

תרגיל ריצה – הסקה ברשת בייסיאנית

הבעיה

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

1. Bayes ball, כמו שמתואר בשקף 75 במצגת.
2. Variable elimination, עם הורדת משתנים מיותרים בהתחלה (כמו שמתואר בשקף 91 במצגת). סדר האלימינציה של המשתנים יינתן כקלט. סדר הכפלת ה-factors כאשר מבטלים משתנה יהיה תמיד מהקטנים לגדולים (מבחינת מספר השורות), כאשר יש לעדכן את הסדר במידה והכפלת מסוימת גרמה לפקטור החדש להיות גדול יותר. לדוגמה, נניח שהיו ארבעה factors שמכילים את המשתנה A שאותו אנחנו רוצים לבטל. נניח שהגדלים שלהם הם: 3, 3, 4, 4. ניקח את שני ה-factors הראשונים (בגודל 3) ונכפיל אותם. אם הפקטור החדש שנוצר יכול 5 שורות, ההכפלה הבאה תהיה של שני ה-factors האחרונים (שמכילים 4 שורות), וכן הלאה.

מימוש

קלט

התוכנית תקרא את כל הקלט שלה מקובץ יחיד - input.txt. הקובץ יכול את מבנה הרשת, ואת השאילתות השונות. לתרגיל מצורף קובץ input.txt לדוגמה, שמכיל את הרשת הבייסיאנית של האזעקה שלמדנו בכיתה (מופיעה בשקף 60 במצגת), וקובץ input2.txt שמכיל רשת בייסיאנית בה חלק מהמשתנים אינם בינאריים.

בכל קובץ השורה הראשונה תתחיל תמיד במילה Network, כדי לסמן שזהו החלק המתאר את הרשת. בשורה השנייה יהיו רשומים שמות המשתנים. לאחר מכן, עבור כל משתנה יהיו רשומים הערכים האפשריים שלו, מי הם ההורים שלו ברשת, וה-CPT. כדי לתאר את ה-CPT תהיה שורה עבור כל צירוף ערכים של הורי המשתנה, ואז ההסתברות עבור כל ערך של המשתנה. לדוגמה, למשתנה B בקובץ input.txt אין הורים, ולכן רשום none בחלק של ה-parents, וב-CPT שלו רשום 0.001, 0.001 = true, שזה אומר $P(B=true) = 0.001$. למשתנה A בקובץ input.txt יש שני הורים, B ו-E, ולכן כשרשומה השורה הבאה ב-CPT של $0.29, 0.29, 0.29, 0.29$ = true, false, true, false זה אומר ש- $P(A=true|B=false, E=true) = 0.29$. למשתנה C בקובץ input2.txt יש 2 הורים, A ו-B, כאשר A הוא בינארי, אבל B ו-C מקבלים 3 ערכים אפשריים, ולכן כשרשומה השורה הבאה ב-CPT של $0.2, 0.3, 0.3, 0.3$ = stay, go, maybe, true זה אומר ש- $P(C=go|A=true, B=maybe) = 0.3$ וגם ש- $P(C=stay|A=true, B=maybe) = 0.2$.

לאחר מכן יתחיל החלק של השאילתות, שיסומן במילה Queries. השאילתה לאלגוריתם Bayes ball תמיד תהיה במבנה הבא: $A-B|E_1=e_1, E_2=e_2, \dots, E_k=e_k$. ז"א השאילתה היא האם A ו-B בלתי תלויים, בהינתן ש- $E_1=e_1, E_2=e_2, \dots, E_k=e_k$. במידה ואין ערכים למשתני evidence השאילתה תיראה כך: $A-B|$. השאילתה 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 השאילתה תמיד תשאל על ההסתברות לקבל ערך עבור משתנה query אחד, כאשר נתונים מספר משתני evidence, ואחרי השאילתה יינתן הסדר בו יש לבצע את האלימינציה של המשתנים החבויים. במידה ואין משתנים חבויים השאילתה תיראה כך: $P(Q=q|E_1=e_1, E_2=e_2, \dots, E_k=e_k)$. אין צורך לבדוק האם תוצאת השאילתה כבר נתונה באחד מה-CPT, אלא האלגוריתם תמיד יחשב אותה.

פלט

הפלט ייכתב לקובץ output.txt. בשורה הראשונה תיכתב תוצאת השאילתה הראשונה, בשורה הבאה תוצאת השאילתה הבאה וכן הלאה. אם השאילתה היא לאלגוריתם Bayes ball, יש לכתוב אך ורק yes אם המשתנים בלתי תלויים, ו-no אחרת. אם השאילתה היא לאלגוריתם Variable elimination יש לכתוב את תוצאת השאילתה (מעוגלת ל-5 מקומות אחרי הנקודה), פסיק, מספר פעולות החיבור שנדרשו ע"י האלגוריתם כדי לענות על השאילתה, פסיק, ואז מספר פעולות הכפל שנדרשו. אין להוסיף שום סימן או תו אחר בקובץ.

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 ביחידים בלבד. אין להעתיק אחד מהשני או ממקור חיצוני. תתבצע בדיקת העתקות. תרגיל מועתק (גם בחלקו) יגרור פסילה מיידית של 2 הצדדים המעורבים, בלי קשר למי העתיק ממי.
- ניתן לכתוב את התוכנית ב-Java בלבד, והיא צריכה להתקמפל ולרוץ בגרסה 1.8.0_144. שם המחלקה בה נמצאת פונקציה 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 וקובץ ה-output.txt יהיו באותה ספרייה בה נמצאת התוכנית, ולכן אין לציין ספרייה ספציפית בקוד שאתם מגישים (במידה וכן, יירדו על כך נקודות).
- יינתן קלט ופלט לדוגמה. ודאו שתכונתכם 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 ואת הקובץ input.txt שניתן כדוגמה לאחת הספריות במחשב, לפתוח command line ולהריץ `javac java * ex1`. התוכנית תיצור את הקובץ output.txt באותה ספרייה והוא צריך להיות זהה לקובץ output.txt שניתן כדוגמה.
- ההגשה נעשית דרך ה-Moodle.
- תאריך הגשת התרגיל- 14.12.17

בהצלחה!