

## מודלים במסחר אלקטרוני (096211)

בחן אמצע מועד א' תשס"ה 18/5/2005

המרצה: פרופ' משה טננהולץ. המתרגל: מר אלון אלטמן.

משך הבחינה: שעתיים וחצי.

בבחינה זו 2 עמודים. אנא ודאו שקיבלתם את כולם. מותר להשתמש בכל חומר עזר. יש להגיש את מחברת הבחינה בלבד.

### שאלה 1

28) שאלה זו עוסקת בדרוג אתרים באינטרנט. הסעיפים אינם תלויים זה בזה.

א. (14 נק') תנו דוגמא לרשת (גרף קישורים) עבורה אלגוריתמי PageRank ו-Hubs and Authorities יביאו לדרוג יחסי שונה של האתרים. עבור PageRank הניחו  $d = 85\%$  (כלומר 15% סיכוי לעבור לאתר אקראי). במקרה של Hubs and Authorities התייחסו לדרוג ה-Authority בלבד. רמז: קיימת דוגמא בעלת 4 צמתים, אחד עם hubbiness גבוה שמצביע (בין השאר) על אחד עם authority גבוה, וצומת אחר עם PageRank גבוה שמוצבע על ידי שני צמתים עם hubbiness נמוך.

ב. (14 נק') שיטת דרוג אתרים  $\mathcal{R}$  תיקרא ניתנת למניפולציה אם קיים גרף  $G = (V, E)$  ואתר  $v \in V$  כך ששינוי ההצבעות היוצאות של  $v$  יגרום לשינוי ב- $\mathcal{R}(v)$ . הוכיחו ש-PageRank ו-Hubs and Authorities הינן ניתנות למניפולציה.

### שאלה 2

28) שאלה זו עוסקת ב-Clustering.

נתונות הנקודות הבאות על ציר המספרים: 1, 2.1, 3, 6, 10, 12.

א. (9 נק') הפעילו את אלגוריתם Heirarchial Clustering על הנקודות לקבלת שני clusters. יש לתאר בפרוט את כל השלבים. מה ההיררכיה המתקבלת?

ב. (9 נק') הפעילו את אלגוריתם  $k$ -means על הנקודות כאשר הצנטרואידים ההתחלתיים הם 1 ו-12. יש לתאר בפרוט את כל השלבים. מה ה-clusters המתקבלים?

ג. (10 נק') אם ה-clusters שהתקבלו בסעיפים הקודמים זהים בין שני האלגוריתמים, הוסיפו נקודה לנתונים כך שיתקבלו clusters שונים. אחרת, הוסיפו נקודה לנתונים כך שיתקבלו clusters זהים בין שני האלגוריתמים. יש לציין מה ה-clusters שיתקבלו בכל אלגוריתם לאחר הוספת הנקודה, אך אין צורך לתאר את כל שלבי ההרצה.

### שאלה 3

(16 נק') שאלה זו עוסקת ב-VCG.

א. (8 נק') יש להפעיל את אלגוריתם VCG על ההערכות הבאות ולפרט את ההקצאה המתקבלת, התשלומים של כל לקוח, ואת הרווח של המוכר:

סל	לקוח 1	לקוח 2	לקוח 3
{A}	10	2	5
{B}	3	5	3
{A, B}	10	5	11

ב. (8 נק') האם ייתכן מצב בו במכרז VCG עם שני מוצרים ושני קונים (א' וב'), הוספת סוכן שלישי(ג') תשנה את הזוכה באחד מהמוצרים מא' לב' אם כן, יש לתת דוגמא. אם לא, יש להוכיח שזה לא ייתכן.

### שאלה 4

(28 נק') שאלה זו עוסקת ב-market-basket.

נתונים הסלים הבאים ומספר הלקוחות שקנו כל סל:

מספר לקוחות	סל	מספר לקוחות	סל
1,100	{A, B, C}	1,000	{C}
500	{A, D}	2,000	{D}
2,000	{B, E}	500	{E}
1,800	{B, D, E}	3,000	{A, B}
200	{A, C, D, E}	1,200	{A, C}
		500	{B, C}

שימו לב: ספירת הלקוחות של תת-סל לא כוללת את הלקוחות שקנו סלים שמכילים אותו. סלים שאינם מצויינים לעיל לא נקנו כלל.

מעוניינים להשתמש באלגוריתם PCY למציאת frequent pairs, כאשר זוג נפוץ הוא כזה שקנו אותו לפחות 4,000 לקוחות.

א. (10 נק') אילו מוצרים עומדים בתנאי A priori? יש לפרט את החישוב.

ב. (12 נק') נתונה פונקציית ערבול  $h$  המערבלת את הזוגות לדליים לפי הטבלה הבאה:

דלי 1	דלי 2	דלי 3	דלי 4	דלי 5
{A, B}	{A, C}	{A, D}	{A, E}	{C, D}
{C, E}	{B, D}	{B, E}	{B, C}	{D, E}

בשלב השני של האלגוריתם משתמשים בפונקציית הערבול  $h$  על הזוגות שנתרו מהפעלת A priori. אילו דליים ייחשבו נפוצים? יש לפרט את החישוב.

ג. (6 נק') בשלב השלישי סופרים את הזוגות מהסלים הנפוצים. מהם הזוגות הנפו-צים בסיים האלגוריתם?