BAYES NETWORK

Machine Learning Assignment – 6

Abstract

Perform WEKA related clustering and solves Bayes Network questions

SathyaNarayanan Srinivasan

Sxs142031@utdallas.edu

Table of Contents

Part - 1

i. **<u>Clustering without PCA</u>**

1. K-means with k=2: SSE =12598.193775711037

   Clustered Instances

   0 33 (87%) ALL(T&B)

   1 5 (13%)(AML)

2. **<u>Confusion matrix for KMeans k=2</u>**

		PREDICTED	
		AML	ALL-(T&B)
ACTUAL	AML	4	7
	ALL-(T&B)	1	26

3. **<u>K-means with k=3:</u>**

   SSE= 11186.928018360504

   Clustered Instances

   0 20 (53%)(ALL B-cell)

   1 3 (8%)(ALL-T-Cell)

   2 15 (39%)(AML)

4. **<u>Confusion matrix for KMeans k=3</u>**

		PREDICTED		
		AML	ALL-(T cell)	ALL-(B cell)
ACTUAL	AML	1	3	7
	ALL-(T cell)	3	0	5
	ALL-(B cell)	11	0	8

5. **<u>Hierarchical Clustering with K=2:</u>**

   Clustered Instances

   0 37 (97%)(ALL-T& B cell)

   1 1 (3%)(AML)

6. **Confusion Matrix for Hierarchical Clustering K=2**

		PREDICTED	
		AML	ALL-(T&B)
ACTUAL	AML	0	11
	ALL-(T&B)	1	26

7. **Hierarchical Clustering with K=3:**

Clustered Instances

0 36 (95%)(ALL B-cell)
1 1 (3%)(ALL T-cell)
2 1 (3%)(AML)

8. **Confusion matrix for Hierarchical Clustering K=3**

		PREDICTED		
		AML	ALL-(T cell)	ALL-(B cell)
ACTUAL	AML	0	0	11
	ALL-(T cell)	0	0	8
	ALL-(B cell)	1	1	17

ii. **Clustering after PCA**

9. **K-means with k=2:**

SSE = 56.3089363060435

Clustered Instances

0 28 (74%)(ALL T& B cell)
1 10 (26%)(AML)

10. **Confusion matrix for KMeans k=2**

		PREDICTED	
		AML	ALL-(T&B)
ACTUAL	AML	4	7
	ALL-(T&B)	6	20

11. **K-means with k=3:**

SSE= 55.22769565994147

Clustered Instances

0 20 (53%)(ALL B CELL)

1 7 (18%)(ALL T CELL)

2 11 (29%)(AML)

12. **Confusion matrix for KMeans k=3**

		PREDICTED		
		AML	ALL-(T cell)	ALL-(B cell)
ACTUAL	AML	3	2	6
	ALL-(T cell)	4	1	3
	ALL-(B cell)	4	4	11

13. **Hierarchical Clustering with K=2:**

Clustered Instances

0 37 (97%)(ALL-T& B cell)

1 1 (3%)(AML)

14. **Confusion Matrix for Hierarchical Clustering K=2**

		PREDICTED	
		AML	ALL-(T&B)
ACTUAL	AML	0	11
	ALL-(T&B)	1	26

15. **Hierarchical Clustering with K=3**:

Clustered Instances

0 36 (95%)(ALL B-cell)

1 1 (3%)(ALL T-cell)

2 1 (3%)(AML)

16. Confusion matrix for Hierarchical Clustering K=3

		PREDICTED		
		AML	ALL-(T cell)	ALL-(B cell)
ACTUAL	AML	1	0	10
	ALL-(T cell)	0	0	8
	ALL-(B cell)	0	1	18

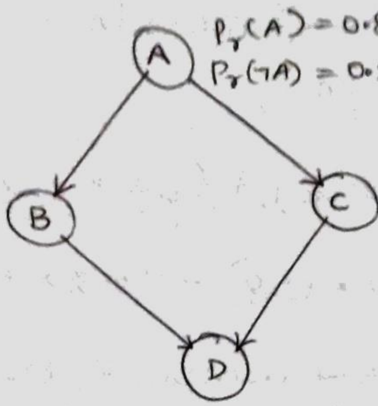
iii. Classification

Algorithm	Parameters	%Training accuracy
Random Forest	Max depth, No.of features, number of trees, seeds	100
Boosting	Classifier, number of iterations, seed	5.2632
J48	Confidence factor, min no.of obj, number of folds, seed, subtree raising	44.7368
Bagging	Bag size percent, number of iterations,classifier ,seed	55.2632

Algorithm	Parameters	%Test Accuracy
Random Forest	Max depth, No.of features, number of trees, seeds	5.2632
Boosting	Classifier, number of iterations,seed	2.6316
J48	Confidence factor, min no.of obj, number of folds, seed, subtree raising	0
Bagging	Bag size percent, number of iterations,classifier ,seed	2.6316

Part – 2

(1)



$Pr(A) = 0.8$
$Pr(\neg A) = 0.2$

$Pr(B|A) = 0.4$
$Pr(B|\neg A) = 0.9$
$Pr(\neg B|A) = 0.6$
$Pr(\neg B|\neg A) = 0.1$

$Pr(C|A) = 0.6$ $\quad Pr(\neg C|A) = 0.4$
$Pr(C|\neg A) = 0.2$ $\quad Pr(\neg C|\neg A) = 0.8$

$Pr(D|B,C) = 0.8$
$Pr(D|B,\neg C) = 0.5$
$Pr(D|\neg B,C) = 0.6$
$Pr(D|\neg B,\neg C) = 0.3$

$$
\begin{aligned}
Pr(C \wedge D) &= Pr(C=1, D=1) \\
&= Pr(A=a, B=b, C=1, D=1) \\
&= \sum_{a,b} Pr(A=a)\, P(B=b \mid A=a) P(C=1 \mid A=a) \\
&\qquad P(D=1 \mid B=b, C=1) \\
&= \{ P(A=1) \cdot P(B=1 \mid A=1)\, P(C=1 \mid A=1) \\
&\qquad P(D=1 \mid B=1, C=1) \} + \\
&\quad \{ P(A=1) \cdot P(B=0 \mid A=1) \cdot P(C=1 \mid A=1) \cdot \\
&\qquad P(D=1 \mid B=0, C=1) \} + \\
&\quad \{ P(A=0)\, P(B=1 \mid A=0) \cdot P(C=1 \mid A=0) \cdot \\
&\qquad P(D=1 \mid B=1, C=1) \} +
\end{aligned}
$$

Page 1

$$\{P(A=0) \cdot P(B=0 \mid A=0)\ P(C=1 \mid A=0)$$
$$P(D=1 \mid B=0, C=1)\}$$

$$= \{(0.8 \times 0.4 \times 0.6 \times 0.8) + (0.8 \times 0.6 \times 0.6 \times 0.6) +$$
$$(0.2 \times 0.9 * 0.2 \times 0.8) + (0.2 \times 0.1 \times 0.2 \times 0.6)\}$$

$$= 0.1536 + 0.1728 + 0.0288 + 0.0024$$

$$= 0.3576$$

(2) $P(HIV) = 0.05\% = 0.0005,$

$P(\neg HIV) = 0.995$

$P(+ \mid HIV) = 0.98$

$P(- \mid HIV) = 0.02$

$P(+ \mid \neg HIV) = 0.03$

$P(- \mid \neg HIV) = 0.97$

$P(HIV \mid +) = ?$

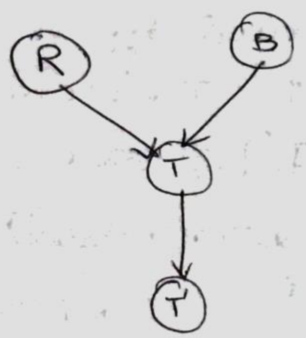
$$P(HIV \mid +) = \frac{P(+ \mid HIV) \cdot P(HIV)}{P(+)}$$

②

$$P(+) = P(+|HIV) \cdot P(HIV) + P(+|\neg HIV) \cdot P(\neg HIV)$$
$$= (0.98 \times 0.0005) + (0.03 \times 0.995)$$
$$= 0.00049 + 0.029985$$
$$= 0.030475$$

$$P(HIV|+) = \frac{0.98 \times 0.0005}{0.030475}$$
$$= \frac{0.00049}{0.030475}$$
$$= 0.016 \text{ (ANS)}$$

③



(i) $R \perp\!\!\!\perp B$

$$P(R,B) = P(R=1, B=1, T=t, T'=t')$$
$$= \sum P(R)P(B) P(T|R,B) P(T'|T)$$

As $P(R,B)$ cannot be expressed as R & B alone, they are CONDITIONALLY INDEPENDENT

(2)

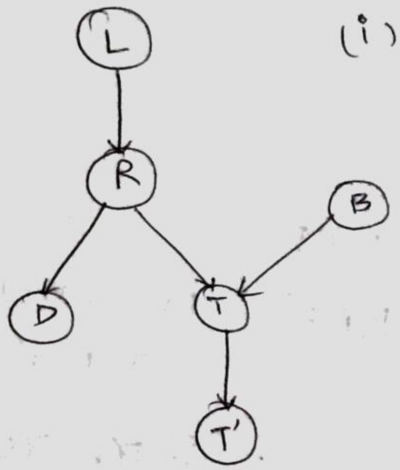
(ii) $R \perp\!\!\!\perp B \mid T$

Path from $R \rightarrow B$ is NOT BLOCKED as node $\rightarrow T \leftarrow$ descends into conditioning node

Thus R & B are not d-separated by T'

Thus R&B are not conditionally independent given T'

(4)



(i) $L \perp\!\!\!\perp T' \mid T$

$L \rightarrow T'$ is BLOCKED.

Thus L & T' are d-separated by T & $L \perp\!\!\!\perp T' \mid T$. Hence; $L \perp\!\!\!\perp T' \mid T$ are CONDITIONALLY INDEPENDENT given T

(ii) $L \perp\!\!\!\perp B$

$$P(L,B) = P(l=1, B=1)$$
$$= \sum P(L,R,D,T,B,T')$$
$$= \sum P(L)\,P(R|L)\,P(D|R)\,P(B)\,P(T|R,B)\,P(T'|T)$$

They can be separated into terms of L & B. Thus, conditionally independant.

iii) $L \perp\!\!\!\perp B \mid T$

$L \rightarrow B$ is not BLOCKED as node $\rightarrow (T) \leftarrow$ is a conditional node. Thus, L & B are not conditionally independant given T.

iv) $L \perp\!\!\!\perp B \mid T'$

Path $L \rightarrow B$ is NOT BLOCKED. Thus L & B are not d-separated by T'. Thus L & B are not conditionally independant given T'

v) $L \perp\!\!\!\perp B \mid T, R$

Path from $L \rightarrow B$ is blocked. Thus, L & B are conditionally independent given T, R

5) A) a) $2 \times 20 = \boxed{40}$

b) $20 \times 5 = \boxed{100}$

c) $500 \times 10 = \boxed{5000}$

B) a) $2 + 2^2 + (2^3 \times 2) + 2^4 \times 16$

$= \boxed{278}$

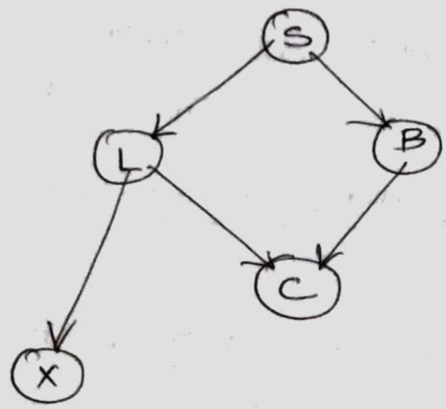
b) $5 + 5^2 + (5^3 \times 2) + 5^4 \times 16$

$= \boxed{10280}$

c) $10+10^2+(10^3\times 2)+(10^4\times 496)$

$= 4962110$

(6)



(i) Given S, pairs that are conditionally independant are:

A) L & B
B) B & X

(ii) Given L, pairs that are conditionally independent are:

A) S & X
B) B & X
C) C & X

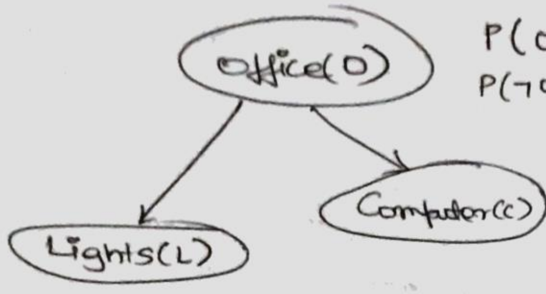
iii) Given {L, B} pairs that are conditionally independent are:

A) S & C
B) S & X
C) C & X

Page 6

(4)

(7)



$P(O) = 0.6$

$P(\neg O) = 0.4$

$P(C|O) = 0.8$

$P(C|\neg O) = 0.1$

$P(\neg C|O) = 0.2$

$P(\neg C|\neg O) = 0.9$

$P(L|O) = 0.5$

$P(L|\neg O) = 0.05$

$P(\neg L|O) = 0.5$

$P(\neg L|\neg O) = 0.95$

$$P(L=1|C=1) = \frac{P(L=1, C=1)}{P(C=1)}$$

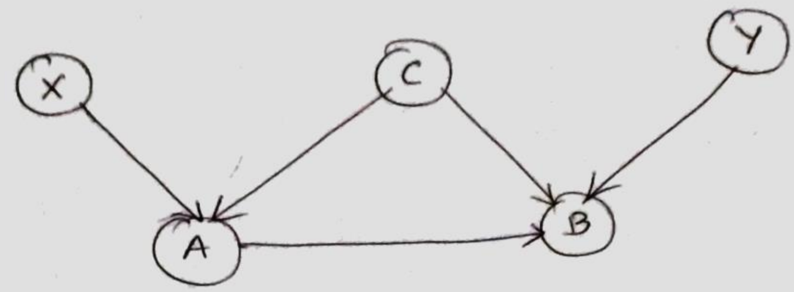
$$= \frac{\sum_{o} P(O=o, L=1, C=1)}{\sum_{O,L} P(O=o, L=l, C=1)}$$

$$= \frac{\sum_{o} P(O=o) \times P(L=1|O=o) \times P(C=1|O=o)}{\sum_{O,L} P(O=o)\, P(L=l|O=o)\, P(C=1|O=o)}$$

$$= \frac{(0.6 \times 0.5 \times 0.8) + (0.4 \times 0.05 \times 0.1)}{\left\{ (0.4 \times 0.95 \times 0.1) + (0.4 \times 0.05 \times 0.1) + (0.6 \times 0.5 \times 0.8) + (0.6 \times 0.5 \times 0.8) \right\}}$$

$= \boxed{0.465 \text{ (ANS)}}$

(8)



Possible sets of nodes of A, B, C are:
{A} {B} {C} {A, C} {B, C} {C, A} {A, B, C}
& {∅}.

{A} → For $X \perp\!\!\!\perp Y \mid A$, paths are X – A – C – B – Y and X – A – B – Y & both paths are blocked at A & B
Thus, {A} D-SEPARATES X & Y

{C} → For $X \perp\!\!\!\perp Y \mid C$ paths are X – A – C – B – Y & X – A – B – Y & both paths are blocked at C & B. Thus, {C} D-SEPARATES X & Y

(5)

$\{A, C\}$ → For $X \perp\!\!\!\perp Y \mid A, C$ paths are

X - A - C - B - Y & X - A - B - Y & both paths

are blocked at C & A. Thus $\{A, C\}$

D-SEPARATES X & Y

$\{(A, B, C)\}$ → For $X \perp\!\!\!\perp Y \mid A, B, C$ paths are

X - A - C - B - Y & X - A - B - Y & both paths

are blocked at A & C. Thus $\{A, B, C\}$

d-separates X & Y.

Page 9