

קורס 67800 תש"ע בחינה מועד א'

הנחיות:

- בבחינה חמש שאלות. עליכם לענות על שלוש מתוכם (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 המחברת.

משך הבחינה: שעתיים וחצי

בהצלחה !

1. נניח שבידינו עץ קליקות T מכויל (calibrated) מעל המשתנים X_1, \dots, X_N . בשאלה זו נבחן את השאלה כיצד לבנות עץ קליקות T' מכויל על חלק מהמשתנים $Y \subset X$ על ידי שימוש ב-T. מטרתנו היא כי T' יהיה קטן ככל האפשר אבל אין ברצוננו לבנות ולכיל את עץ הקליקות מההתחלה.

1. (8 נקודות) נניח כי T מורכב משתי קליקות: $C_1 = \{A, B, C\}$, $C_2 = \{C, D\}$. מה יהיה T' עבור $y = \{B, C\}$? שימו לב שעליכם לתאר לא רק את המבנה של T', אלא גם כיצד נעשה שימוש בהודעות המכוילות שחושבו עבור T.

2. (7 נקודות) עבור אותו T, מה יהיה T' עבור $y = \{A, C, D\}$?

3. (18 נקודות) האם גודל הקליקה המקסימלית ב-T' בהכרח קטן מזה שב-T כאשר מוציאים משתנה אחד מ-T? הוכיחו או הפריכו.

2. נניח כי אנו מעוניינים לבחון האם ערך מסוים C של פונקציה $f(X)$ של משתנה מקרי X הוא קיצוני כאשר X מתפלג לפי $P(X)$. כלומר ברצוננו לחשב

$$p\text{-value}(C) = E_{P(X)}[1\{f(x) > C\}]$$

1. (2 נקודות) נניח כי ביכלתכם לדגום ישירות מ- $P(X)$. תארו כיצד תחשבו את $p\text{-value}(C)$

2. (4 נקודות) על מנת לזהות תופעות לא סבירות, בדרך כלל נתעניין בערכי C עבורם מתקבל $p\text{-value}$ קטן. מה תהיה הבעיה בגישה שהצעת ב(א) במצב זה?

3. (5 נקודות) בכיתה ובתרגול ראינו כי Importance Sampling היא שיטה המאפשרת לקרב תוחלות של פונקציה $f(x)$ ביחס להתפלגות $P(X)$ על ידי דגימה מהתפלגות אחרת $Q(X)$. על מנת לתקן את הבעיה בציינת ב(ב) נרצה להשתמש בשיטה זו גם כאשר ניתן לדגום ישירות מ- $P(X)$. תאר איכותית מה נדרש מ- $Q(X)$ שיעזור לפתור את הבעיה.

4. (5 נקודות) נניח כעת כי בידך $Q(X)$ כפי שתארת ב(ג). מה הבעייתיות בשימוש ב- $Q(X)$ לחישוב ערכי $p\text{-value}$ גבוהים?

5. (12 נקודות) על מנת לפתור את כל הבעיות, מוצע לבצע את חישוב התוחלת על ידי דגימה מההתפלגות

$$Q'(X) = \alpha P(X) + (1 - \alpha) Q(X)$$

כאשר α הינו קבוע בין 0 ל-1. תאור כיצד יתבצע Importance Sampling במקרה זה ורשום נוסחה מפורשת למשקלות של כל דגימה.

6. (5 נקודות) מה היתרון בשיטה בשיטה המוצעת ב(ה)

3. בכיתה ראינו שניתן להסתכל על פונקציית הציון (model selection score) הבייטיאנית כמכפלה של הסתברויות ניבוי הדוגמא ה- i ית בהינתן הקודמות. בשאלה זו נבחן פונקציה עם מוטיביציה דומה (Leave one out cross validation score)

$$LOOCV(D:G) = \prod_{m=1}^M P(X[m] | G, D_{-m})$$

כאשר D_m הוא אוסף הנתונים ללא הדוגמה ה- m -ית.

1. (18 נקודות) עבור משתנה בינארי X , נשתמש בפריור דיריכלה $(\theta_0, \theta_1) \sim Dir(\alpha_0, \alpha_1)$, כאשר $\theta_0 = P(X=0)$. נשתמש ב- $M[0]$ עבור מספר הפעמים ש- X שווה ל-0 ב- D , ובדומה עבור $M[1]$. כמו כן נסמן ב- $x[i]$ את הערך אותו מקבל המשתנה בדוגמא ה- i -ית. הוכח כי במקרה זה

$$LOOCV(D:G) = \frac{1}{(M + \alpha_0 + \alpha_1 - 1)^M} \times (M[0] + \alpha_0 - 1)^{M[0]} \times (M[1] + \alpha_1 - 1)^{M[1]}$$

2. (5 נקודות) הכלל את (א) עבור המקרה המולטינומי.

3. (10 נקודות) הכלל את (א) עבור רשת עם שני משתנים בינאריים בה X הורה של Y .

4. נתון מודל פרמטרי $p(x_1, \dots, x_n; \theta)$. אתם מקבלים מדגם בגודל M המכיל רק חלק מהמשתנים $o \in \{1, \dots, n\}$. נסמן את המדגם ב- $\{x_o^m\}_{m=1}^M$ ואת קבוצת המשתנים שאינם נצפים ב- $h \in \{1, \dots, n\}$ (הניחו ש o ו h זרות ואיחודן הוא $\{1, \dots, n\}$).

א. (7 נקודות) הסבירו כיצד ניתן להשתמש באלגוריתם EM לאמידת הפרמטרים θ במקרה זה.

מהי הפונקציה שהאלגוריתם מנסה למקסם וכיצד מחושבים העדכונים של הפרמטר?

ב. (16 נקודות) נגדיר את הפונקציה הבאה: $F(\theta, q) = \sum_{m=1}^M \sum_{x_h} q_m(x_h) \log \frac{p(x_h, x_o^m; \theta)}{q_m(x_h)}$

כאשר q הוא אוסף של M פילוגים על המשתנים x_h . הראו שעבור θ קבוע, המקסימום של F ביחס ל q נתון ע"י: $q_m(x_h) = p(x_h | x_o^m; \theta)$.

ג. (10 נקודות) נגדיר את האלגוריתם הבא לקבלת סדרת פרמטרים $\theta, \theta^2, \dots, \theta^t$:

- אתחלו את θ^0 לערך כלשהו

- חשבו $q^{t+1} = \arg \max_q F(\theta^t, q)$

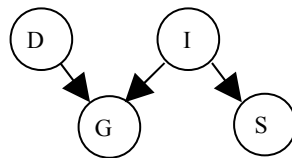
- חשבו $\theta^{t+1} = \arg \max_{\theta} F(\theta, q^{t+1})$

הראו שזו סדרת הפרמטרים שהיה מקבל אלגוריתם EM עם אותו אתחול (השתמשו בסעיף ב.).

5. א. (11 נקודות) הראו שאם פילוג $p(x)$ מקיים את תכונות חוסר התלות המרקוביות הלוקאליות בגרף
מכונן ללא מעגלים G אזי $p(x)$ ניתנת לתיאור ע"י רשת בייסיאנית על G . מהם הפרמטרים של רשת זו?

ב. (12 נקודות) הוכיחו שגם ההיפך מתקיים (כלומר רשת בייסיאנית מקיימת את תכונות חוסר התלות המרקוביות הלוקאליות)

ג. (7 נקודות) נניח כי פילוג $p(x)$ מתפרק לפי הגרף שבצירוף. ציינו את כל תכונות חוסר התלות אשר בהכרח מתקיימות ב p והן מהצורה $W \perp Y|Z$, כאשר W, Y, Z משתנים בודדים או הקבוצה הריקה.



ד. (3 נקודות) האם ייתכן פילוג שמתפרק לפי הגרף בסעיף הקודם ומקיים תכונות שאינן ברשימה שנתתם? הסבירו.