

תרגיל ריצה – הסקה ברשת בייסיאנית

הבעיה

עליכם לממש רשת בייסיאנית ואלגוריתמי הסקה הסתברותית. האלגוריתמים שעליכם לממש:

1. Bayes ball.
2. Variable elimination, עם הורדת משתנים מיותרים בהתחלה (כמו שמתואר בשקפים 91-92 במצגת). סדר האלימינציה של המשתנים יינתן כקלט.

מכיוון שזמני הריצה שונים ממחשב למחשב, הפרמטר שאותו התוכנית תמדוד הוא מספר פעולות הכפל והחיבור שנעשו.

שימו לב: באלגוריתם 2 יש צורך לקבוע את סדר הכפלת ה-factors כאשר מבטלים משתנה. לשם הפשטות, הסדר יהיה מהקטן לגדול. במידה ויש כמה פקטורים באותו גודל, הם יסודרו לפי סכום ערכי ה-ASCII של המשתנים שלהם (מהקטן לגדול).

מימוש

קלט

הקלט לתוכנית מורכב מהקובץ input.txt ומקובץ נוסף בפורמט xml שמכיל את מבנה הרשת. לתרגיל מצורף קובץ input.txt לדוגמה, ו-2 קבצי xml לדוגמה.

בקובץ input.txt השורה הראשונה תכיל את שם קובץ ה-xml שמכיל את מבנה הרשת. שאר הקובץ יכיל את השאלות.

השאלות לאלגוריתם Bayes ball תמיד תהיה במבנה הבא: $A-B|E_1=e_1, E_2=e_2, \dots, E_k=e_k$. ז"א השאלה היא האם $A-B$ בלתי תלויים, בהינתן ש- $E_1=e_1, E_2=e_2, \dots$. במידה ואין ערכים למשתני evidence השאלה תיראה כך: $A-B|$.

השאלות לאלגוריתם Variable elimination תמיד תהיה במבנה הבא:

$$P(Q=q|E_1=e_1, E_2=e_2, \dots, E_k=e_k) H_1-H_2-\dots-H_j$$

ז"א השאלה תמיד תשאל על ההסתברות לקבל ערך עבור משתנה query אחד, כאשר נתונים מספר משתני evidence, ואחרי השאלה יינתן הסדר בו יש לבצע את האלימינציה של המשתנים החבויים. במידה ואין משתנים חבויים השאלה תיראה כך:

$$P(Q=q|E_1=e_1, E_2=e_2, \dots, E_k=e_k)$$

במידה ותוצאת השאלה כבר נתונה באחד מה- CPT, אין לחשב אותה אלא לתת אותה מיד.

מבנה קובץ ה-xml הוא כדלהלן. הקובץ יתחיל בתגית <NETWORK> ויסתיים ב-</NETWORK>. בתחילת הקובץ נתונים המשתנים. כל משתנה יסומן בתג <VARIABLE>, השם שלו יופיע אחרי התג <NAME> והערכים האפשריים שהוא יכול לקבל יופיעו אחרי התג <OUTCOME>. לדוגמה, המשתנה הבינארי E ייכתב כך בקובץ:

```
<VARIABLE>
  <NAME>E</NAME>
  <OUTCOME>T</OUTCOME>
  <OUTCOME>F</OUTCOME>
</VARIABLE>
```

אחרי רשימת כל המשתנים יופיעו ההגדרות של ה-CPT. כל CPT יסומן בתג <DEFINITION>. אחרי יופיע התג <FOR> שיציין באיזה משתנה מדובר, ואחרי התג <GIVEN> לכל הורה של המשתנה. בסוף יופיע התג <TABLE> ואחרי יופיעו כל הערכים של ה-CPT, מופרדים על ידי רווחים, ומסודרים לפי הסדר בו הופיעו הערכים של ההורים ושל המשתנה. לדוגמה, אם הגדרנו משתנים בינאריים B ו-A בדיוק כמו שהגדרנו את E בדוגמה הקודמת, ונרצה לתאר את ה-CPT של A כאשר E ו-B הם ההורים שלו, ההגדרה תיראה כך:

```
<DEFINITION>
  <FOR>A</FOR>
  <GIVEN>E</GIVEN>
  <GIVEN>B</GIVEN>
  <TABLE>0.95 0.05 0.29 0.71 0.94 0.06 0.001 0.999</TABLE>
</DEFINITION>
```

ניתן לראות בדוגמה זו ש- $P(A=T|E=T, B=T)=0.95$ וגם ש- $P(A=F|E=F, B=T)=0.06$.

צירפתי למטלה קובץ עם דוגמה נוספת כדי שיהיה מובן מהו סדר הערכים בטבלה.

פלט

הפלט ייכתב לקובץ output.txt. בשורה הראשונה תיכתב תוצאת השאלתה הראשונה, בשורה הבאה תוצאת השאלתה הבאה וכן הלאה.

אם השאלתה היא לאלגוריתם Bayes ball יש לכתוב אך ורק yes אם המשתנים בלתי תלויים, ו- no אחרת. אם השאלתה היא לאלגוריתם Variable elimination יש לכתוב את תוצאת השאלתה (מעוגלת ל- 5 מקומות אחרי הנקודה), פסיק, מספר פעולות החיבור שנדרשו ע"י האלגוריתם כדי לענות על השאלתה, פסיק, ואז מספר פעולות הכפל שנדרשו. אין להוסיף שום סימן או תו אחר בקובץ. שימו לב: בפעולות הנירמול יש לספור את פעולות החיבור, אך אין לספור שם פעולות כפל (כי יש שם חילוק).

אופן הניקוד

- קוד נכון, שמממש את האלגוריתמים, ומחזיר את התוצאה המבוקשת על כל הקלטים החוקיים.
- קוד מתועד וקריא (שמות משתנים ופונקציות משמעותיים).
- הגשה בזמן.

פרטי ההגשה

- ההגשה לפי כל כללי הגשת המטלות, כפי שמתואר בנוהל הגשת המטלות (נמצא במודל).
- ניתן לכתוב את התוכנית ב- Java בלבד, והיא צריכה להתקמפל ולרוץ בגרסה 1.8.0_232. שם המחלקה בה נמצאת פונקציית ה- main יהיה Ex1. יש להשתמש ב- default-package בלבד (ללא תתי תיקיות). חובה להגיש את קבצי המקור.
- אין לממש GUI. קובץ ה- input.txt, קובץ ה- XML שמתאר את הרשת, וקובץ ה- output.txt יהיו באותה ספרייה בה נמצאת התוכנית, ולכן אין לציין ספרייה ספציפית בקוד שאתם מגישים (במידה וכן, ירדו על כך נקודות).
- ייתן קלט ופלט לדוגמה. ודאו שתוכניתכם עובדת אתו כמו שצריך, אך זהו לא הקלט היחיד אתו תיבדק התוכנית.
- התוכנית תיבדק דרך ה- command line ולא ב- eclipse. לכן, כדי לוודא שהתוכנית שלכם עובדת עליכם להעתיק את קבצי המקור, את הקובץ input.txt שניתן כדוגמה, ואת קובץ ה- xml המתאים לאחת הספריות במחשב, לפתוח command line ולהריץ `javac *.java` ואז `java Ex1`. התוכנית תיצור את הקובץ output.txt באותה ספרייה והוא צריך להיות זהה לקובץ output.txt שניתן כדוגמה.
- ההגשה נעשית דרך מערכת הבדיקות האוטומטית. בכתובת: <http://submit.org.il/ariel> שם המשתמש שלכם הוא שם המשתמש שהוקצה לכם באי-מייל שלכם באריאל, כמו שכתוב במכלול. לדוגמה, אם הדואר שלכם במכלול הוא israel.israeli@msmail.ariel.ac.il, אז שם המשתמש שלכם הוא: israel.israeli. הסיסמה היא 4 הספרות האחרונות של תעודת הזהות.
- ברגע שתגישו את המטלה תקבלו למייל הפרטי שלכם שרשום במכלול (לא המייל באריאל) את תוצאות ההגשה- האם התרגיל התקמפל, ואם כן מה היה הפלט שלכם בהשוואה למה הפלט הרצוי. תוכלו כמובן להגיש שוב ולדרוס את ההגשה הקודמת. שימו לב, המייל מהמערכת עלול להגיע לתיקיית הספאם. בדקו שם וסמנו אותו כ- "not spam". בנוסף, המערכת לא עובדת כמו שצריך עם מיילים של walla.
- חשוב להבהיר: הציון של הבדיקה האוטומטית הוא לא הציון הסופי שלכם במטלה. בבדיקה עצמה ייבדקו קלטים נוספים וייבדק הקוד שלכם. הציון כן מיועד לתת לכם אינדיקציה לגבי מצב התרגיל שלכם.
- תאריך הגשת התרגיל- 13.06.24

בהצלחה!