

### חלק יבש- תרגיל בית 3:

#### שאלה 1:

יתכן שה-SSE יהיה שווה ל-0 אם מתקיים כי עבור כל קלסטר, כל הנקודות בו זהות. במקרה זה, בתוך כל קלסטר הצנטרואיד שווה לכל אחת מהנקודות, וכך סכום המרחקים  $G_i$  יהיה שווה לאפס לכל  $i$ .

#### שאלה 2:

2.1) מנוע החיפוש  $S$  יקבל ערך  $accuracy$  ששווה ל-1 ויאחזר 0 מסמכים כאשר כל המסמכים שבמאגר אינם רלוונטיים לשאילתא  $q$ .

$$RI = \frac{TP + TN}{TP + TN + FP + FN} = \frac{TP + TN}{\binom{n}{2}}$$

מכיוון שה- $accuracy$  שווה ל-1, נובע כי הסכום  $TP+TN$  שווה לכלל המסמכים במאגר, לכן מנוע החיפוש סיווג את הפריטים בצורה נכונה. נתון כי הוא אחזר 0 מסמכים, לכן כל המסמכים אינם רלוונטיים לשאילתא.

2.2) מדד ה- $accuracy$  מטעה בדוגמה שהצגנו משום שקיבלנו את ה- $accuracy$  הגבוה ביותר שניתן, ולכן לכאורה יכולנו להסיק כי מנוע החיפוש מניב תוצאות טובות. אך בפועל, לכל שאילתא לא נקבל אף מסמך.

לכן נסתכל על המדדים  $precision$  ו- $recall$  שערכיהם אפס, דבר המשקף בצורה טובה יותר את איכות מנוע החיפוש.

$$Precision = \frac{TP}{FP + TP} \qquad Recall = \frac{TP}{TP + FN}$$