# CS7641 ML Problem Set II by Georgia Tech

*T. Ruzmetov*

*December 5, 2017*



## Problem 1

**You have to communicate a signal in a language that has 3 symbols A, B and C. The probability ofobserving A is 50% while that of observing B and C is 25% each. Design an appropriate encoding for this language. What is the entropy of this signal in bits?**

a. Build a variable width encoding by constructiang a tree which assigns bits based on probabilities as shown in Fig.1. Then average bit size would be $0.5 * 1 + 2 * 0.25 + 2 * 0.25 = 1.5$

b. $Entropy = -\sum_i P_i \log_2 P_i = -(0.5 \log_2 0.5 + 0.25 \log_2 0.25 + 0.25 \log_2 0.25) = 1.5$

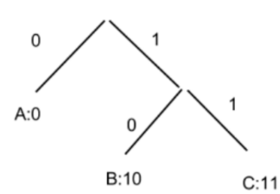


Figure 1:

## Problem 2

**Show that the Kmeans procedure can be viewed as a special case of the EM algorithm applied to an appropriate mixture of Gaussian densities model.**

In K-means we assign labels to data points based on how close they are to a centroid and iterate by moving centroids to new optimum location(center of class) until it convergence is acheived. For EM, we first calculated expectation values per data point for all clusters. Closer the data point to a centroid location higher expectation it has.

$$P(x = x_i | \mu = \mu_j) = e^{-\frac{(x_i - \mu_j)^2}{2\sigma^2}}$$

$$E(Z_{ij}) = \frac{P(x = x_i | \mu = \mu_j)}{\sum_i P(x = x_i | \mu = \mu_j)}$$

Then we calculate new centroid location per cluster using previously calculated expectation values as a weight for our statistical average. In contrast to K-means, here clusters don't need to claim data points based on closeness because it is kind of taken care of, where far points have very low expectaion(weights) so they don't contribute to $\mu_j$ update as much as close points.

$$\mu_j = \frac{\sum_i E(Z_{ij}) x_i}{\sum_i E(Z_{ij})}$$

Basically, say we have 2 clusters, then all data points belong to a cluster 1 with some probability, where far points have low probability and close points have high probability respectively. Same reasoning applies to cluster 2. This simply inplies that boundary between clusters is soft. In contrast, Kmeans has a strict boundary condition based on distance.

Now, we can take EM Gaussian densities model and change soft boundary condition to a hard one by claiming data points based on some strict condition. Then it becomes same as K-means.

## Problem 3

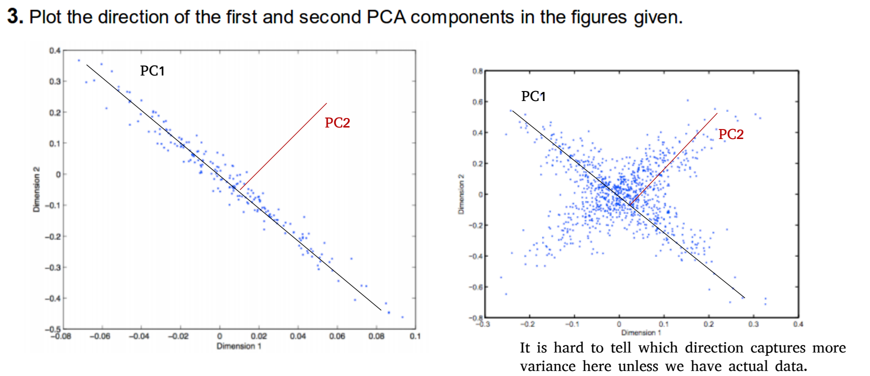


Figure 2:

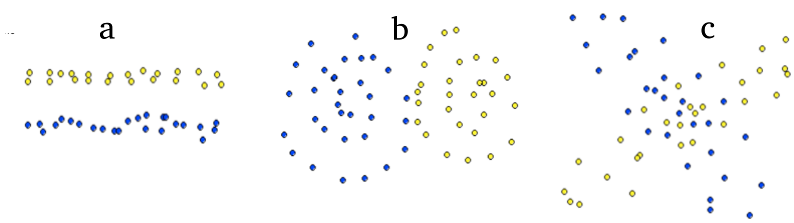


Figure 3:

## Problem 4

**Which clustering method(s) is most likely to produce the following results at k = 2? Choose the most likely method(s) and briefly explain why it/they will work better where others will not in at most 3 sentences. Here are the five clustering methods you can choose from:**

- **Hierarchical clustering with single link**
- **Hierarchical clustering with complete link**
- **Hierarchical clustering with average link**
- **K-means**
- **EM**

a. Hierarchical clustering with single linkage works best for this case, because neighbouring points are very close to each other. EM with elliptic gaussian will also work, but migh need smart initialization.

b. K-means and EM GMM will work best.

c. EM is most likely the best candidate here, since points are mixed.

## Problem 7

**Consider the following simple grid world problem. (Actions are N, S, E, W and are deterministic.) Our goal is to maximize the following reward:**

1. **10 for the transition from state 6 to G**
2. **10 for the transition from state 8 to G**
3. **0 for all other transitions**

| | | |
|---|---|---|
| S | 2 | 3 |
| 4 | 5 | 6 |
| 7 | 8 | G |

a. **Draw the Markov Decision Process associated to the system.**

b. **Compute the value function for each state for iteration 0, 1, 2 and 3 with** $\gamma = 0.8$

| | | |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 0 | 0 |
| 0 | 0 | 10 |

Figure 4: iter0

| | | |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 0 | 8/3 |
| 0 | 8/3 | 10 |

Figure 5: iter1

| | | |
|---|---|---|
| 0 | 0 | 16/15 |
| 0 | 16/15 | 8/3 |
| 16/15 | 8/3 | 10 |

Figure 6: iter2

| | | |
|---|---|---|
| 0 | 0.57 | 16/15 |
| 0.57 | 16/15 | 3.23 |
| 16/15 | 3.23 | 10 |

Figure 7: iter3