22 - 51 - 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 $C_4, C_5 \rightarrow C_{4,5}$

③ Iteration 2
$$d_{0,45} = \tfrac{1}{2} \times (4+5) = 4.5$$
$$d_{20,45} = \tfrac{1}{2} \times (16+15) = 15.5$$

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

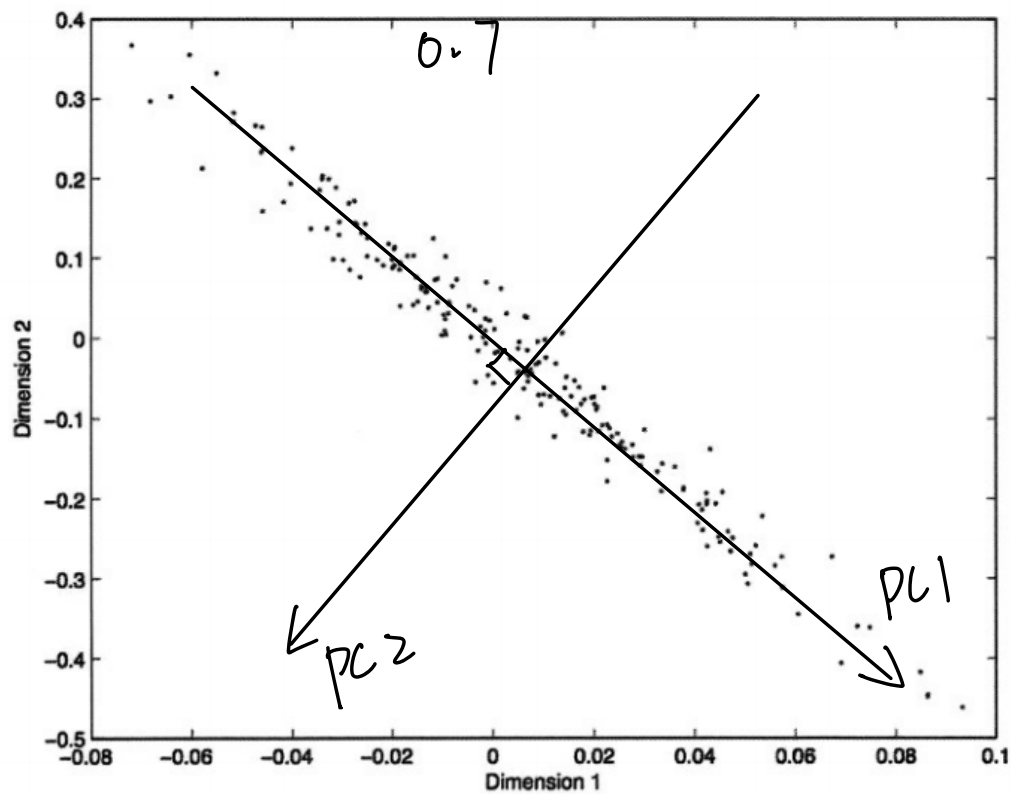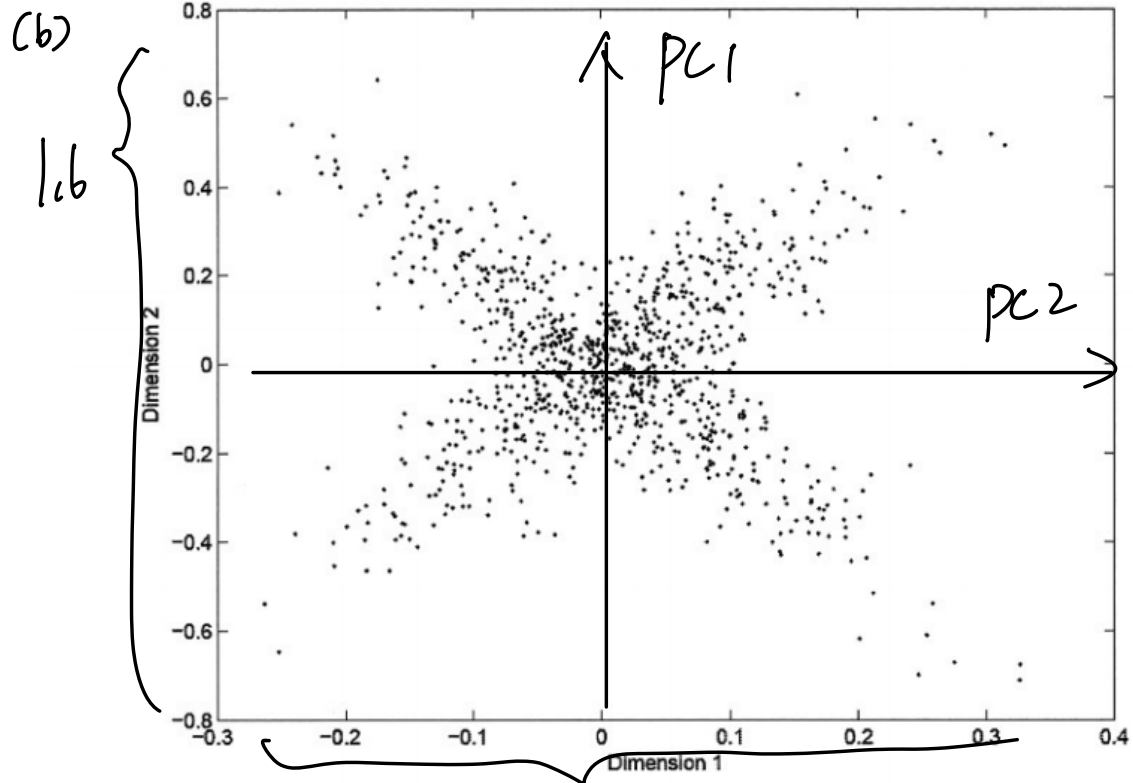$C_{0,4,5,20,25}$ $C_{39,43,44}$
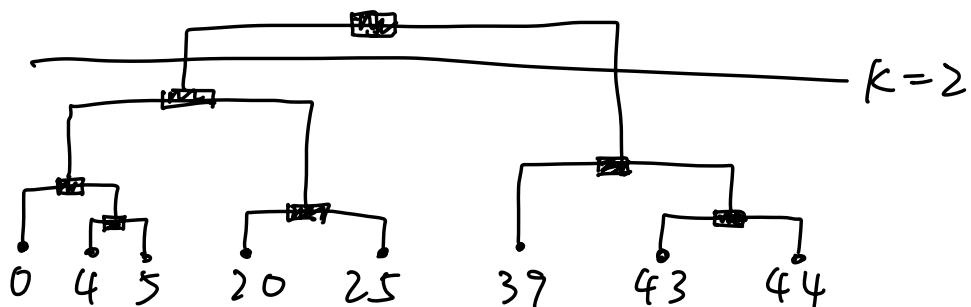
Figure 5. Plots of the two datasets of 2D points.

(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 $P(D|X) = \dfrac{P(X|D)\ P(D)}{P(X)}$

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

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

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

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

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

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

$p(D) = \dfrac{4}{8} = \dfrac{1}{2}$ $p(c) = \dfrac{1}{2}$

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

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

⑤ $g(D|X) > g(C|x)$

So we classify it to Dog.

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

P(First head occures on toss $2, 4, 6 \cdots$)

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

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

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

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

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

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

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

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

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

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

$$\frac{1}{1-\text{rate}}$$

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