23-S1-Q4

Q (a)(i) HAC - single.

(ii) complete

(b) PCA

problem

achieve.

(c) (i) Naive. → 1 0 1

(ii) $p(ps) = 20\%$

60% ps → ≥ 25h

30% n-ps → ≥ 25%

posterior probability

Solution (a) (i)

① Initialization : Each object is a cluster

② Iteration 1 : Merge two clusters with the min distance distant

	1	3	6	7	11	12
1	0					
3	2	0				
6	5	3	0			
7	6	4	1	0		
11	10	8	5	4	0	
12	11	9	6	5	1	0

we find {6} and {7} has the min distant 1

{11} and {12} has the min distant 1

So we randomly choose {6} and {7} and merge them to cluster {6,7}

③ Iteration 2.

	1	3	6, 7	11	12
1	0				
3	2	0			
6,7	5	3	0		
11	10	8	4	0	
12	11	9	5	1	0

$\{11\}\ \{12\} \rightarrow \{11, 12\}$

④ Iteration 3

	1	3	6, 7	11, 12
1	0			
3	2	0		
6,7	5	3	0	
11,12	10	8	4	0

$\{1\}\ \{3\} \rightarrow \{1, 3\}$

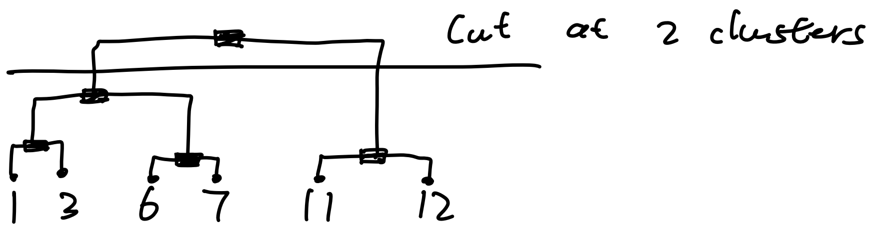
⑤ Iteration 4

	1, 3	6, 7	11, 12
1,3	0		
6,7	3	0	
11,12	8	4	0

$\{1,3\}, \{6,7\} \rightarrow \{1,3,6,7\}$

⑥ Stopping: only 2 clusters are left

$C_1 = \{1,3,6,7\}\ C_2 = \{11, 12\}$



(ii)

① Initialization: Each object is a cluster

② Iteration 1: Merge two clusters with the min distance distant

	1	3	6	7	11	12
1	0					
3	2	0				
6	5	3	0			
7	6	4	1	0		
11	10	8	5	4	0	
12	11	9	6	5	1	0

we find {6} and {7} has the min distant 1

{11} and {12} has the min distant 1

So we randomly choose {6} and {7} and merge them to cluster {6,7}

③ Iteration 2.

	1	3	6, 7	11	12
1	0				
3	2	0			
6,7	6	4	0		
11	10	8	5	0	
12	11	9	6	1	0

{11} {12} → {11,12}

④ Iteration 3

	1	3	6,7	11,12
1	0			
3	2	0		
6,7	6	4	0	
11,12	11	9	6	0

{1} {3} → {1,3}

⑤ Iteration 4

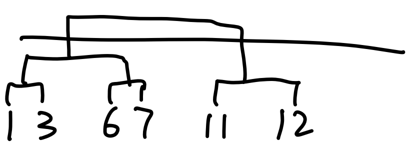
	1,3	6,7	11,12
1,3	0		
6,7	6	0	
11,12	11	6	0

{1,3} {6,7} → {1,3,6,7}

{6,7} {11,12} → {6,7,11,12}

random choose the first one

⑥ Stop : only two clusters left C_1={1,3,6,7} C_2={11,12}



(b) ① concept : PCA is a simple and popular method and unsupervised Learning, to learn the best low dimensional subspace for data projection.

② Problem: PCA want to achieve two objective simultaneously.

Objective

(1) the mean square error (MSE) of the projected data is minimized.

(2) The variance of the projected data is max

③ procedure: Apply SVD to the data matrix, Select the unit weight vector V_k from the V matrix, and calculate $(V_k^T x_i)V_k$

④ Reason: The variance-max criterion is equivalent to minimising the reconstruction error under orthogonal projection.

(c) ① study = 1 free = 0 money = 1

② $p(spam \mid study = 1, free = 0, money = 1)$

$$= \frac{p(study = 1 \mid spam)\, p(free = 0 \mid spam)\, p(money = 1 \mid spam)\, p(spam)}{p(study = 1, free = 0, money = 1)}$$

$$= \frac{0 \times 1 \times \frac{4}{8} \times \frac{8}{12}}{\frac{0}{12}}$$

$= 0$

Without smoothing the spam-class likelihood is zero, because the classifier havent trained by {study = 1, free = 0, money = 1} sample

(ii) ① we defind

S = purchase a premium subscription

H = spent more than 25 hours in last week

$P(S) = 0.2$

$P(H|S) = 0.6$

$P(H|\neg S) = 0.3$

$P(S|H) = ?$

② $P(H) = P(H|S) \times P(S) + P(H|\neg S)P(\neg S)$

$= 0.6 \times 0.2 + 0.3 \times (1 - 0.2)$

$= 0.12 + 0.24$

$= 0.36$

③ $P(S|H) = \dfrac{P(H|S) \times P(S)}{P(H)}$

$= \dfrac{0.6 \times 0.2}{0.36}$

$= 0.3333$