Arvind Sai DOODA A20553046

CS584 Machine Learning

# Assignment-3

1. Problem1: Lloyd's Method

Given a data set with seven data points $\{x_1, \ldots, x_7\}$ and the distance between all pairs of data points are in the following table.

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ |
|---|---|---|---|---|---|---|---|
| $x_1$ | 0 | 5 | 3 | 1 | 6 | 2 | 3 |
| $x_2$ | 5 | 0 | 4 | 6 | 1 | 7 | 8 |
| $x_3$ | 3 | 4 | 0 | 4 | 3 | 5 | 6 |
| $x_4$ | 1 | 6 | 0 | 4 | 3 | 5 | 6 |
| $x_5$ | 6 | 1 | 3 | 7 | 0 | 8 | 9 |
| $x_6$ | 2 | 7 | 5 | 1 | 8 | 0 | 1 |
| $x_7$ | 3 | 8 | 6 | 2 | 9 | 1 | 0 |

Assume

no. of cluster $K=2$.

cluster centres are initialized to be $x_3$ and $x_6$.

Q. 5 ~~Points~~..

Given,

no of clusters $(K) = 2$

initialized cluster centres

Cluster 1 $(x_3) = 3$

cluster 2 $(x_6) = 6$

from above the distance matrix:

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ |
|---|---|---|---|---|---|---|---|
| $x_1$ | 0 | 5 | 6 | 27 | 3 | 4 | 6 |
| $x_2$ | 5 | 0 | 4 | 6 | 1 | 7 | 8 |
| $x_3$ | 3 | 4 | 0 | 4 | 3 | 5 | 6 |
| $x_4$ | 1 | 6 | 1 | 0 | 7889 | 8 | 9 |
| $x_5$ | 2 | 7 | 5 | 1 | 0 | 1 | 3 |
| $x_6$ | 7 | 8 | 9 | 7889 | 1 | 0 | 2 |
| $x_7$ | 3 | 8 | 6 | 2 | 1 | 0 | 1 |

1. 5 Points What's the two clusters formed at the end of the first iteration of Lloyd's algorithm?

First Iteration:

In first iteration of the clustering Process:

cluster 1 ( $x_3$ center ) contains data points:

$$\{x_1, x_3, x_5\}$$

cluster 2 ( $x_6$ center ) contain data points:

$$\{x_2, x_4, x_6, x_7\}$$

∴ The clusters are updated as.

The center for cluster one (1) calculated as.

mean of its data point : $(x1 + x3 + x5)/3$

$$= \frac{x_1 + x_3 + x_5}{3}$$ ∴ cluster for center 1 is 3.

$$= \frac{3+3+3}{3} \Rightarrow \frac{9}{3} = 3$$

The center for cluster 2. calculated as

mean of its data point : $(x_2 + x_4 + x_6 + x_7)/4$

$$= \frac{x_2 + x_4 + x_6 + x_7}{4}$$

$$= \frac{6+1+1+1}{4} = \frac{9}{4} = 2.25$$

Cluster for Centre 2 = 2.25

2. 5 Points. What's the Two clusters formed at the end of the second iteration of lloyd's algorithm?

In second Iteration of clustering Process.

cluster 1 ( $x_3$ (center=5)):

Data Points will be : { $x_1, x_3, x_5$ }

cluster 2 (center=2.25)

Data points will be { $x_2, x_4, x_6, x_7$ }

To get new cluster centers we should do the mean of data points in each cluster.

* New cluster 1 center $= \dfrac{x_1 + x_3 + x_5}{3}$

$= \dfrac{3+3+3}{3} \Rightarrow 3$

cluster 1 center = 3.

* New cluster 2 center $= \dfrac{x_2 + x_4 + x_6 + x_7}{4}$

$= \dfrac{6+1+1+1}{4}$

$= 2.25$

∴ we can see that there is no change in

the both Iteration 1 and Iteration 2. It can leads to converge of the algorith (It converged).

so we can add 2 and 3 is 5.

3. 10 Points what's the two clusters formed when the Lloyd's algorithm converges?

result after convergence:

upon convergence of the lloyd's algorithm Two distinct clusters emerge:

cluster 1 (center at 3) contains data points $\{x_1, x_3, x_5\}$.

cluster 2 (center at 2.25) contains data points $\{x_2, x_4, x_6, x_7\}$

These are the Two clusters formed when the lloyd's algorith converges.

Problem 2 (15 Points):

Solution:

In order to establish that $P(x)$ conforms to a Gaussian Mixture Model (GMM), we must calculate $P(x)$ by aggregating $P(z)$ $P(x|z)$ across all feasible $z$ values.

$$P(z) = \pi K = 1K$$
$$= 1$$

Now, let's consider $P(x|z)$. it's given

$$P(x|z) = (\pi K = 1 K N(x | \mu K, \Sigma K) zK$$

Since each $zK$ can take value 0 or 1 the sum over all possible value of $z$ can be written as sum of all $zK$ possible combinations.

$$P(x) = \Sigma z_1 \Sigma z_2 \ldots \Sigma zK P(z) P(x|z)$$

by using definition of $P(z)$ and $P(x|z)$.

$$P(x) = \Sigma z_1 \Sigma z_2 \ldots \Sigma zK (\pi K = 1 K \pi K zK)$$
$$(\pi K = 1 K N(x | \mu K, \Sigma K) zK)$$

$\therefore$ $z_K$ is 1 and all other are 0.

$z_K \in \{0, 1\}$

$\Rightarrow \Sigma K = 1 \, K z_K = 1$

$\therefore$ $P(x) = \Sigma K = 1 K (\pi_K N (x | \mu_K, \Sigma_K))$

$\therefore$ $P(x)$ is indeed a GMM with specific form.

Hence Proved.

$P(x)$, obtaining by summin $P(z) \, P(x|z)$ over all possible values of $z$, is a GMM

$$P(x) = \sum_{z} P(z) \, P(x|z)$$

$$= \sum_{K=1}^{K} \pi_K N(x | \mu_K, \Sigma_K)$$

$\therefore$