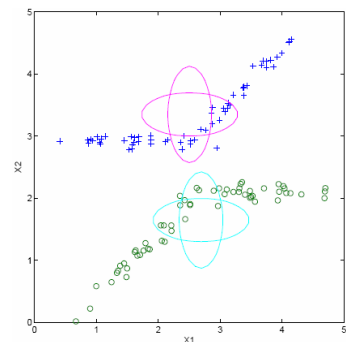
3. (30 points) In the following plots '+' and 'o' denote positive and negative labels, respectively. We want to use two Gaussian distributions for positive examples and two Gaussian distributions for negative examples. We use EM algorithm to iteratively update parameters and try different initializations. The top row shows three different initializations and the bottom row shows three possible models after the first iteration. For each initialization on the top, draw an arrow to the model on the bottom that will result after the first EM iteration.



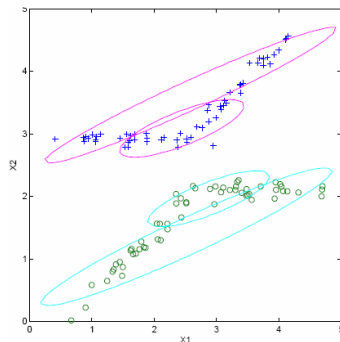
a1



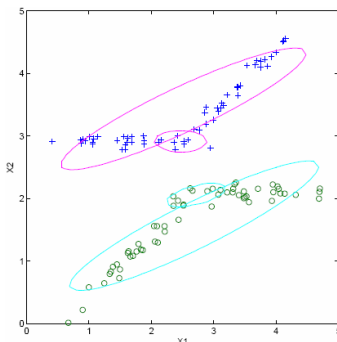
a2



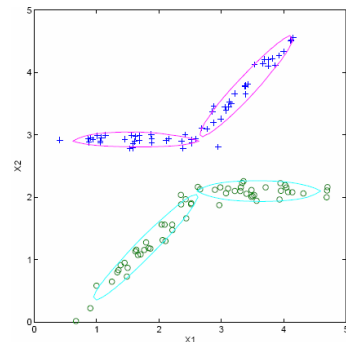
a3



b1



b2



b3

Sol: a1: b1, a2: b3, a3: b2