22-S1-Q4

Q: (a) HAC → 2 average

(b) PCA : PC1 PC2

(c) (i) NB + classify

(ii) P

Solution: (a) HAC

① Initialize: Each point as an individual cluster

② Iteration 1: calculate the distant between clusters

	0	4	5	20	25	39	43	44
0	0							
4	4	0						
5	5	1	0					
20	20	16	15	0				
25	25	21	20	5	0			
39	39	35	34	19	14	0		
43	43	39	38	23	18	4	0	
44	44	40	39	24	19	5	1	0

merge the two clusters withe the mimimum average linkage

we merge {4} and {5} into cluster {4,5} with distant 1

and denote it as $C_4, C_5 \longrightarrow C_{4,5}$

③ Iteration 2

$d_{0,45} = \frac{1}{2} \times (4+5) = 4.5$

$d_{20,45} = \frac{1}{2} \times (16+15) = 15.5$

and so on

	0	4, 5	20	25	39	43	44
0	0						
4,5	4.5	0					
20	20	15.5	0				
25	25	20.5	5	0			
39	39	34.5	19	14	0		
43	43	38.5	23	18	4	0	
44	44	39.5	24	19	5	1	0

$C_{43}, C_{44} \rightarrow C_{43,44}$

④ Iteration 3

	0	4, 5	20	25	39	43, 44
0	0					
4,5	4.5	0				
20	20	15.5	0			
25	25	20.5	5	0		
39	39	34.5	19	14	0	
43,44	43.5	39	23.5	18.5	4.5	0

$C_{39}, C_{43,44} \rightarrow C_{39,43,44}$

⑤ I4

	0	4, 5	20	25	39, 43, 44
0	0				
4,5	4.5	0			
20	20	15.5	0		
25	25	20.5	5	0	
39, 43, 44	42	37.5	22	17	0

$C_0, C_{4,5} \rightarrow C_{0,4,5}$

⑥ I5

	0, 4, 5	20	25	39, 43, 44
0, 4, 5	0			
20	17	0		
25	22	5	0	
39, 43, 44	39	22	17	0

C_{20} C_{25} $\rightarrow$ $C_{20,25}$

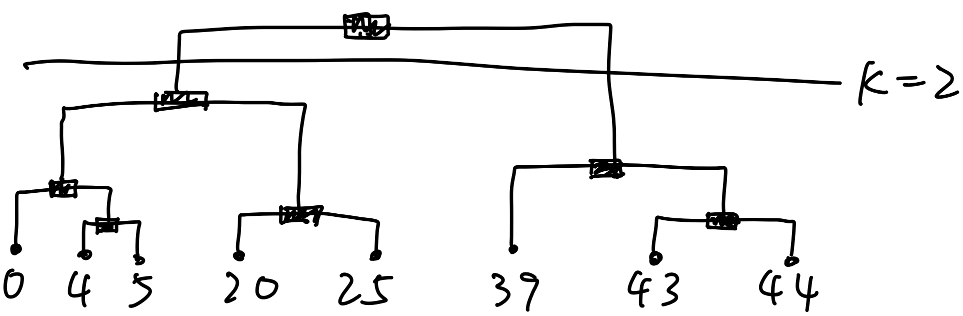
⑦ 26

	0, 4, 5	20 25	39, 43, 44
0, 4, 5	0		
20, 25	19.5	0	
39, 43, 44	39	19.5	0

$C_{0,4,5}$ $C_{20,25}$ $\rightarrow$ $C_{0,4,5,20,25}$

⑧ Stopping : only 2 clusters left

$C_{0,4,5,20,25}$ $C_{39,43,44}$



(b)

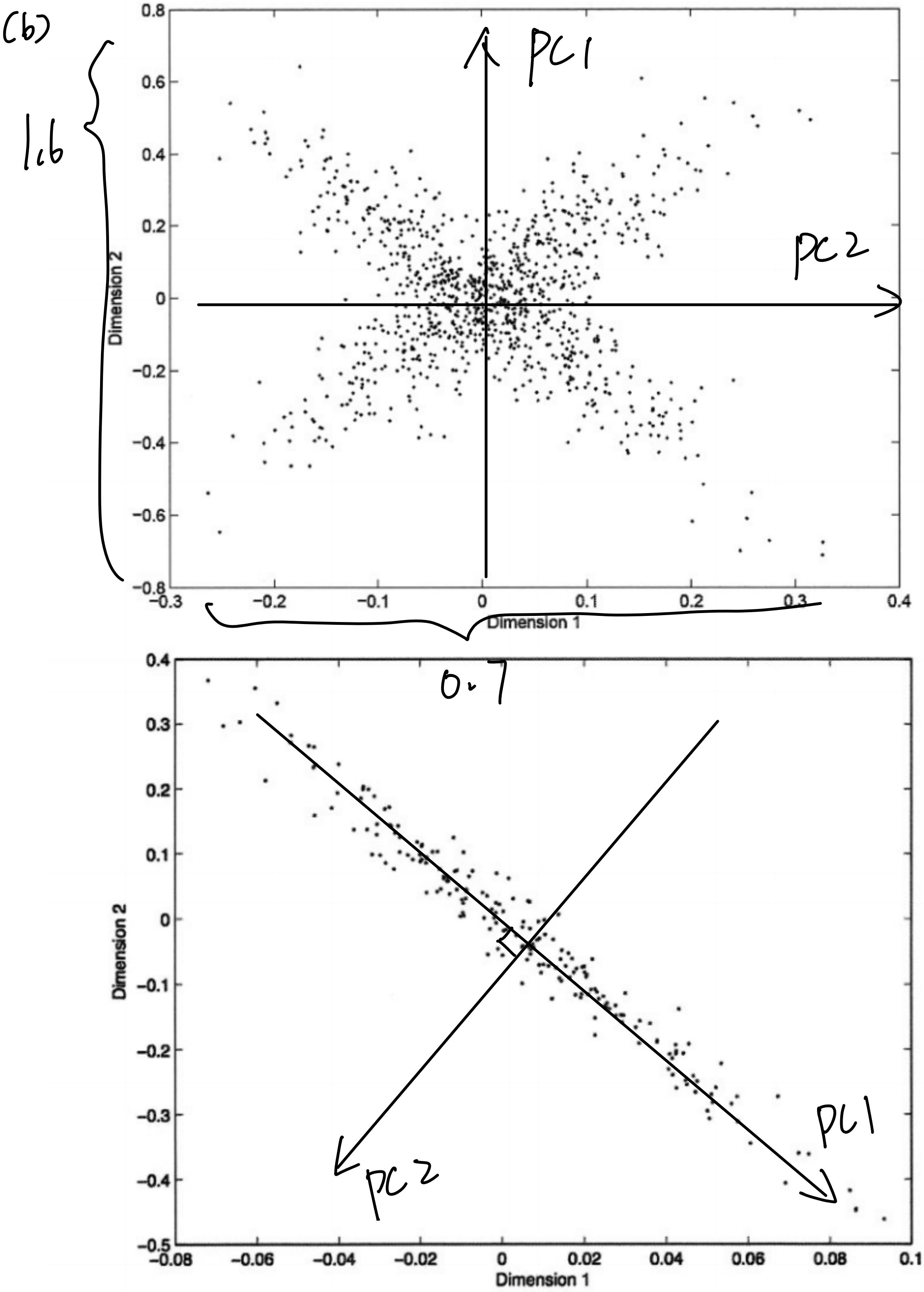
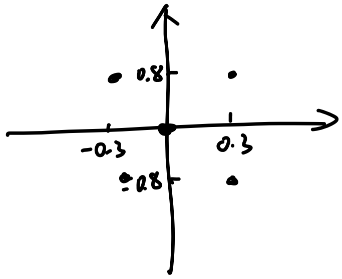


Figure 5. Plots of the two datasets of 2D points.

① we can calculate a simple sample to find PC1&2



$\mu=(0,0)$

center the data

covariance matricx Σ

$$\sigma_{11}^2 = \frac{1}{5} \times \left[(-0.3)^2+(-0.3)^2+0^2+0.3^2+0.3^2\right] = 0.072$$

$$\sigma_{22}^2 = \frac{1}{5} \times \left[(0.8)^2+(0.8)^2+0^2+0.8^2+0.8^2\right] = 0.512$$

$$\sigma_{12}^2 = \frac{1}{5} \times \left[(-0.3\times 0.8)+[-0.3\times(-0.8)]+0.3\times 0.8+0.3\times(-0.8)\right]$$

$$= 0$$

$$\Sigma = \begin{bmatrix} 0.072 & 0 \\ 0 & 0.512 \end{bmatrix}$$

$$|\lambda I - \Sigma| = (\lambda - 0.072)(\lambda - 0.512) = 0$$

$\lambda_1 = 0.512$ ✓ $\qquad \lambda_2 = 0.072$

$\Sigma v_1 = \lambda_1 v_1 \quad$ let $v_1 = \begin{pmatrix} a \\ b \end{pmatrix}$

$$\begin{bmatrix} 0.072 & 0 \\ 0 & 0.512 \end{bmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{bmatrix} 0.072\,a \\ 0.512\,b \end{bmatrix} = 0.512\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 0.512a \\ 0.512b \end{pmatrix}$$

$$\Rightarrow \begin{cases} a = 0 \\ b = \forall b \end{cases} \Rightarrow v_1 = \begin{pmatrix} 0 \\ a \end{pmatrix} \Rightarrow v_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

similary $\Rightarrow v_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

(c)(i) Naive Bayes

① we denote

$X = \{Sound = Bark, Fur = Coarse, Color = Brown\}$

$SB = \{Sound = Bark\}$

$FC = \{Fur = Coarse\}$

$CB = \{Color = Brown\}$

$D = \{Class = Dog\}$

$C = \{Class = Cat\}$

② $$P(D|X) = \frac{P(X|D)\,P(D)}{P(X)}$$

$$= \frac{P(SB|D)\,P(FC|D)\,P(CB|D)\,P(D)}{P(X)}$$

$$P(C|X) = \frac{P(X|C)\,P(C)}{P(X)}$$

$$= \frac{P(SB|C)\,P(FC|C)\,P(CB|C)\,P(C)}{P(X)}$$

ignore $P(X)$

we defing $g(D|X) = P(SB|D)\,P(FC|D)\,P(CB|D)\,P(D)$

$g(C|X) = p(SB|C)\, p(FC|C)\, p(CB|C)\, p(C)$

③ $p(SB|D) = \frac{3}{4}$ $\quad p(SB|C) = \frac{1}{4}$

$p(FC|D) = \frac{3}{4}$ $\quad p(FC|C) = \frac{1}{4}$

$p(CB|D) = \frac{1}{2}$ $\quad p(CB|D) = \frac{1}{2}$

$p(D) = \frac{4}{8} = \frac{1}{2}$ $\quad p(C) = \frac{1}{2}$

④ $g(D|X) = \frac{3}{4} \times \frac{3}{4} \times \frac{1}{2} \times \frac{1}{2} = 0.140625$

$g(C|X) = \frac{1}{4} \times \frac{1}{4} \times \frac{1}{2} \times \frac{1}{2} = 0.015625$

⑤ $g(D|X) > g(C|X)$

So we classify it to Dog.

(ii) we define $q = 1-p$

A = First head occures on toss 2, 4, 6 ...

$P(A)$

$$= qp + q^3p + q^5p + \cdots q^{2k+1}p$$

$$= \sum_{k=0}^{\infty} q^{2k+1} p$$

$$= qp \sum_{k=0}^{\infty} q^{2k}$$

$$= qp \sum_{k=0}^{\infty} (q^2)^k$$

$$= qp \frac{1}{1-q^2}$$

$$= \frac{(1-p)p}{1-(1-p)^2}$$

$$= \frac{(1-p)p}{1-(1-2p+p^2)}$$

$$= \frac{(1-p)p}{2p-p^2}$$

$$= \frac{(1-p)p}{p(2-p)}$$

$$= \frac{1-p}{2-p}, \quad p \in (0,1]$$

$$\frac{1}{1-公比}$$

$$\sum_{k=0}^{n} q = \frac{a_1(1-q^n)}{1-q}$$

If $p=0$, $P(A)=0$

All in all

$$P(A) = \begin{cases} \frac{1-p}{2-p} & , p \in (0,1] \\ 0 & , p=0 \end{cases}$$