

cse7/4641 Problem Set 2

Bin Xie

April 2019

Answer 1

Since the probability of observing A is 50% while that of observing B and C is 25% each, we can encode A with one bit as 0, B with 2 bits as 10 and C with 2 bits as 11.

The entropy of this signal in bits is: $1 * 0.5 + 2 * 0.25 + 2 * 0.25 = 1.5$ bits.

Answer 2

EM algorithm applied to the appropriate mixture of Gaussian densities model can provide probabilistic interpretation for K-means. If we assume that all Gaussian components have σI as their covariance matrices and $\sigma \rightarrow 0$, then we only have to estimate μ_k , i.e. the EM algorithm for Gaussian Mixture Model(GMM) parameter estimation is simplified to K-means. So the EM algorithm for GMM can be the soft K-means and it gives the probabilities for each data to clusters represented by the corresponding Gaussian distribution while K-means just distributes each data into the corresponding cluster.

Answer 3

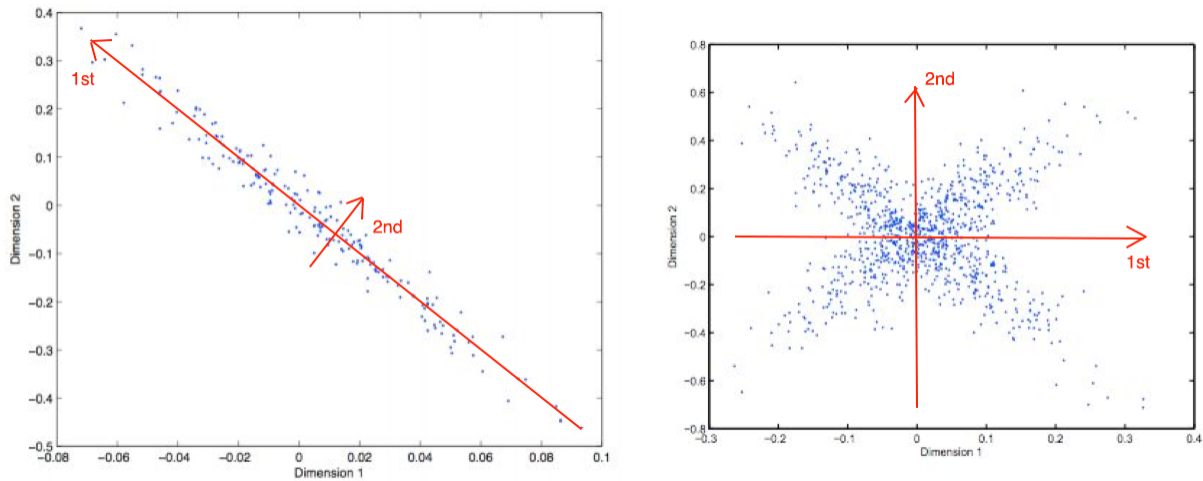


Figure 1: First and second PCA components in the figures given

Answer 4

a. Hierarchical clustering with single link is most likely to produce the following results at $k = 2$ since the data has a sparse distribution. K-means and EM will have problem in finding the right centroids of the clusters and assigning the cluster for points which are from different clusters but have shorter distance than

the points from the same cluster. Hierarchical clustering with complete link or with average link may also work but has more computation.

b. K-means or EM is most likely to produce the following results at $k = 2$ since the centroids of two clusters are distinguished. All hierarchical clustering methods can have problem in assigning the clusters for points from two different clusters but have close distance.

c. EM is most likely to produce the following results at $k = 2$ since it gives the probabilities for points to each cluster, which supports mixed membership. However, the other four methods can not handle with the overlapping points since the hierarchical clustering can have problem in assigning the overlapping points which have close distance while K-means can have problem in finding the centroids.

Answer 6

For total exploration:

$$Q(s, a1) = 1 + 1 * 0.25 + 2 * 0.25 + 1 * 0.25 + 2 * 0.25 = 2.5$$

$$Q(s, a2) = 0 + 1 * 0.25 + 2 * 0.25 + 0 * 0.25 + 6 * 0.25 = 2.25$$

For greedy exploration:

$$Q(s, a1) = 1 + 1 * 0.25 + 2 * 0.25 + 1 * 0.25 + 2 * 0.25 = 2.5$$

$$Q(s, a2) = 0 + 0 * 0.5 + 6 * 0.5 = 3$$

Answer 7

a. The Markov Decision Process associated to the system is shown as follows:

b. The value functions for each state for iteration 0, 1, 2 and 3 with $\gamma = 0.8$ are shown as follows:

iteration0:

| | | |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 0 | 0 |
| 0 | 0 | G |

iteration1:

| | | |
|---|----|----|
| 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 |

Answer 8