

שאלה 1

שאלה זו עוסקת באלגוריתם multistage PCY למציאת frequent pairs.

א. הוצע, לשם יעילות, להשתמש באותה פונקציית hash בכל שלבי האלגוריתם. כיצד תשפיע פעולה זו על ביצועי האלגוריתם? נמקו את תשובתכם!

ב. לשם חסכון בזיכרון, הוצע שבכל שלב יישמר רק ה-bitmap מהשלב הקודם ולא יישמרו bitmaps אחרים ו/או רשימת ה-frequent items. כיצד תשפיע פעולה זו על ביצועי האלגוריתם? נמקו את תשובתכם!

ג. הוצע לשפר את האלגוריתם ולשמור בכל שלב את תוצאות הספירה עצמה של ה-hash table ולא רק bitmap. כיצד ישפיע שינוי זה על ביצועי האלגוריתם, אם אנו מניחים שכל הזיכרון הפנוי מנוצל לאחסון ה-hash table?

שאלה 2

נתונה טבלת רכישות מוצרים על ידי לקוחות ונתונות 4 פונקציות hash על מספרי הלקוחות:

f_4	f_3	f_2	f_1	4	3	2	1	לקוח/מוצר
5	7	4	1		*		*	1
7	6	1	2	*		*		2
3	3	6	3		*		*	3
2	4	7	4			*	*	4
1	2	3	5		*			5
4	1	5	6		*	*	*	6
6	5	2	7	*		*		7

א. חשבו את ה-Jaccard Measure בין זוגות המוצרים (1,2) ו-(2,3).

ב. חשבו את חתימות ה-Min hash לכל אחד מהמוצרים (לפי כל אחת מהפונקציות).

ג. חשבו את מידת הדמיון כפי שתתקבל מאלגוריתם Min hash בין כל זוג מוצרים.

ד. האם ייתכן מצב שבו מידת הדמיון שמתקבלת ב-Min hash תהיה קטנה יותר ממידת הדמיון בפועל? אם כן, תנו דוגמא. אם לא, יש להוכיח.

שאלה 3

(28 נק') 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 4

לכל אחד מהסעיפים הבאים, בחרו את האלגוריתם או השיטה היעילים ביותר לפתרון הבעיה המוצגת מבין האפשרויות להלן. נמקו בקצרה את תשובותיכם.
PageRank, Multistage PCY, Min Hash, Recommender Systems, k-means, ID3

- א. מציאת סלי המוצרים הנמכרים ביותר בחנות מקוונת.
- ב. הצעת הנחות על זוגות מוצרים למתעניינים במוצר מסויים.
- ג. מציאת האנשים המקושרים ביותר ברשת חברתית.
- ד. זיהוי של קבוצות שונות של לקוחות לפי מאפייני הרכישה שלהם ומאפיינים דמוגרפיים.
- ה. הצעת הנחות על מוצרים בהינתן הסטוריית הרכישות של הלקוח.
- ו. זיהוי לקוחות אשר עלולים לנטוש את החברה לפי נתוני נטישה של לקוחות אחרים ומאפיינים דמוגרפיים.

שאלה 5

נתונה הסיטואציה הבאה: לשני שחקנים יש 3 מטבעות כל אחד. הם צריכים להחליט בו זמנית אם לקחת מטבע אחד לשחקן האחר. אם אחד השחקנים מחליט לקחת מטבע, הוא מקבל אותו על חשבון השחקן האחר. אך, אם שני השחקנים מחליטים לקחת מטבע, מאבדים שניהם את כל המטבעות.

- א. תארו את הסיטואציה כמשחק.
- ב. מצאו את כל נקודות שיווי המשקל באסטרטגיות טהורות במשחק שהתקבל.
- ג. מצאו שיווי משקל בו שני השחקנים משחקים אסטרטגיות מעורבות (לא טהורות).
- ד. מה התוחלת של מספר המטבעות שבידי כל שחקן בשיווי המשקל הנ"ל?