## Assi5

## Xiaolong Yang

# Conceptual Questions (5 points)

## 1.a

K‐Means is sensitive to noisy data and outliers, because it calculated the center by average of the dataset, and the outliers or noise can affect the mean, thus the K centers would be meaningless.

## 1.b

No.

## 1.c

Hierarchical clustering can be applied in more distance function while k-means get a good result merely in Euclidean distance.

Hierarchical clustering did not need the specific K – the number of clusters.

# Advanced Classification: Perceptron (5 points)

|  |  |  |
| --- | --- | --- |
| *x*1 | *x*2 | **y** |
| 0 | 0 | + |
| 0 | 1 | + |
| 1 | 0 | + |
| 1 | 1 | - |

Table 1: Data points with class labels

*If y != sign(wTx), update w = w + η\*x\*y*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | iter1 | sign | *η\*x\*y-4* | iter2 | sign | *η\*x\*y-1* | iter3 | sign |
| W0 | 0.25 |  | -0.5 | -0.25 |  | 0.5 | 0.25 |  |
| W1 | 0.25 |  | -0.5 | -0.25 |  | 0 | -0.25 |  |
| W2 | 0.25 |  | -0.5 | -0.25 |  | 0 | -0.25 |  |
| Y1 | 0.25 | 1 |  | -0.25 | -1 |  | 0.25 | 1 |
| Y2 | 0.5 | 1 |  | -0.5 | -1 |  | 0 | 1 |
| Y3 | 0.5 | 1 |  | -0.5 | -1 |  | 0 | 1 |
| Y4 | 0.75 | 1 |  | -0.75 | -1 |  | -0.25 | -1 |

# Hierarchical Agglomerative Clustering and B-Cubed Evaluation (8 points)

|  |  |  |  |
| --- | --- | --- | --- |
| **Point** | **x** | **y** | **Ground Truth** |
| P1 | 1 | 1 | C1 |
| P2 | 1 | 2 | C1 |
| P3 | 2 | 1 | C1 |
| P4 | 5 | 1 | C2 |
| P5 | 3 | 2 | C1 |
| P6 | 5 | 2 | C2 |
| P7 | 3 | 3 | C1 |

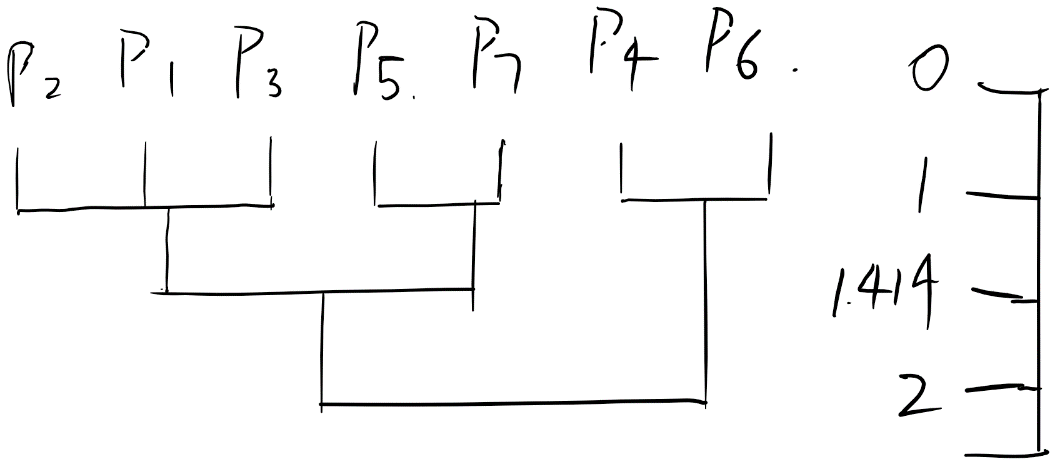
Table 2: Data Points

## 3.a.

L1 (value = 1): C1(P1, P2, P3), C2(P4, P6), C3(P5, P7);

L2 (value = 1.414): C4(P1, P2, P3, P5, P7), C5(P4, P6);

L3 (value = 2): C6(P1,P2, P3, P4, P5, P6, P7).



## 3.b

There would be C1(P2, P1, P3), C2(P4, P6), C3(P5, P7)

## 3.c

Precision 1= 3/3 =1

Precision 2= 2/2 = 1

Precision 3= 2/2 = 1

Final Precision = 1

Recall1 = 3/5 = 0.6

Recall2 = 2/2 = 1

Recall3 = 2/5 = 0.4

Final Recall = (Recall 1\*3+ Recall 2\*2+ Recall 3\*2) \* (3+2+2) = 0.657