CS 5525
Solutions to Homework Assignment 4
Meghendra Singh

November 8, 2016

[15] 1.

(a) The estimated conditional probabilities are:

Cond. prob.	**Value**
P(A|+)	$3/5 = 0.6$
P(B|+)	$1/5 = 0.2$
P(C|+)	$4/5 = 0.8$
P(A|-)	$2/5 = 0.4$
P(B|-)	$2/5 = 0.4$
P(C|-)	$5/5 = 1$

(b) By Bayes theorem we have:

$$P(C_j|A_1, A_2, ..., A_n) = \frac{P(C_j) * P(A_1, A_2, ..., A_n|C_j)}{P(A_1, A_2, ...A_n)} \tag{1}$$

and the Naïve (independence among attributes) assumption gives us:

$$P(A_1, A_2, ..., A_n|C_j) = P(A_1|C_j)P(A_2|C_j)...P(A_n|C_j) \tag{2}$$

By (1) and (2) we have,

$$P(C_j|A_1, A_2, ..., A_n) = \frac{P(C_j) * P(A_1|C_j)P(A_2|C_j)...P(A_n|C_j)}{P(A_1, A_2, ...A_n)}$$

After ignoring the constant denominator we can estimate the `posterior` probability for a class C_j as:

$$P(C_j|A_1, A_2, ..., A_n) = P(C_j) * P(A_1|C_j)P(A_2|C_j)...P(A_n|C_j)$$

Therefore,

$$\begin{aligned} P(+|A=0, B=1, C=0) &= P(A=0|+)P(B=1|+)P(C=0|+)P(+) \\ &= \frac{2}{5} * \frac{1}{5} * \frac{1}{5} * \frac{1}{2} \\ &= \frac{1}{125} = 0.008 \end{aligned}$$

And,

$$\begin{aligned}P(-|A=0,B=1,C=0) &= P(A=0|-)P(B=1|-)P(C=0|-)P(-)\\ &= \frac{3}{5}*\frac{2}{5}*0*\frac{1}{2}\\ &= 0\end{aligned}$$

Since, $P(+|A=0,B=1,C=0) > P(-|A=0,B=1,C=0)$, the test sample will be labeled as '+'.

(c) M-estimate is given by: $P(A_i|C) = \frac{N_{ic}+m*p}{N_c+m}$, given $m=4$ and $p=1/2$ we get the following conditional probabilities:

Cond. prob.	**Value**
P(A|+)	$5/9 = 0.555$
P(B|+)	$3/9 = 0.333$
P(C|+)	$6/9 = 0.666$
P(A|-)	$4/9 = 0.444$
P(B|-)	$4/9 = 0.444$
P(C|-)	$7/9 = 0.777$

(d)The `posterior` probabilities for the two classes, given the test sample $(A=0,B=1,C=0)$ are as follows:

$$\begin{aligned}P(+|A=0,B=1,C=0) &= P(A=0|+)P(B=1|+)P(C=0|+)P(+)\\ &= \frac{4}{9}\times\frac{3}{9}\times\frac{3}{9}\times\frac{1}{2}\\ &= \frac{18}{729} = 0.0246\end{aligned}$$

$$\begin{aligned}P(-|A=0,B=1,C=0) &= P(A=0|-)P(B=1|-)P(C=0|-)P(-)\\ &= \frac{5}{9}\times\frac{4}{9}\times\frac{2}{9}\times\frac{1}{2}\\ &= \frac{20}{729} = 0.0274\end{aligned}$$

Since, $P(+|A=0,B=1,C=0) < P(-|A=0,B=1,C=0)$, the test sample will be labeled as '$-$'.

(e) In the first method, absence of a training sample can make the joint conditional probability zero. While the second method (using m-estimates) compensates for such cases and prevents the `posterior` probability expresion from becoming zero just becasue of one of the conditional probabilities was zero. (b) and (d) above are an example of such a case. Therefore, the second method (m-estimates) is better than the first method.

[15] 2.

(a) The contingency tables for the six rules are as follows:

$\{b\} \rightarrow \{c\}$	c	$\overline{c}$
b	3	4
$\overline{b}$	2	1
$Support = \frac{3}{10} = 0.3$		
$Confidence = \frac{3}{7} = 0.428$		

$\{a\} \rightarrow \{d\}$	d	$\overline{d}$
a	4	1
$\overline{a}$	5	0
$Support = \frac{4}{10} = 0.4$		
$Confidence = \frac{4}{5} = 0.8$		

$\{b\} \rightarrow \{d\}$	d	$\overline{d}$
b	6	1
$\overline{b}$	3	0
$Support = \frac{6}{10} = 0.6$		
$Confidence = \frac{6}{7} = 0.857$		

$\{e\} \rightarrow \{c\}$	c	$\overline{c}$
e	2	4
$\overline{e}$	3	1
$Support = \frac{2}{10} = 0.2$		
$Confidence = \frac{2}{6} = 0.333$		

$\{c\} \rightarrow \{a\}$	a	$\overline{a}$
c	2	3
$\overline{c}$	3	2
$Support = \frac{2}{10} = 0.2$		
$Confidence = \frac{2}{5} = 0.4$		

(b) Rules ranked in decreasing order according to support:

$\{b\} \rightarrow \{d\} > \{a\} \rightarrow \{d\} > \{b\} \rightarrow \{c\} > \{e\} \rightarrow \{c\} = \{c\} \rightarrow \{a\}$

Rules ranked in decreasing order according to confidence:

$\{b\} \rightarrow \{d\} > \{a\} \rightarrow \{d\} > \{b\} \rightarrow \{c\} > \{c\} \rightarrow \{a\} > \{e\} \rightarrow \{c\}$

---

[15] 3.

The following table gives the support, confidence and lift for the six rules, along with the respective ranks according to each measure (greater magnitude results in lower rank):

Rule	Support	Rank	Confidence	Rank	Lift	Rank
bread $\rightarrow$ milk	0.32	1	0.4	4	1	2
milk $\rightarrow$ bread	0.32	1	0.8	1	1	2
coke $\rightarrow$ pepsi	0.08	2	0.167	6	0.42	3
pepsi $\rightarrow$ coke	0.08	2	0.2	5	0.42	3
wine $\rightarrow$ caviar	0.06	3	0.43	3	5.36	1
caviar $\rightarrow$ wine	0.06	3	0.75	2	5.36	1

---

[10] 4.

The support and confidence for the three rules are as follows:

Rule 1:
$\{(5 \leq A \leq 8), B = 1\} \rightarrow \{C = 1\}$: Support $= \frac{2}{12} = 0.166$; Confidence $= \frac{2}{2} = 1$

Rule 2:
$\{(A \textit{ is odd}), B = 1\} \rightarrow \{C = 1\}$: Support $= \frac{2}{12} = 0.166$; Confidence $= \frac{2}{3} = 0.66$

Rule 3:
$\{A \textit{ is even}), C = 1\} \rightarrow \{B = 1\}$: Support $= \frac{2}{12} = 0.166$; Confidence $= \frac{2}{4} = 0.5$

---

[15] 5.

Part I: The following table shows the Euclidean distance between each point and the three initial centroids ($A_1,\ B_1$ *and* C_1) and the cluster assignment after the first iteration, A_1 is considered the centroid of the first cluster, B_1 the centroid of the second cluster and C_1 the centroid of the third cluster:

Euclidean distance	A_1	B_1	C_1	Cluster assignment
A_1	0	3.606	8.062	1
A_2	5	4.243	3.162	3
A_3	8.485	5	7.280	2
B_1	3.606	0	7.211	2
C_1	8.062	7.211	0	3
C_2	2.236	1.414	7.616	2

New centroids after the first iteration are: $(2, 10)$, $(5.67, 7)$ and $(1.5, 3.5)$

Part II: The final cluster assignments after *k-means* clustering are:

$\{1, 2, 3, 5\} \rightarrow$ cluster 1 and $\{9\} \rightarrow$ cluster 2

(i) The final centroids are: **2.75** and **9** for clusters 1 and 2 respectively
(ii) Cohesion or within clusters sum of squares is **8.75**
(iii) Separation or between clusters sum of square is **31.25**

---

[20] 6.

(a) The initial proximity matrix is as follows:

	0.1	**1**	**1.7**	**3.4**	**3.9**	**4.7**
0.1	0	0.9	1.6	3.3	3.8	4.6
1	0.9	0	0.7	2.4	2.9	3.7
1.7	1.6	0.7	0	1.7	2.2	3.0
3.4	3.3	2.4	1.7	0	0.5	1.3
3.9	3.8	2.9	2.2	0.5	0	0.8
4.7	4.6	3.7	3.0	1.3	0.8	0

(i) Cluster memberships obtained after single linkage, agglomerative, hierarchical clustering are: $\{0.1, 1, 1.7\} \rightarrow$ cluster 1 and $\{3.4, 3.9, 4.7\} \rightarrow$ cluster 2.
(ii) The dendrogram is as follows:

(b) The initial proximity matrix computed for the four data points A: (0 2 0 0); B. (2 0 1 2); C: (2 1 0 2); D: (2 2 1 0), using the cosine similarity measure is as follows:

	A	**B**	**C**	**D**
A	1	0	0.33	0.66
B	0	1	0.88	0.55
C	0.33	0.88	1	0.66
D	0.66	0.55	0.667	1

Upon single linkage hierarchical clustering we get the following proximity matrices after two successive merge operations:

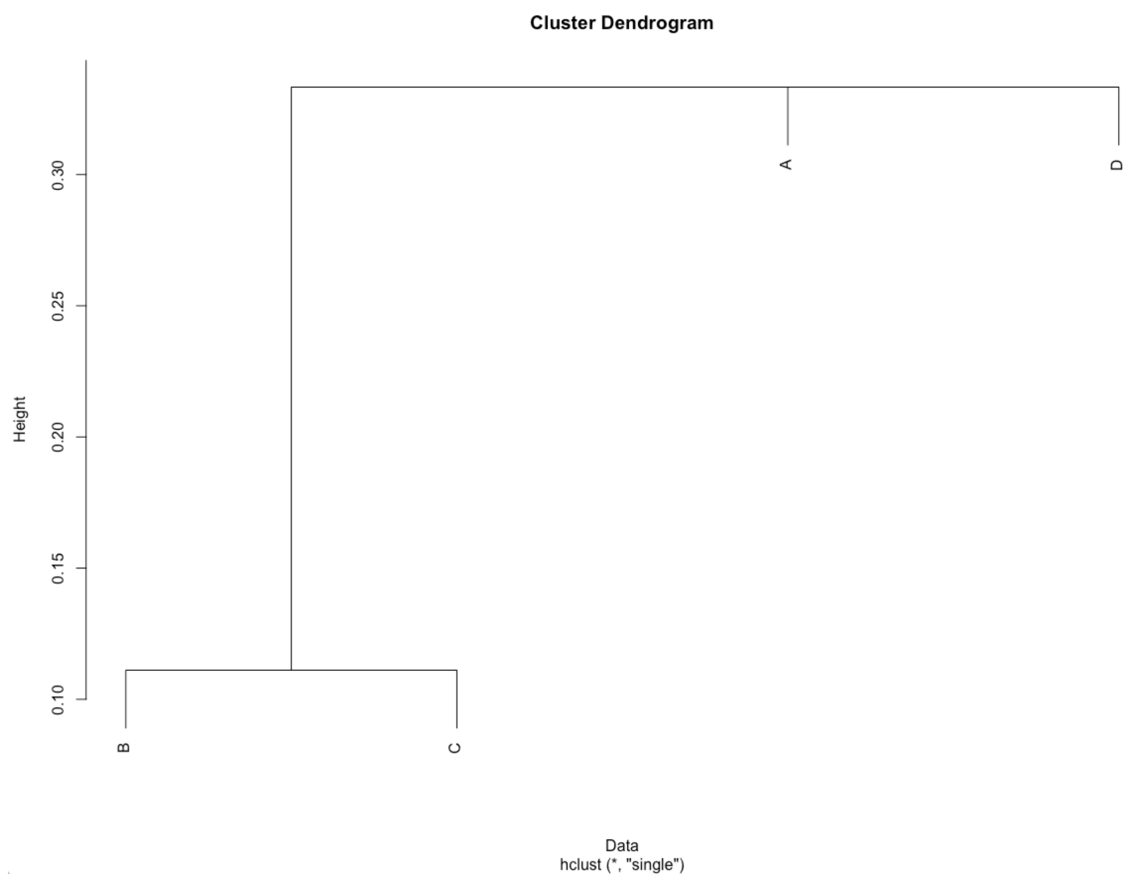
	A	**BC**	**D**
A	1	0.33	0.66
BC	0.33	1	0.66
D	0.66	0.66	1

	AD	**BC**
→ **AD**	1	0.66
BC	0.66	1

or

	A	**BCD**
A	1	0.66
BCD	0.66	1

There, are two merge operations are possible since we have **0.66** as the maximum similarity in the proximity matrix between **A** and **D** as well as between **BC** and **D**. The dendrogram for the first case is as follows:



(c) Following are the proximity matrices, there reductions on successive merge operations and the cluster dendrograms for the single and complete linkage hierarchical clustering:
(i) **Single linkage:**

	p1	p2p5	p3	p4
p1	1	0.35	0.41	0.55
p2p5	0.35	1	0.85	0.76
p3	0.41	0.85	1	0.44
p4	0.55	0.76	0.44	1

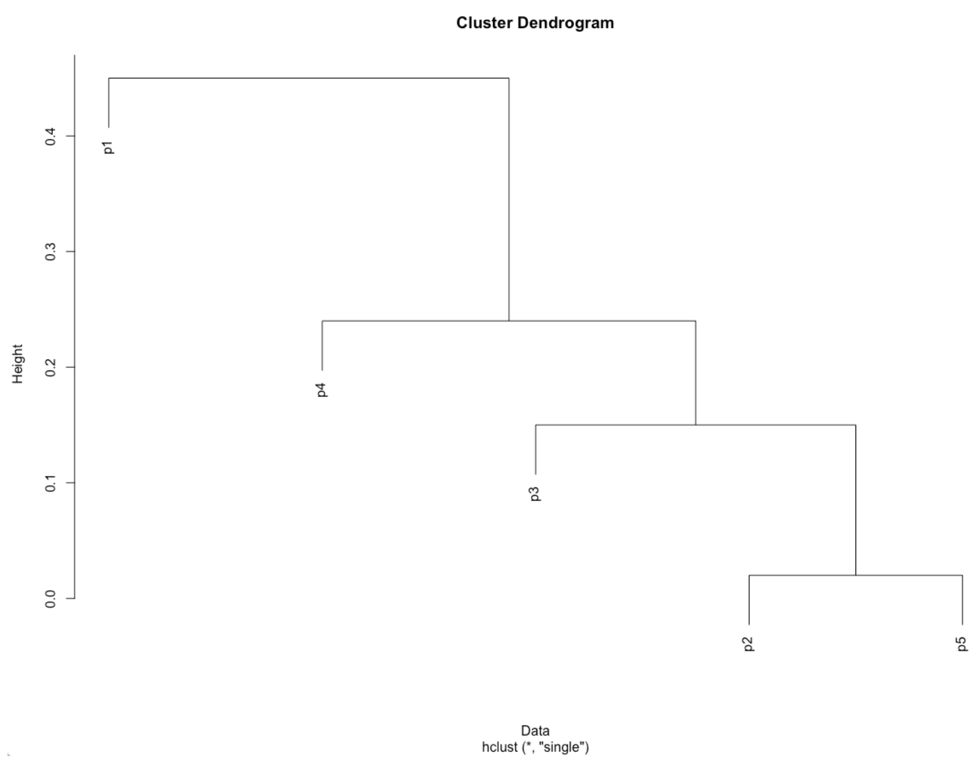
$\rightarrow$

	p1	p2p3p5	p4
p1	1	0.41	0.55
p2p3p5	0.41	1	0.76
p4	0.55	0.76	1

$\rightarrow$

	p1	p2p3p4p5
p1	1	0.55
p2p3p4p5	0.55	1

The resultant cluster dendrogram is as follows:



(i) **Complete linkage:**

	p1	p2p5	p3	p4
p1	1	0.1	0.41	0.55
p2p5	0.1	1	0.64	0.47
p3	0.41	0.64	1	0.44
p4	0.55	0.47	0.44	1

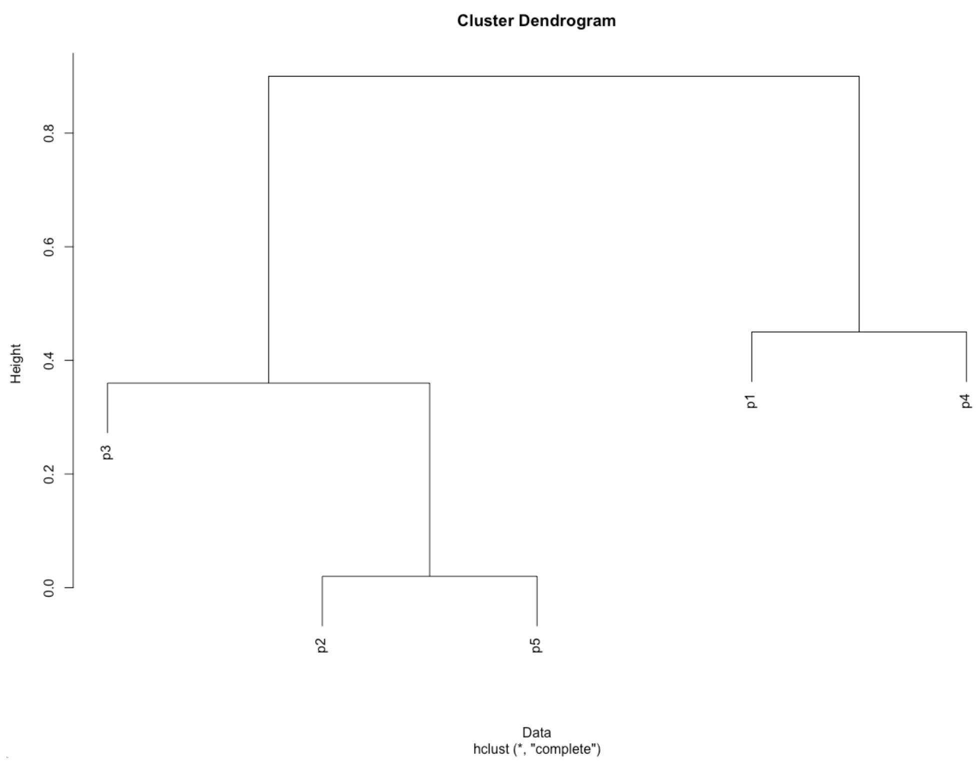
$\rightarrow$

	p1	p2p3p5	p4
p1	1	0.1	0.55
p2p3p5	0.1	1	0.44
p4	0.55	0.44	1

$\rightarrow$

	p1p4	p2p3p5
p1p4	1	0.1
p2p3p5	0.1	1

The resultant cluster dendrogram is as follows:



[10] 7.

Upon applying DBSCAN with $\varepsilon = 0.15$ and MinPts $= 4$ we get the following core, border and noise points:

Core points	**Border points**	**Noise points**
x, t, r, q, s, a, b, c, d, e, f, g, h, i, j, k and l	w, v, y, z, u, p and m	n and o

The following figure shows the core points, border points and noise points in blue, green and red colors respectively: