להגשה עד:
18/05/2021

חישוביות וקוגניציה
מרצה: ד"ר אורן שריקי, מתרגל: אביב דותן

דף תרגיל 5

1. (30 נק') נתונה רשת בעלת נוירון קלט יחיד ונוירון פלט יחיד לא לינארי, כמתואר בתרשים. הקלטים מגיעים מהתפלגות יונימודאלית עם ממוצע 0 ושונות σ^2. נוירון הפלט הוא נוירון סיגמואידלי ופונקציית התמסורת שלו היא הפונקציה הלוגיסטית:

$$s = g(h) = \frac{1}{1+e^{-h}}, \quad h = wx$$

הרשת לומדת למידת אונליין באמצעות אלגוריתם למידה לא מפוקחת המבצע מקסימיזציה לאינפורמציה המשותפת בין הקלט והפלט.

א. השתמשו בכלל הלמידה שנלמד בכיתה לגבי רשתות מסוג זה עם קלט ופלט N-ממדיים (שבהן w היא מטריצה):

$$\Delta w = \eta[(w^T)^{-1} + yx^T], \quad y = \frac{g"(h_l)}{g'(h_l)}$$

ורשמו אלגוריתם למידה אונליין למשקל w.
פשטו את הביטוי שקיבלתם ב- א' על-ידי שימוש בעובדה שהפונקציה הלוגיסטית מקיימת $g''(h) = g'(h)(1-2g(h))$.

ב. השתמשו בניתוח התכנסות בממוצע, וקבלו משוואה שפתרונה הוא הערך שאליו צפוי המשקל w להתכנס (אין צורך לקבל ביטוי מפורש ל-w).

ג. הראו כי כאשר ערכי הקלט x קטנים יחסית, המשקל גדל ביחס הפוך לסטיית התקן של הקלט, והסבירו את משמעות התוצאה.
הדרכה: ניתן להשתמש בקרובים הבאים:
  i. $e^{-u} \cong 1-u, \quad u \ll 1$
  ii. $\frac{1}{1-u} \cong 1+u, \quad u \ll 1$

ד. נניח כעת כי ממוצע הקלטים הוא $\mu \neq 0$ וכי מוסיפים איבר הטיה (bias) לחישוב הפלט, כלומר $h = wx + w_0$. ניתן להראות כי כלל הלמידה ל – w_0 הוא:

$$\Delta w_0 = \eta\big(1 - 2g(h)\big)$$

מהו הערך שאליו צפוי w_0 להתכנס? בטאו את תשובתכם בעזרת הגדלים w ו-μ.
הסבירו את משמעות התוצאה. הדרכה: קבעו תחילה בעזרת שיקולים של התכנסות בממוצע מה אמור להיות הפלט הממוצע והסיקו מה יהיה הערך הממוצע של h.

2. (40 נק') בשאלה זו עליכם להשתמש ב-ICA כדי להפריד תערובות של קבצי קול (עם או בלי רעש) לשלושה מקורות שונים. את קבצי הקול תוכלו להוריד מאתר הקורס.

א. ממשו אלגוריתם infomax ICA, והפרידו **שלוש** תערובות **ללא רעש** למקורות השונים.

ב. חזרו על סעיף א', אך הפעם השתמשו ב**שלוש** תערובות **עם רעש**. האם הרשת הצליחה להפריד את המקורות? אם לא – מדוע?

ג. כעת, השתמשו ב**שבע** תערובות שונות **עם רעש**.
  i. השתמשו בכלל Sanger כדי להוריד את המימד לשלוש תערובות בלבד.
  ii. השתמשו באלגוריתם PCA מוכן (למשתמשי Matlab – פונקציית pca), והשוו את הוקטורים העצמיים לאלו שקיבלתם בתת-הסעיף הקודם. אם קיבלתם וקטורים שונים, הסבירו מדוע.
  iii. השתמשו באלגוריתם infomax ICA להפרדת הפלט של אלגוריתם Sanger למקורות השונים. השוו לתוצאותיכם מסעיפים א' ו-ב', והסבירו את ההבדלים.

<u>הבהרה:</u> בכל סעיף, יש להגיש את קבצי הקול של התוצאות (למשתמשי Matlab – כדאי להכיר את הפונקציות `audioread` ו-`audiowrite`), וכן את מטריצת מקדמי הקורלציות בין התוצאות למקורות האמיתיים (למשתמשי Matlab – פונקציית `corr`). את הקוד יש להגיש בנפרד.

3. (30 נק') נתונים P קלטים ממימד n: $\forall 1 \le \mu \le P$, $\vec{x}^{\mu} \in \mathbb{R}^n$ המתפלגים לפי התפלגות גאוסיאנית רב-מימדית: $X \sim \mathcal{N}(\vec{\mu}, C)$:

$$f(\vec{x}) = \frac{1}{\sqrt{(2\pi)^n \det C}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T C^{-1}(\vec{x}-\vec{\mu})}$$

הראו שלאחר PCA יתקבלו ערוצים $\vec{y}$ בלתי תלויים סטטיסטית:

א. השתמשו במשפט הלכסון האורתוגונלי (תרגיל 1) והראו כי:

$$C^{-1} = QD^{-1}Q^T$$

כאשר Q היא מטריצה אורתוגונלית שעמודותיה הן הוקטורים העצמיים של C ו-D^{-1} היא מטריצה אלכסונית שערכי האלכסון שלה הופכיים לערכים העצמיים של C.

ב. היעזרו בסעיף א', ובצעו החלפת משתנה בפונקציית צפיפות ההסתברות של $\vec{x}$ כדי לקבל את פונקציית צפיפות ההסתברות של $\vec{y}$. (לא לשכוח את היעקוביאן!)

ג. הראו כי ערוצי הפלט בלתי תלויים סטטיסטית:

$$\Pr\left(\bigwedge_i (a_i < y_i < b_i)\right) = \prod_{i=1}^{n} \Pr(a_i < y_i < b_i)$$

או, באופן שקול:

$$\int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_n}^{b_n} f(\vec{y})\overrightarrow{dy} = \prod_{i=1}^{n} \int_{a_i}^{b_i} f_i(y_i) dy_i$$