

תרגיל ריצה – הסקה ברשת בייסיאנית

הבעיה

עליכם לממש רשת בייסיאנית ואלגוריתמי הסקה הסתברותית שונים, ולהשוות ביניהם. האלגוריתמים שעליכם לממש:

1. הסקה פשוטה, ללא שום שיפורים, כמו שמתואר בשקף 67 במצגת.
2. Variable elimination, עם הורדת משתנים מיותרים בהתחלה (כמו שמתואר בשקף 91 במצגת), כאשר סדר האלימינציה של המשתנים הוא לפי סדר ה-ABC.
3. כמו 2, כאשר אתם קובעים בצורה היוריסטית את סדר האלימינציה של המשתנים.

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

שימו לב: אין צורך לממש את Bayes-ball, ולכן באלגוריתמים 2 ו-3 אין צורך להוריד משתנים מיותרים ע"י שימוש באי-תלות מותנית (כמו שמתואר בשקף 92 במצגת). אומנם, באלגוריתמים אלו 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 השאילתות.

יכולות להיות 2 סוגי שאילתות:

1. שאילתה לחישוב joint probability (=תא בטבלה) - השאילתה תהיה במבנה הבא:

$$P(E_1=e_1, E_2=e_2, \dots, E_n=e_n)$$
 ז"א השאילתה תשאל על ההסתברות המשותפת של כל המשתנים. במקרה זה הערכים של כל המשתנים יהיו נתונים בשאילתה, ויש לבצע חישוב ישיר כמו שמתואר בשקף 56 במצגת.
2. שאילתה לאלגוריתם - השאילתה תהיה במבנה הבא:

$$P(Q=q|E_1=e_1, E_2=e_2, \dots, E_k=e_k), 1$$
 ז"א השאילתה תשאל על ההסתברות לקבל ערך עבור משתנה query אחד, כאשר נתונים מספר משתני evidence. אחרי השאילתה מצוין באיזה אלגוריתם להשתמש: 1, 2, או 3, לפי מה שנכתב קודם. במידה ותוצאת השאילתה כבר נתונה באחד מה-CPT (ז"א שהשאילתה זהה בדיוק למה שנמצא כבר באחת הטבלאות) אין לחשב אותה אלא לתת אותה מיד, ללא תלות באלגוריתם שנבחר.

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 בקובץ יש לכתוב את תוצאת השאלתה הראשונה, פסיק, מספר פעולות החיבור שנדרשו כדי לענות על השאלתה, פסיק, ואז מספר פעולות הכפל שנדרשו. בשורה הבאה יש לכתוב את אותם הנתונים עבור השאלתה השנייה וכן הלאה. אין להוסיף שום סימן או תו אחר בקובץ. את ההסתברויות יש לכתוב עם 5 ספרות אחרי הנקודה. במידה ולא היה צריך לעשות פעולות חיבור ו/או כפל יש לכתוב 0 במקום המתאים. שימו לב: בפעולת הנירמול יש לספור את פעולות החיבור, אך אין לספור שם פעולות כפל (כי יש שם חילוק).

בנוסף יש להגיש דרך המודל קובץ Word או Pdf בשם details. בתחילת הקובץ יש לכתוב את פרטי המגיש (שם ות.ז.). לאחר מכן יש לתאר במילים את היוריסטיקה בה בחרתם להשתמש כדי לקבוע את סדר האלימינציה של המשתנים באלגוריתם 3.

אופן הניקוד

- קוד נכון, שמממש את האלגוריתמים, ומחזיר את התוצאה המבוקשת על כל הקלטים החוקיים.
- איכות הפונקציה היוריסטית בה בחרתם להשתמש.
- קוד מתועד וקריא (שמות משתנים ופונקציות משמעותיים). אין לכתוב הערות בעברית.
- 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. 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.
- עליכם להניח שקובץ ה-input.txt נמצא באותה ספרייה בה נמצאת התוכנית, ולכן אין לקרוא את המיקום שלו כארגומנט או לציין ספרייה ספציפית בקוד שאתם מגישים (במידה וכן, ירדו על כך נקודות).
- קובץ ה-output.txt צריך להיכתב באותה ספרייה בה נמצאת התוכנית, ולכן אין לקרוא את המיקום שלו כארגומנט או לציין ספרייה ספציפית בקוד שאתם מגישים (במידה וכן, ירדו על כך נקודות). יש להניח שהקובץ לא קיים, ולכן עליכם ליצור אותו בספרייה בה התוכנית רצה ולכתוב לתוכו את הפלט.
- ייתן קלט ופלט לדוגמה. ודאו שתוכניתכם עובדת אתו כמו שצריך, אך זהו לא הקלט היחיד אותו 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 `Ex1.java`. 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.
- הגשת קובץ ה-details נעשית דרך המודל, באותו תאריך של הגשת קבצי המקור במערכת ה-submit.
- תאריך הגשת התרגיל- 29.04.25
- שימו לב: אין לחכות לרגע האחרון כדי להגיש. לקראת סוף זמן ההגשה השרת עמוס, ולכן ה-feedback מתקבל בשלב מאוחר יותר (אחרי שתאריך ההגשה כבר עובר).

בהצלחה!