



תאריך הבחינה :	11/07/2011
שם המרצה :	פרופ' מרק לסט
שם הקורס :	כריית נתונים ומחשני נתונים
מספר הקורס :	372-1-3105
שנה :	2011
סמסטר :	ב' מועד : א'
משך הבחינה :	3 שעות
חומר עזר :	דף נוסחאות (מצורף לבחינה) + מחשבון
גרסה מס' :	1

חלק 1 [50 נקודות]

- יש לענות על כל השאלות
- משקל של כל שאלה – 5 נקודות
- יש לרשום את התשובה בכתב-יד ברור במקום המיועד לכך על-גבי שאלון הבחינה בלבד
- תשובה לא מנומקת (במידה ונדרשת הנמקה) תקבל ציון של אפס

א. יש להסביר בקצרה מהי תופעת Overfitting.

ב. נכון/לא נכון : אם כל תתי הקבוצות האפשריות של קבוצת פריטים (itemset) הן שכיחות (frequent), אז גם ה-itemset עצמו הוא frequent. יש לנמק בקצרה את תשובתכם.

ג. נכון/לא נכון : ב-Association Rules, אם נעביר פריט אחד מהצד הימני של החוק לצד השמאלי של החוק, אז ה-confidence של החוק לעולם לא יהיה נמוך יותר. (דוגמא : מעבר מחוק $X \rightarrow Y, Z$ לחוק $X, Y \rightarrow Z$). יש לנמק בקצרה את תשובתכם.

ד. מהי ההנחה הבסיסית של Naïve Bayes Classifier?



ה. יש להציג את שלוש התכונות של אנתרופיה.

1. _____

2. _____

3. _____