

## קורס 67800 תשע"ג

### בחינה מועד א'

הנחיות:

- בבחינה חמש שאלות. עליכם לענות על שלוש מתוכן (משקל כל שאלה 33 נקודות).
- לא ייבדקו יותר משלוש שאלות כך שאם פתרתם שאלות נוספות עליכם לבחור אילו שאלות ייבדקו ולסמן זאת בצורה ברורה.
- עליכם לנמק היטב כל תשובה ולהוכיח באופן מדויק כל טענה שנדרשתם להוכיח. תשובה נכונה ללא נימוק ו/או דרך לא תזכה אתכם בנקודות.
- עליכם לענות על כל השאלות הבאות באופן עצמאי. תלמיד שיתפס מעתיק, או שיהיה קיים חשש כי העתיק יועבר לטיפול רשויות האוניברסיטה.
- אין להשתמש בחומר עזר מכל סוג שהוא.
- (1 נקודות) ודאו כי מספר הזהות שלכם רשום על מחברת הבחינה.

משך הבחינה: שלוש שעות

בהצלחה !

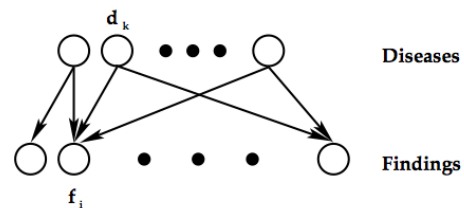
1. בשאלה זו נעסוק בגודל הפקטור המקסימלי בעת ביצוע אלימינציה למשתנים (variable elimination). נגדיר רוחב של סדר אלימינציה עבור גרף מעל  $X_1, \dots, X_n$  להיות מספר המשתנים המופיעים בפקטור הגדול ביותר בחישוב (לפני סכימה) פחות אחד. שני סדרי אלימינציה יהיו מאותה מחלקת שקילות אם יש להם אותו הרוחב. לדוגמא, עבור הרשת  $X_1-X_2$ , ישנם שני סדרי אלימינציה (קודם 1 ואח"כ 2 או להיפך) ושניהם באותה מחלקת שקילות. גודל מחלקת שקילות יהיה מספר סדרי האלימינציה במחלקה.

(א) (7 נקודות) נתון גרף שרשרת  $X_1-X_2-X_3$ . תארו את מחלקות השקילות של סדרי אלימינציה עבור רשת זו ואת גודלן.

(ב) (13 נקודות) נתון גרף שרשרת עם  $n$  משתנים. תארו את מחלקות השקילות של סדרי אלימינציה עבור גרף זה ואת גודלן.

(ג) (13 נקודות) נתון גרף לא מכוון בצורת בו כל אחד מ  $X_2, \dots, X_n$  מחובר רק ל  $X_1$ . מהן מחלקות השקילות בעלות רוחב מינימלי ומקסימלי עבור גרף זה ומה גודלן?.

2. בשאלה זו נבחן את השימוש באלגוריתם גיבס לביצוע inference ברשת לניבוי הסיכוי לגורמי תחלואה שונים בהינתן סימפטומים. במודל הדו-שכבתי כל משתנה מחלה  $d_i$  הוא הורה של סימפטומים  $f_j$  היכולים להיגרם מהמחלה (ראו ציור להמחשה).



(א) (8 נקודות) לצורך סעיף זה בלבד, נניח כי הייצוג של כל סימפטום בהינתן ההורים הוא OR דטרנמיניסטי. האם אלגוריתם גיבס הרגיל עלול להיתקע בהשמה מסוימת? ניתן לתת נימוק מפורט בהקשר של רשת קטנה עם שתי מחלות וסימפטום אחד.

(ב) (10 נקודות) נניח כי על מנת להאיץ את גיבס, רוצים לדגום שני משתנים בכל פעם. עבור רשת בייסיאנית כללית, באילו מהטבלאות ברשת יהיה חובה להשתמש עבור ביצוע החישובים הנדרשים לצמד משתנים מסוים? נמקו!

(ג) (15 נקודות) עבור הרשת הרפואית המתוארת לעיל, הוצע אלגוריתם גיבס הבא: בהינתן השמה למחלות, דוגמים את כל הסימפטומים יחדיו, ולאחר מכן דוגמים כל מחלה באופן נפרד כמו בגיבס רגיל. תארו את החישובים המעורבים באלגוריתם זה והסבירו מדוע נרצה להשתמש בו.

3. בשאלה זו נעסוק בלימוד יעיל של מבנה רשת בייסיאנית בעזרת פונקציית ציון הנראות. הניחו פרמטריזציה מלאה של המודלים ושאלו ערכים חסרים בנתונים.

(א) **(7 נקודות)** להזכירכם, מודל TAN הוא רשת בייסיאנית בה קיים צומת C שהוא הורה של צמתים  $X_1, \dots, X_n$ , ובנוסף לקשתות המחברות בין צמתים אלו מבנה של עץ פורס. הסבירו כיצד ניתן ללמוד את העץ הזה ביעילות.

(ב) **(13 נקודות)** נסמן את מבנה הרשת הבייסיאנית משאלה 2 ב G (אין צורך לפתור את שאלה 2 לצורך סעיף זה). כעת רוצים ללמוד גרף G' שהוא G ועוד קשתות של עץ פורס כלשהו על **משתני הסימפטומים בלבד**. האם ניתן ללמוד את G' האופטימלי ביעילות? אם כן, הסבירו כיצד לעשות זאת. אם לא, הסבירו היטב מדוע לא (אין צורך בהוכחת קושי פורמלית).

(ג) **(13 נקודות)** נניח כי כעת רוצים ללמוד מבנה רשת כללי בו לכל משתנה לכל היותר K הורים. כמו כן נתון הסדר הטופולוגי הנכון של המשתנים. הראו שניתן ללמוד את המבנה האופטימלי בזמן פולינומיאלי במספר המשתנים עבור K קבוע.

4. בשאלה זו נתייחס לגרפים G שהם **מכוונים** וחסרי מעגלים (כלומר DAGs). בכיתה הגדרנו את אוסף תכונות חוסר התלות  $I_{LM}(G)$  עבור גרפים כאלו. נגדיר כעת אוסף תכונות  $I_{pair}(G)$  כך:  

$$I_{pair}(G) = \{X_i \perp X_j \mid X_{ND(i) \setminus j} : j \in ND(i) \setminus Pa(i)\}$$
 כאשר  $ND(i)$  הם הצמתים שאינם צאצאים של i (כלומר גם ההורים של i הם ב  $ND(i)$ ).

א. **(11 נקודות)** הראו שעבור פילוג חיובי ממש  $\rho$  (כלומר אין מאורע בהסתברות אפס) על משתנים  $W, X, Y, Z$  מתקיים:  $(X \perp Y \mid Z, W) \& (X \perp W \mid Z, Y) \Rightarrow (X \perp Y, W \mid Z)$ .

ב. **(11 נקודות)** נתון פילוג  $p(x_1, \dots, x_n)$  וגרף G. הראו שאם  $I(p) \supseteq I_{LM}(G)$  אזי  $I(p) \supseteq I_{pair}(G)$ . אפשר להשתמש בעובדה ש:  $(W \perp X, Y \mid Z) \Rightarrow (W \perp X \mid Y, Z)$ .

ג. **(11 נקודות)** נתון פילוג  $p(x_1, \dots, x_n)$  חיובי ממש וגרף G. הראו שאם  $I(p) \supseteq I_{pair}(G)$  אזי  $I(p) \supseteq I_{LM}(G)$ .

5. שאלה זו תתייחס ללמידת פרמטרים ואלגוריתם EM.

א. **(13 נקודות)** נתון מודל פרמטרי  $p(x_h, x_o; \theta)$  כאשר  $x_h, x_o$  שתי קבוצות של משתנים, ורק  $x_o$  נצפה. הראו שאלגוריתם EM עבור  $\theta$  יעצור אם ורק אם  $\theta$  היא נקודה סטציונרית של הנראות של  $x_o$ . הניחו שבפעולת המקסימיזציה יש  $\text{argmax}$  יחיד, ושלומדמים עם נקודת מדגם יחידה.

ב. **(13 נקודות)** נתונה משפחה פרמטרית  $p(x_1, x_2, x_3; \theta) = \theta_a(x_1)\theta_a(x_2)\theta_b(x_3 \mid x_1, x_2)$  כלומר פרמטריזציה של רשת בייסיאנית עם מבנה V, כאשר לפילוג על  $x_1$  ו  $x_2$  יש את אותם פרמטרים. נתון מדגם המגדיר התפלגות אמפירית  $p_D(x_1, x_2, x_3; \theta)$ . מצאו ביטוי סגור לפרמטרי Maximum Likelihood כפונקציה של פילוג זה.

## ג. (13 נקודות) נתונה משפחה פרמטרית

כלומר פרמטריזציה מלאה של רשת  $p(x_1, x_2, x_3; \theta) = \theta_1(x_1)\theta_2(x_2)\theta_3(x_3 | x_1, x_2)$  בייסיאנית עם מבנה  $V$ . כעת הניחו שהמשתנים הנצפים היחידים הם  $x_1$  ו  $x_2$  והפילוג האמפירי שלהם הוא  $p_D(x_1, x_2; \theta)$ . אתם משתמשים ב EM להעריך את הפרמטרים של המודל ומתחילים בפרמטרים  $\theta^{(1)}$ . תוך כמה איטרציות יתכנס האלגוריתם ולאילו פרמטרים? בטאו תשובתכם כפונקציה של הפרמטרים ההתחלתיים והפילוג האמפירי הנתון.