ת.ז. ____________________ תאריך: 28/02/2012

קורס 67800 תשע"ב

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 בתכונות פונקציות ציון (Score Functions) לרשתות בייסיאניות. פונקצית ציון תקרא שקולת ציון אם היא מחזירה אותו ערך לשתי רשתות שונות מאותה מחלקת שקילות. להזכירכם, שתי רשתות הן באותה מחלקת שקילות אם הן מייצגות את אותן אי-תלויות. בכל הסעיפים נניח שהפרמטריזציה מלאה (כלומר כל ה CPD מיוצגים על ידי טבלאות מלאות).

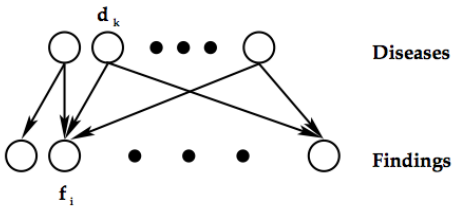
א. (11 נקודות) הוכיחו כי אם שתי רשתות הן באותה מחלקת שקילות אז ציון הנראות שלהן זהה.

ב. (11 נקודות) נניח שיש לנו רק שני משתנים X ו Y. הוכיחו כי אם שתי רשתות הן באותה מחלקה שקילות אז ציון ה BIC זהה לשתי הרשתות. להזכירכם, BIC שווה לציון הנראות מינוס $\frac{\log M}{2} Dim(G)$ כאשר M מספר הדוגמאות ו $Dim(G)$ מספר הפרמטרים ה<u>חופשיים</u> במודל.

ג. (11 נקודות) פריור K2 הוא פריור דיריכלה ל- $\theta_i(x_i \mid x_{Pa(i)})$ כך שכל הפרמטרים של הפריור (ה- α השונים) שווים ל-1. הוכיחו כי הציון הבייסיאני עם פריור K2 אינו שקול ציון.
רמז: מספיק להראות דוגמא בה שתי רשתות מאותה מחלקת שקילות מקבלות ציון שונה.
להזכירכם, הציון הבייסיאני הוא:

$$\log P(D \mid G) = \sum_i \sum_{x_{Pa(i)}} \left(\log \frac{\Gamma(\sum_{x_i} \alpha_{x_i \mid x_{Pa(i)}})}{\Gamma\left(\sum_{x_i} \alpha_{x_i \mid x_{Pa(i)}} + M[x_{Pa(i)}]\right)} + \sum_{x_i} \log \frac{\Gamma\left(\alpha_{x_i \mid x_{Pa(i)}} + M[x_i, x_{Pa(i)}]\right)}{\Gamma(\alpha_{x_i \mid x_{Pa(i)}})} \right)$$

2. בשיעור האחרון ראינו מערכת מומחה לניבוי הסיכוי לגורמי תחלואה שונים בהינתן סימפטומים המבוססת על מודל גרפי דו-שכבתי בו כל משתנה מחלה d_i הוא הורה של סימפטומים f_i היכולים להיגרם מהמחלה (ראו ציור להמחשה).



בשאלה זו נבחן את האפשרות לבצע הסקה בעזרת אלגוריתם Loopy Belief Propagation ברשת זו. לצורך נוחות הניחו לאורך כל השאלה שכל המשתנים ברשת הם בינאריים.

א. (11 נקודות) מהו תנאי הכרחי על מבנה הרשת כך שבהינתן תצפיות על f_1,f_2 אלגוריתם Loopy תמיד יחזיר תשובה מדויקת?

ב. (11 נקודות) נניח כעת כי התנאי ב (א) לא מתקיים. <u>בהינתן כל</u> המשתנים מסוג f_i, אילו הודעות צריך לחשב במסגרת האלגוריתם? תנו נוסחה לכל סוג הודעה. אילו הודעות קשות לחישוב במקרה הכללי?

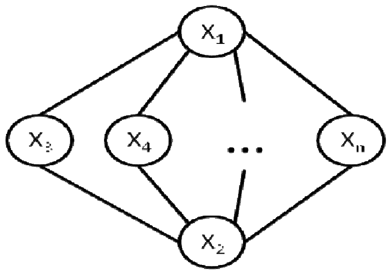
ג. (11 נקודות) לסטודנט עירני הסתבר כי להתפלגות של f_i בהינתן הוריו צורה מיוחדת: אם לפחות אחד מההורים שווה ל-1, אזי $P(f_i=1|par_i)=\alpha$, אחרת, ההסתברות ש f_i יהיה 1 שווה ל β. האם ניתן במקרה זה לחשב את ההודעות מקליקות גדולות (של f_i והוריו) למשתנה f_i יותר ביעילות? אם לא, הסבירו מדוע. אם כן, תארו במדויק כיצד יתבצע החישוב היעיל.

3. בשאלה זו נבחן את אלגוריתם ה Cutset conditioning , דרך אלטרנטיבית לביצוע הסקה מדוייקת. השיטה מתבססת על התניה על חלק מהמשתנים, כך שבהינתן המשתנים שנבחרו, קל לבצע חישובים. בשאלות נניח שנתון גרף לא מכוון G וקבוצת קדקודים C כך שאם מוציאים את C והקשתות שלהם מ G, מקבלים עץ T. נסמן ב k את הגודל של C.

א. (11 נקודות) נניח שרוצים לחשב את הפילוג השולי של אחד המשתנים שאינו ב C. נסמן משתנה זה ב Y, ואת שאר המשתנים ב-X. הסבירו כיצד ניתן לחשב את הסכום

$P(Y=y)=\sum_{x,c} p(x,y,c)$ באופן יעיל תוך שימוש בהנחה על C. הניחו ש k הוא "קטן" כך שניתן לעשות מספר פעולות אקספוננציאלי ב k.

ב. (11 נקודות) נתון הגרף הבא:



מהו k המינימלי? מהו ה tree-width של הגרף? מה מסקנותיכם לגבי יעילות שיטת ה cutset conditioning? נמקו היטב את כל תשובותיכם!

ג. (11 נקודות) דרך אחרת לחישוב פילוג שולי היא על ידי Variable Elimination. הסבירו עבור גרף כללי מדוע מבחינת הזיכרון הנדרש לחישוב ייתכן שנעדיף את שיטת ה cutset. הניחו שה tree-width של הגרף G שווה ל k+1.

4. נתון מודל מכוון $X_1 \to X_2 \to X_3$. רוצים ללמוד את הפרמטרים שלו מתוך מדגם. נסמן ב $p_D(x_1,x_2,x_3)$ את הפילוג האמפירי. הניחו את הפרמטריזציה הבאה:
$p(x_{i+1} | x_i)=\theta_p(x_{i+1} | x_i), p(x_1)=\theta_s(x_1)$ (כלומר משתמשים באותה טבלה לכל הקשתות).

א. (11 נקודות) מצאו ביטוי סגור לפרמטרים הנראות המירבית כאשר כל המשתנים נצפים. בטאו את הפרמטרים הללו כפונקציה של ההתפלגויות השוליות האמפיריות על משתנים בודדים וזוגות של משתנים.

ב. (11 נקודות) כעת הניחו שנתון מדגם שבו המשתנה X_2 תמיד חבוי. כתבו במפורש את נוסחאות העדכון של אלגוריתם EM עבור הפרמטרים של המודל. הניחו פרמטריזציה כמו בסעיף א.

ג. (11 נקודות) נניח כעת מודל לא מכוון:
$p(x_1,x_2,x_3;\theta)\propto e^{\theta_s(x_1)+\theta_s(x_2)+\theta_s(x_3)+\theta_p(x_1,x_2)+\theta_p(x_2,x_3)}$ על אותו גרף. שימו לב שמשתמשים באותם פרמטרים לכל הפוטנציאלים הבודדים ולכל הפוטנציאלים על הזוגות. כתבו את המשוואה אותה אמורים לקיים פרמטרי ML. פתרו אותה עבור מדגם אמפירי אחיד (בו כל ההשמות האפשריות מופיעות אותו מספר פעמים). וודאו שמתקבלת תוצאה הגיונית.

5. בהנתן גרף **מכוון** G ללא מעגלים על n קדקודים נאמר שהוא P-map עבור פילוג $p(x_1,x_2,...,x_n)$ אם מתקיים $I(p)=I_{d-sep}(G)$. בכל סעיפי השאלה גרפים מכוונים יהיו ללא מעגלים גם כאשר לא מציינים זאת.

א. (11 נקודות) מהו ה n המקסימאלי כך שלכל פילוג $p(x_1,x_2,...,x_n)$ קיים גרף G שהוא P-map עבורו? הוכיחו את טענתכם.

ב. (11 נקודות) נניח שעבור $p(x_1,x_2,...,x_n)$ מסוים ידוע ש G^* הוא P-map עבור p. נתונים שני משתנים X_1,X_2 כך שהקשת $X_1\rightarrow X_2$ היא ב G^*. הוכיחו שלכל סט משתנים U שאינו מכיל את X_1,X_2 מתקיים $(X_1\perp X_2\mid U)\notin I(p)$.

ג. (11 נקודות) נניח שעבור $p(x_1,x_2,...,x_n)$ מסוים ידוע ש G^* הוא P-map עבור p. נתונים שלושה משתנים X_1,X_2,X_3 כך שהמבנה $X_1\rightarrow X_3\leftarrow X_2$ הוא ב G^*. נתון סט משתנים U שמכיל את X_3, מה מהבאים מתקיים? הוכיחו.

a. $(X_1\perp X_2\mid U)\notin I(p)$

b. $(X_1\perp X_2\mid U)\in I(p)$

c. שתי האופציות אפשריות.