מספר התלמיד הנבחן רשום את כל תשע הספרות

האוניברסיטה

ט"ז בתמוז תשע"ד

14

ביולי 2014

סמסטר 2014ב

20595 / 4

מס' שאלון - 493

86 מס' מועד

שאלון בחינת גמר

20595 - כריית מידע

הדבק כאן את מדבקת הנבחן

משך בחינה: שעות

> בשאלון זה 5 עמודים

מבנה הבחינה:

בבחינה שני חלקים.

יש לענות על כל השאלות.

:חלק א

ענו על ארבע השאלות שבחלק זה.

90 נקודות לחלק זה.

חלק ב:

ענו על השאלה שבחלק זה.

10 נקודות לחלק זה.

משקל כל שאלה מפורט בגוף השאלון.

אפשר להשתמש בהנחות במהלך הפתרון, אם תזדקקו לכך.

פרטו את הנחותיכם!

חומר עזר:

כל חומר עזר מותר בשימוש.

השימוש במחשב נישא או במכשיר כלשהו שבאמצעותו אפשר להתחבר

לאינטרנט אסור.

בהצלחה !!!

החזירו

למשגיח את השאלון

וכל עזר אחר שקיבלתם בתוך מחברת התשובות



חלק א (90 נקודות)

בחלק זה ארבע שאלות. ענו במפורט על ארבע השאלות.

(decision trees) עצי החלטה – עצי (מקודות) – עצי שאלה 1

: נתונה טבלת נתוני אימון

היקף השריפה	הצלחה בכיבוי	רוח	מזג אוויר	מועד שריפה	מס׳ קריאה
גדול מאוד	לא	לא	נאה	בוקר	1
גדול מאוד	כן	לא	מעונן	בוקר	2
קטן	לא	לא	נאה	ערב	3
גדול מאוד	לא	כן	גשום	צהריים	4
גדול מאוד	כן	כן	נאה	צהריים	5
קטן	י לא	לא	נאה	ערב	6
קטן	לא	לא	מעונן	ערב	7
	לא	לא	·		8
בינוני			גשום	צהריים	
גדול מאוד	לא	כן	נאה	צהריים	9
גדול מאוד	כן	כן	מעונן	ערב	10
קטן	לא	לא	גשום	ערב	11
גדול מאוד	לא	כן	מעונן	בוקר	12
בינוני	לא	לא	נאה	צהריים	13
גדול מאוד	כן	כן	גשום	צהריים	14

א. בנו עץ החלטה חלקי, הכולל את רמת השורש ורמה אחת נוספת בלבד, לנתוני האימון שבטבלה לחיזוי החלטה אם הצליחו לכבות את השריפה. הדגימו בתשובתכם את שלבי בחירת התכונה המפצלת בעץ.

הערה: יש לכלול חישוב של אחד המדדים, כדוגמת אנטרופיה, Gain ratio, מדד גיני.

ב האם אפשר לגזום את עץ ההחלטה שבניתם בסעיף אי נמקו תשובתכם.

(association rules) שאלה 2 (15 נקודות) – חוקי הקשר

נתונה הטבלה:

Items	TransID		
A, B, C, D	T100		
A, B, C, E	T200		
A, B, E, F, H	T300		
A, C, H	T400		

: הניחו

Min_support=50%

Min_confidence=60%

חשבו את מדד הענין (lift) עבור כל הקבוצות התדירות.

שאלה 3 (clustering) - ניתוח אשכולות (clustering)

בטבלה שלפניכם נתונים ארבעה סוגי תרופות. לכל תרופה שתי תכונות (weight - I pH-Index).

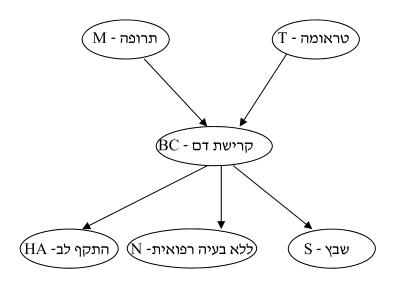
Medicine	Weight	pH-Index
A	1	1
В	2	1
C	4	3
D	5	4

בצעו אשכול לנתונים אלו תוך שימוש באלגוריתם k-means במטריקה של מרחק אוקלידי.

 $c_1=A, c_2=B:$ תנאי התחלה

שאלה 4 (25 נקודות) – למידה בייסיאנית

באיור שלפניכם נתונה רשת בייסיאנית:



.FALSE מציין TRUE להלן טבלאות ההסתברויות, כאשר

P(M=T)	0.2
P(M=F)	0.8
P(T=T)	0.05
P(T=F)	0.95

M	T	P(BC=T)	P(BC=F)	
T	T	0.95	0.05	
Т	F	0.3	0.7	
F	T	0.6	0.4	
F	F	0.9	0.1	

BC	P(HA=T)	P(HA=F)	P(S=T)	P(S=F)	P(N=T)	P(N=F)
T	0.4	0.6	0.35	0.65	0.25	0.75
F	0.15	0.85	0.1	0.9	0.75	0.25

מהי ההסתברות שנבדק יפתח בעיית קרישת דם כתוצאה מתרופה ומטראומה, אך ללא כל בעיה רפואית!

המשך הבחינה בעמוד הבא

חלק ב (10 נקודות)

בחלק זה שאלה אחת. ענו עליה במפורט.

שאלה 5 (10 נקודות)

שרטטו נתונים בדו-ממד בעלי שני אשכולות, הניתנים לזיהוי ברמת דיוק של 100% בעזרת שרטטו נתונים בדו-ממד בעלי שני אשכולות, הניתנים k-means אלגוריתם

בהצלחה!