קורס: למידה חישובית מתקדמת

מרצה: דר' שי פיין

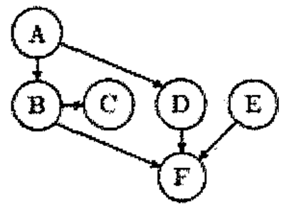
משך הבחינה: 3 שעות

המבחן בחומר סגור (<u>ללא</u> ספרים ומחברות)

בחינה סופית מועד א

ענה/י על 4 מתוך 5 השאלות הבאות (25 נקודות כל אחת)

1. נתונה הרשת הבייסיאנית הבאה



a. כתוב את הפילוג המשותף $P(A,B,C,D,E,F)$ תוך שימוש באי התלויות המותנות (conditional independencies) שהרשת מייצגת

b. בהינתן שכל המשתנים ברשת הבייסיאנית A,B,C,D,E,F הם משתנים בוליאניים, כלומר יכולים לקבל רק שני ערכים: "yes" ו- "no". מה הוא מספר הפרמטרים המינימלי הנדרש כדי להגדיר במפורש את הרשת הבייסיאנית שהמבנה שלה נתון למעלה?

- רמז: זכור שלמשל $P(E = yes) = 1 - P(E = no)$

c. מה הוא מספר הפרמטרים המינימלי הנדרש כדי לבטא את הפילוג המשותף של A,B,C,D,E,F אם לא ניתן היה להניח את אי התלויות המותנות שהרשת מייצגת?

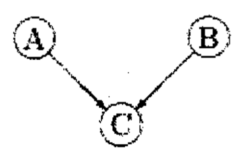
d. מה מהטענות הבאות נכונה עבור הרשת הבייסיאנית הנתונה?
הראה מסלול אקטיבי (הטענה שקרית) או שכל המסלולים חסומים (הטענה נכונה)

- $D \perp\!\!\!\perp E \mid \emptyset$
- $C \perp\!\!\!\perp E \mid F$
- $C \perp\!\!\!\perp E \mid A$
- $A \perp\!\!\!\perp F \mid \{B,D\}$

שים לב: "$X \perp\!\!\!\perp Y \mid Z$" מייצג את הטענה "$X$ ו- Y הם d-separated ע"י Z"

e. נתון הפילוג המשותף הבא מעל שלושת המשתנים האקראיים A,B,C

p_{ABC}	$A = a_1$		$A = a_2$	
	$B = b_1$	$B = b_2$	$B = b_1$	$B = b_2$
$C = c_1$	$3/32$	$1/4$	$1/40$	$3/32$
$C = c_2$	$1/32$	$1/8$	$1/10$	$9/32$



האם הגרף מימין לטבלה לעיל מייצג את הפילוג המשותף המתואר בטבלה? מדוע?

2. מידול סדרות

a. מודלים מרקוביים תבויים (HMM)

i. הסבר את המוטיבציה למעבר משרשרת מרקובית גלויה ל- HMM

ii. מהם הפרמטרים של מודל HMM ומה הם מייצגים?

iii. פרוצדורת forward הוא אלגוריתם הסקה (inference) יעיל בעל סיבוכיות חישובית $O(N^2T)$, כאשר N הוא מספר המצבים ו- T אורך סדרת התצפיות. הסבר כיצד האלגוריתם הזה פועל ומשיג סיבוכיות פולינומית, כלומר לא סיבוכיות אקספוננציאלית $O(N^TT)$

b. שני ניסויי הטלות מטבע סיפקו את סדרות התצפיות הבאות

ניסוי 1: $O_1 = (H\ H\ H\ H\ T\ H\ T\ T\ T\ T)$

ניסוי 2: $O_2 = (H\ T\ T\ H\ T\ H\ H\ T\ T\ H)$

i. בהנחה שסדרת התצפיות בכל ניסוי היא תוצאה של הטלות מטבע בלתי תלויות (כלומר i.i.d)

- עבור כל ניסוי, כתוב את הפילוג של המטבע (כלומר $P(H), P(T)$), וחשב את הנראות (likelihood) של סדרת התצפיות (כלומר O_1, O_2)
- האם נעשה שימוש באותו מטבע בשני הניסויים?

ii. בהינתן HMM עם שלושה מצבים, כאשר הסתברויות הפלט בכל מצב

	State 1	State 2	State 3
$P(H)$	0.5	0.75	0.25
$P(T)$	0.5	0.25	0.75

כל הסתברויות המעבר הם 1/3 וכל הסתברויות התחלת-מצב הם 1/3

- עבור כל ניסוי ענה על השאלות הבאות
  a. מהי סדרת המצבים המסתברת ביותר?
  b. מה ההתפלגות המשותפת של סדרת התצפיות וסדרת המצבים המסתברת ביותר?
  c. מה ההסתברות שסדרת המצבים נוצרה במלואה ע"י "State 1"?
- האם ה- HMM הנתון הוא מודל טוב עבור שני הניסויים, או האם תעדיף להשתמש במודל HMM שונה בכל ניסוי? הסבר

3. אנטרופיה ודחיסה

a. נתון משתנה אקראי X שהאנטרופיה שלו היא $H(X)$, ונתון $Y(X)$ פונקציה דטרמיניסטית של X

i. מהי האנטרופיה $H(Y)$?

ii. מהי האנטרופיה המותנה $H(Y|X)$?

iii. מהי האנטרופיה המותנה $H(X|Y)$?

iv. מהי האנטרופיה המשותפת $H(X,Y)$?

b. הנח ש- X הוא משתנה אקראי קטיגוריאלי בעל 5 ערכים אפשריים $\{\alpha,\beta,\gamma,\delta,\epsilon\}$. בהינתן שני פילוגים $p(x)$ ו- $q(x)$, ושתי סכימות קידוד $C_1(x)$ ו- $C_2(x)$

Symbol	$p(x)$	$q(x)$	$C_1(x)$	$C_2(x)$
α	1/2	1/2	0	0
β	1/4	1/8	10	100
γ	1/8	1/8	110	101
δ	1/16	1/8	1110	110
ϵ	1/16	1/8	1111	111

i. חשב את $H(p), H(q)$ ואת האנטרופיות היחסיות (KL divergence) $D(p||q)$ ו- $D(q||p)$

ii. הראה שאורך מילת הקוד הממוצעת בקידוד C_1 תחת הפילוג p שווה לאנטרופיה $H(p)$, ולפיכך C_1 הוא הקידוד האופטימלי תחת הפילוג p. הראה ש- C_2 הוא הקידוד האופטימלי תחת הפילוג q.

iii. אם נשתמש בקידוד C_2 כאשר ההסתברות היא p, בכמה תגדל אורך מילת הקוד הממוצעת מעבר לאנטרופיה $H(p)$? השתמש ב- $D(p||q)$

iv. אם נשתמש בקידוד C_1 כאשר ההסתברות היא q, בכמה תגדל אורך מילת הקוד הממוצעת מעבר לאנטרופיה $H(q)$? השתמש ב- $D(q||p)$

4. מקסימום אנטרופיה (MaxEnt)
   a. מהו הפילוג שממקסם אנטרופיה (ללא אילוצים)? הוכח
   b. בהינתן
      i. דוגמאות $|\mathcal{X}| = N$,$x_1, x_2, \dots, x_m \in \mathcal{X}$ שנדגמו i.i.d מפילוג לא ידוע
      ii. n פיצ'רים $f_1, f_2, \dots, f_n$ כאשר $f_j: \mathcal{X} \to \mathbb{R}$
      iii. ממוצעים אמפיריים של הפיצ'רים $\hat{E}[f_j] = \frac{1}{m}\sum_i f_j(x_i)\ ,\forall j$

   הראה שפילוג גיבס (Gibbs distribution) ממקסמת את האנטרופיה בכפוף לאילוצי תוחלת על הפיצ'רים, $E_p[f_j] = \hat{E}[f_j],\quad \forall j$

   (שים לב: פילוג גיבס הוא מהצורה $q(x) = \frac{1}{Z}\exp\left(\sum_j \lambda_j f_j(x)\right)$ כאשר $\lambda_j \in \mathbb{R}$ ו- Z הוא גורם נורמליזיציה)

   c. מהו הפילוג שממקסם את האנטרופיה ומקיים את האילוצים $E_p[X] = 0$ ו- $E_p[X^2] = \sigma^2$?

5. דחיסה לא משמרת ועיקרון צוואר הבקבוק של האינפורמציה
   a. מה זה דחיסה לא משמרת (lossy compression)? מה ההבדל בין דחיסה משמרת (lossless) ודחיסה לא משמרת (lossy)?
   b. הגדר את בעיית האופטימיזציה של קצב העיוות (Rate Distortion)
      i. כתוב (ללא הוכחה) את הפתרון האופטימלי ותאר את אלגוריתם Blahut-Arimoto שמשיג פתרון זה
         - האם האלגוריתם יתכנס תמיד לפתרון הגלובלי?
      ii. הסבר את האיזון (tradeoff) בין דחיסה (compression) ועיוות (distortion)
         - השתמש בהסבר שלך ב- Rate Distortion curve וכופל לגראנז' β
   c. הגדר את עיקרון צוואר הבקבוק של האינפורמציה (Information Bottleneck (IB))
      i. מה החסרונות ב- Rate Distortion ש- IB מנסה לטפל?
      ii. הגדר את בעיית האופטימיזציה של IB, כתוב (ללא הוכחה) את הפתרון, ונתח את תכונותיו
         - השתמש ב- Relevance Compression curve כדי להסביר את האיזון (tradeoff) בין דחיסה (compression) ואינפורמציה רלוונטית (relevant information)
         - האם האלגוריתם iterative optimization יתכנס לפתרון אופטימלי?