

22-S1-Q4

Q: (a) HAC \rightarrow 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 distance 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 with the minimum average linkage

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

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

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 74

	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 75

	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

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

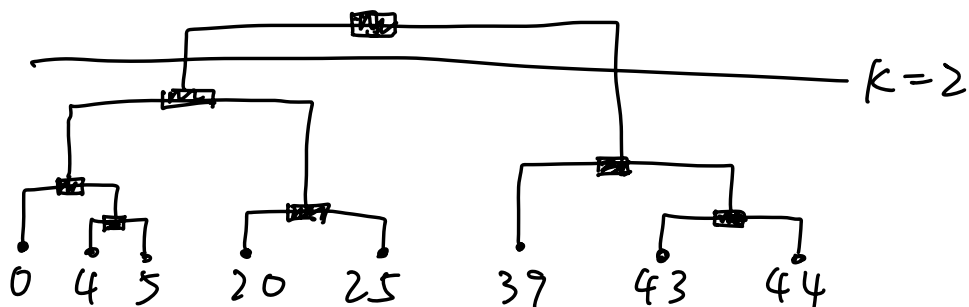
⑦ 26

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

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

⑧ Stopping : only 2 clusters left

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



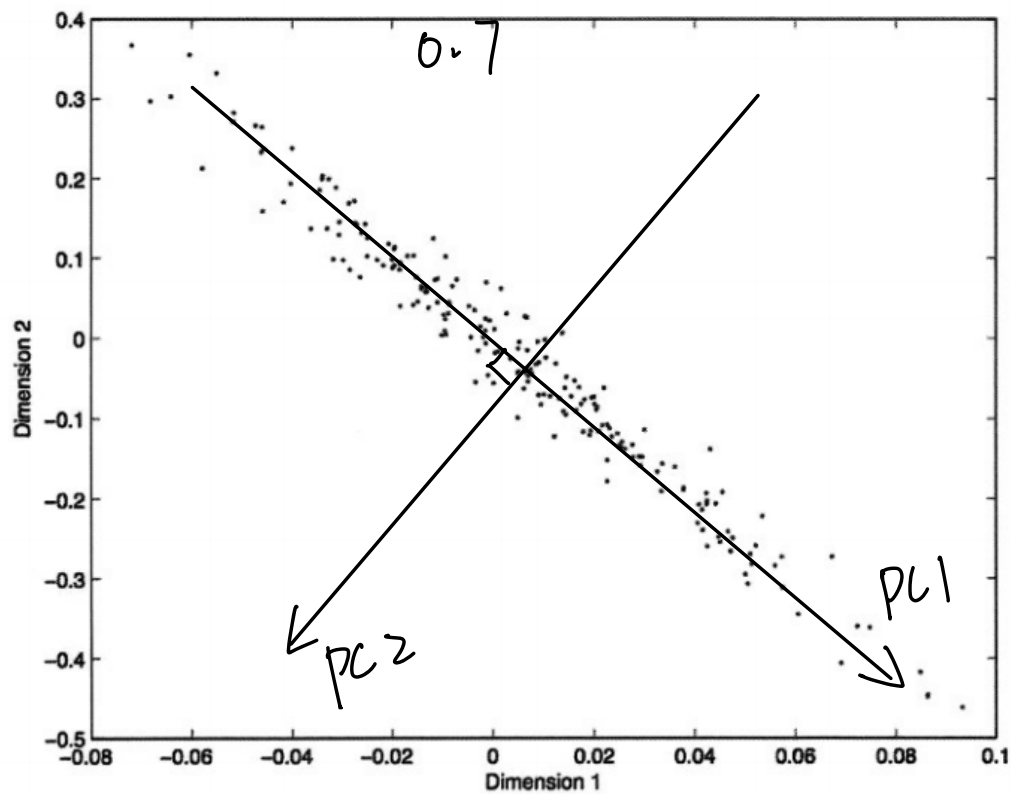
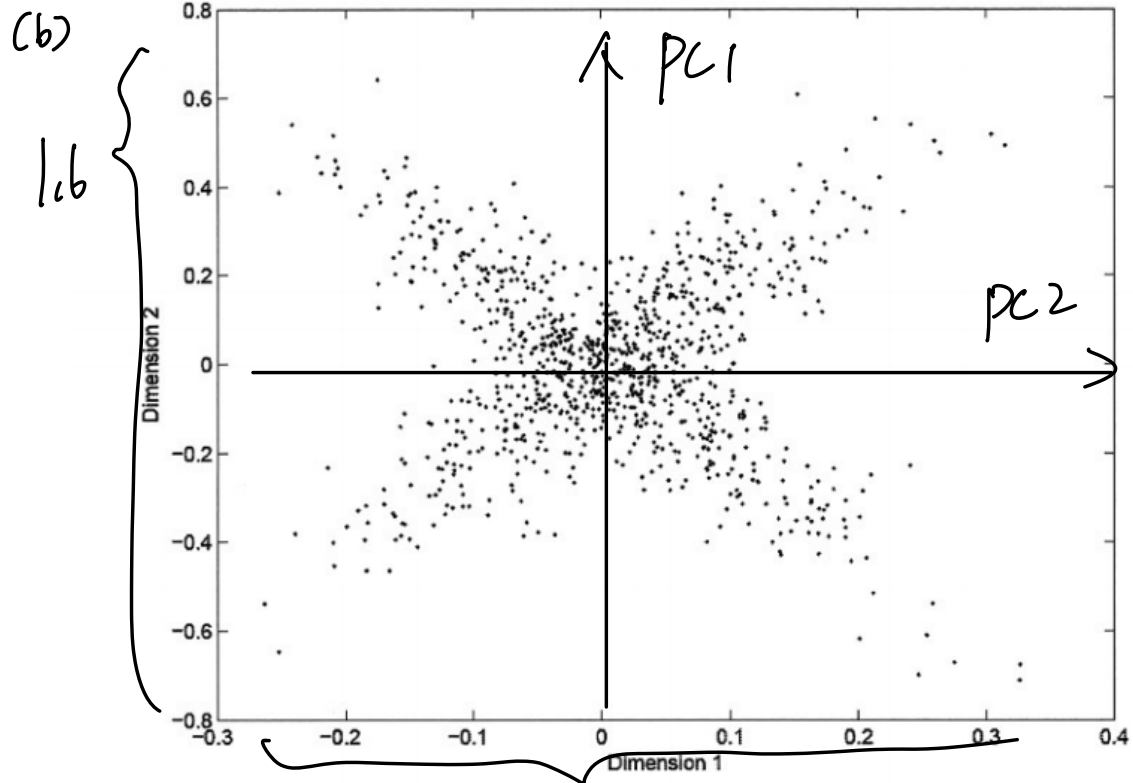
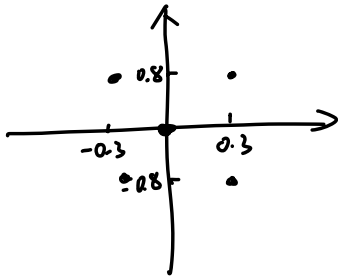


Figure 5. Plots of the two datasets of 2D points.

① we can calculate a sample to find PC1 & 2



$$\mu = (0, 0)$$

center the data

covariance matrix Σ

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

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

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

$$= 0$$

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

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

$$\lambda_1 = 0.512 \quad \checkmark \quad \lambda_2 = 0.072$$

$$\Sigma v_i = \lambda_i v_i \quad \text{let } v_i = \begin{pmatrix} a \\ b \end{pmatrix}$$

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

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

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

CC)(i) Naive Bayes

① we denote

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

$$SB = \{ \text{Sound} = \text{Bark} \}$$

$$FC = \{ \text{Fur} = \text{Coarse} \}$$

$$CB = \{ \text{Color} = \text{Brown} \}$$

$$D = \{ \text{Class} = \text{Dog} \}$$

$$C = \{ \text{Class} = \text{Cat} \}$$

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

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

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

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

ignore $P(X)$

we defining $g(D|X) = P(SB|D) P(FC|D) P(CB|D) P(D)$

$$g(C|x) = p(SB|C) p(FC|C) p(CB|C) p(C)$$

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

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

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

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

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

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

$$\textcircled{5} g(D|x) > g(C|x)$$

So we classify it to Dog.

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

$p(\text{First head occurs on toss } 2, 4, 6 \dots)$

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

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

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

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

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

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

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

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

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

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

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

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