תאריך העבודה: 18.7.2022
שם המרצה: ד"ר קובי טודרוס
שם הקורס: נושאים באנליזה סטטיסטית מרובת משתנים
מספר הקורס: 361.2.2130
שנה: 2022, סמסטר: ב'
משך העבודה: שמונה ימים

אוניברסיטת בן-גוריון בנגב
Ben-Gurion University of the Negev

עבודה בקורס נושאים באנליזה סטטיסטית מרובת משתנים

יש לענות באופן מפורט על כל השאלות. תשובות לא מנומקות לא תקבלנה ניקוד.

בהצלחה !!!

<u>שאלה 1: הרחבה אורתונורמלית של וקטור אקראי (15 נק')</u>

נתון וקטור אקראי $\mathbf{x}\in\mathbb{R}^{N}$ המקיים את המודל הבא:

$$\mathbf{x}=\mathbf{A}\mathbf{s}+\mathbf{w}$$

המטריצה $\mathbf{A}\in\mathbb{R}^{N\times N}$ היא מטריצה דטרמיניסטית לא-סינגולרית המקיימת את הפירוק הבא:

$$\mathbf{A}=\mathbf{B}\mathbf{C}\mathbf{F}$$

כאשר $\mathbf{B}\triangleq\left[\mathbf{b}_{1},\ldots\mathbf{b}_{N}\right]$ ו-$\mathbf{F}\triangleq\left[\mathbf{f}_{1},\ldots\mathbf{f}_{N}\right]$ הן מטריצות יוניטריות אשר מורכבות מסט וקטורי בסיס אורתונורמליים הפורשים את $\mathbb{R}^{N}$ ו-$\mathbf{C}\triangleq\begin{bmatrix}c_{1}&\cdots&0\\\vdots&\ddots&\vdots\\0&\cdots&c_{N}\end{bmatrix}$ היא מטריצה דיאגונלית עם ערכים ממשיים על האלכסון הראשי כך שמתקיים $|c_{1}|>\ldots>|c_{N}|$. הוקטורים $\mathbf{s}\in\mathbb{R}^{N}$ ו-$\mathbf{w}\in\mathbb{R}^{N}$ הם וקטורים אקראיים חסרי קורלציה עם תוחלת $\mathbf{0}$ ומטריצות קווריאנס $\Sigma_{\mathbf{s}}=\sigma_{\mathbf{s}}^{2}\mathbf{I}_{N}$ ו-$\Sigma_{\mathbf{w}}=\sigma_{\mathbf{w}}^{2}\mathbf{I}_{N}$, בהתאמה, כאשר $\mathbf{I}_{N}\in\mathbb{R}^{N\times N}$ מסמנת מטריצת יחידה.

א. (3 נק') מצאו באופן מפורש את וקטורי הבסיס של הרחבת Karhunen-Loeve.

ב. (4 נק') מצאו באופן מפורש את הוריאנסים של מקדמי הרחבת Karhunen-Loeve.

ג. (8 נק') נתון וקטור אקראי $\mathbf{y}\triangleq T\left[\mathbf{x}\right]\in\mathbb{R}^{M}$, $M<N$, המתקבל באמצעות אופרטור דחיסה רועש $T\left[\mathbf{x}\right]\triangleq\mathbf{H}^{T}\mathbf{x}+\mathbf{z}$, כאשר $\mathbf{H}\triangleq\left[\mathbf{h}_{1},\ldots,\mathbf{h}_{M}\right]\in\mathbb{R}^{N\times M}$ היא מטריצה המכילה M וקטורי

בסיס אורתונורמליים הפורשים תת-מרחב לינארי של $\mathbb{R}^N$ ו- $\mathbf{z} \in \mathbb{R}^M$ הוא וקטור אקראי עם תוחלת $\mathbf{0}$ ומטריצת קווריאנס $\Sigma_{\mathbf{z}} = \sigma_{\mathbf{z}}^2 \mathbf{I}_M$. מניחים כי $\mathbf{x}$ ו- $\mathbf{z}$ הם חסרי קורלציה. מצאו באופן מפורש את המטריצה $\mathbf{H}$ שעבורה אנרגית הוקטור הדחוס $E\left[\|\mathbf{y}\|^2\right]$ תהיה מקסימאלית. מצאו באופן מפורש את האנרגיה המתקבלת עבור מטריצה זו.

שאלה 2: Canonical Correlation Analysis (15 נק')

יהיו $\mathbf{x} \triangleq \left[x_1, x_2, \ldots, x_p\right]^T$ ו- $\mathbf{y} \triangleq \left[y_1, y_2, \ldots, y_q\right]^T$, $p \le q$, וקטורים אקראיים ממשיים עם מטריצות קווריאנס לא-סינגולריות $\Sigma_{\mathbf{x}}$ ו- $\Sigma_{\mathbf{y}}$, בהתאמה. מטריצת הקרוס-קווריאנס נתונה לפי $\Sigma_{\mathbf{xy}} = \mathbf{g}\mathbf{h}^T$, כאשר $\mathbf{g} \in \mathbb{R}^p$ ו- $\mathbf{h} \in \mathbb{R}^q$. בנוסף, נגדיר את הוקטורים האקראיים $\mathbf{w} \triangleq \mathbf{G}\mathbf{x} + \mathbf{c}$ ו- $\mathbf{z} \triangleq \mathbf{H}\mathbf{y} + \mathbf{d}$, כאשר המטריצה $\mathbf{G} \in \mathbb{R}^{p \times p}$ הינה מטריצה דטרמיניסטית הפיכה. המטריצה $\mathbf{H} \in \mathbb{R}^{m \times q}$, $p \le m \le q$ הינה מטריצה דטרמיניסטית כך שלמטריצת הקווריאנס $\Sigma_{\mathbf{z}}$ קיימת הופכית. הוקטורים $\mathbf{c} \in \mathbb{R}^{p \times 1}$ ו- $\mathbf{d} \in \mathbb{R}^{m \times 1}$ הם וקטורים דטרמיניסטיים.

א. (5 נק') נסמן ב- $\rho_1, \ldots, \rho_p$ את מקדמי הקורלציה הקנוניים וב- $\left\{\left(\mathbf{a}_k, \mathbf{b}_k\right)\right\}_{k=1}^{p}$ את כיווני הקורלציה הקנוניים של $(\mathbf{x}, \mathbf{y})$. כתבו ביטויים מפורשים בתלות ב- $\Sigma_{\mathbf{x}}$, $\Sigma_{\mathbf{y}}$, $\mathbf{g}$ ו- $\mathbf{h}$ עבור $\rho_1, \ldots, \rho_r$ ו- $\left\{\left(\mathbf{a}_k, \mathbf{b}_k\right)\right\}_{k=1}^{r}$, כאשר r הוא מספר מקדמי הקורלציה הקנוניים ששונים מאפס.

ב. (5 נק') נסמן ב- $\rho'_1, \ldots, \rho'_p$ את מקדמי הקורלציה הקנוניים וב- $\left\{\left(\mathbf{a}'_k, \mathbf{b}'_k\right)\right\}_{k=1}^{p}$ את כיווני הקורלציה הקנוניים של $(\mathbf{w}, \mathbf{z})$. כתבו ביטויים מפורשים בתלות ב- $\Sigma_{\mathbf{y}}$, $\mathbf{h}$, $\mathbf{H}$, $\mathbf{G}$, $\rho_1, \ldots, \rho_p$ ו- $\left\{\left(\mathbf{a}_k, \mathbf{b}_k\right)\right\}_{k=1}^{p}$ עבור $\rho'_1, \ldots, \rho'_j$ ו- $\left\{\left(\mathbf{a}'_k, \mathbf{b}'_k\right)\right\}_{k=1}^{j}$, כאשר j הוא מספר מקדמי הקורלציה הקנוניים ששונים מאפס.

ג. (5 נק') מצאו תנאי על $\mathbf{H}$ ו- m (מספר השורות של $\mathbf{H}$) שעבורו מקדמי הקורלציה הקנוניים שנמצאו בסעיפים א' ו-ב' מתלכדים.

<u>שאלה 3: אלגוריתם ה-EM (20 נק׳)</u>

תהי $\mathbf{x}^N \triangleq \left\{\mathbf{x}_n \in \mathbb{R}^p\right\}_{n=1}^N$ סדרת מדידות/תצפיות המקיימת את המודל הבא:

$$\mathbf{x}_n = \left(1-u_n\right)\mathbf{s}_n + u_n\mathbf{w}_n$$

כאשר $u^N \triangleq \left\{u_n \in \mathbb{R}\right\}_{n=1}^N$, $\mathbf{s}^N \triangleq \left\{\mathbf{s}_n \in \mathbb{R}^p\right\}_{n=1}^N$ ו- $\mathbf{w}^N \triangleq \left\{\mathbf{w}_n \in \mathbb{R}^p\right\}_{n=1}^N$ הן סדרות אקראיות חבויות (בלתי נצפות). נתון כי כל אחת מהסדרות u^N, $\mathbf{s}^N$ ו- $\mathbf{w}^N$ היא סדרה אקראית i.i.d. כך ש- u_n הוא משתנה אקראי בינארי המקבל ערך של 0 או 1 ומתקיים $\Pr\left[u=1\right]=\varepsilon$, $\mathbf{s}_n$ הוא וקטור אקראי גאוסי עם תוחלת $\mathbf{m}$ ומטריצת קווריאנס $\mathbf{C}$ לא ידועים ו- $\mathbf{w}_n$ הוא וקטור אקראי עם פונקצית צפיפות פילוג ידועה $f_{\mathbf{w}}\left(\cdot\right)$. בנוסף, נתון כי הסדרות u^N, $\mathbf{s}^N$ ו- $\mathbf{c}^N$ הן בת״ס במשותף. פתחו את משוואות העדכון של אלגוריתם ה-EM לשיערוך ε, $\mathbf{m}$ ו- $\mathbf{C}$ מתוך סדרת התצפיות $\mathbf{x}^N$. יש לרשום משוואות עדכון מפורשות ללא תוחלות.

שאלה 4: Multivariate Kernel Density Estimation (15 נק')

נתונה סדרה $\left\{\mathbf{x}_n\right\}_{n=1}^{N}$ של דגימות i.i.d. מהפילוג של וקטור אקראי $\mathbf{x}\in\mathbb{R}^p$ עם פונקצית צפיפות פילוג $f_{\mathbf{x}}(\cdot)$ חסומה וגזירה ברציפות מעל $\mathbb{R}^p$. נגדיר את המשערך של $f_{\mathbf{x}}(\mathbf{r})$, $\mathbf{r}\in\mathbb{R}^p$, כדלקמן:

$$\hat{f}_{\mathbf{x}}(\mathbf{r})\triangleq\frac{1}{Nw_N^p}\sum_{n=1}^{N}K\left(\frac{\mathbf{r}-\mathbf{x}_n}{w_N}\right)$$

כאשר $K(\cdot)$ הינה פונקצית גרעין המקיימת את התכונות הבאות:

1. $\int_{\mathbb{R}^p}K(\mathbf{r})d\mathbf{r}=1$.

2. $\int_{\mathbb{R}^p}K^2(\mathbf{r})d\mathbf{r}<\infty$.

הפרמטר $w_N>0$ הוא פרמטר scaling אשר תלוי במספר הדגימות N.

א. (3 נק') מצאו האם לכל N סופי מתקיים $E\left[\left|\hat{f}_{\mathbf{x}}(\mathbf{r})-f_{\mathbf{x}}(\mathbf{r})\right|^2\right]<\infty$.

ב. (6 נק') מצאו תנאי מספיק על w_N שעבורו לכל $\varepsilon>0$ ולכל $\mathbf{r}\in\mathbb{R}^p$ מתקיים:

$$\Pr\left[\left|\hat{f}_{\mathbf{x}}(\mathbf{r})-f_{\mathbf{x}}(\mathbf{r})\right|>\varepsilon\right]\xrightarrow[N\to\infty]{}0$$

ג. (6 נק') ידוע כי פרמטר רוחב החלון האופטימלי במובן AMISE נתון לפי:

$$w_{\mathrm{o},N}=\left(\frac{p\int_{\mathbb{R}^p}K^2(\mathbf{r})d\mathbf{r}}{N\left(\int_{\mathbb{R}^p}\left(\mathbf{e}_1^T\mathbf{r}\right)^2K(\mathbf{r})d\mathbf{r}\right)^2\int_{\mathbb{R}^p}\mathrm{tr}^2\left[\nabla^2f_{\mathbf{x}}(\mathbf{r})\right]d\mathbf{r}}\right)^{\frac{1}{p+4}}$$

כאשר $\mathbf{e}_1\triangleq[1,0,\ldots,0]^T$ הוא וקטור היחידה הראשון. מצאו האם עבור $w_{\mathrm{o},N}$ זה מתקיים:

$$\Pr\left[\left|\hat{f}_{\mathbf{x}}(\mathbf{r})-f_{\mathbf{x}}(\mathbf{r})\right|>\varepsilon\right]\xrightarrow[N\to\infty]{}0$$

לכל $\varepsilon>0$ ולכל $\mathbf{r}\in\mathbb{R}^p$?

שאלה 5: גילוי אותות ברעש גאוסי באמצעות Mardia's tests (15 נק')

נתונה בעיית ההחלטה הבאה עבור גילוי אות אקראי ברעש גאוסי:

$$\boxed{\begin{array}{ll} H_0: & \mathbf{x}_n = \mathbf{w}_n, \quad n = 1,\ldots,N \quad \text{(signal does not exist)} \\ H_1: & \mathbf{x}_n = \mathbf{s}_n + \mathbf{w}_n, \quad n = 1,\ldots,N \quad \text{(signal exists)} \end{array}}$$

כאשר $\left\{\mathbf{x}_n \in \mathbb{R}^p\right\}$ היא סדרת תצפיות אקראיות. התהליך $\left\{\mathbf{s}_n \in \mathbb{R}^p\right\}$ הוא תהליך סיגנל אקראי i.i.d. בלתי-נצפה עם פילוג לא-גאוסי סימטרי מסביב לראשית. התהליך $\left\{\mathbf{w}_n \in \mathbb{R}^p\right\}$ הוא תהליך רעש i.i.d. בלתי-נצפה עם פילוג גאוסי. מניחים כי:

1. התהליכים $\left\{\mathbf{s}_n \in \mathbb{R}^p\right\}$ ו-$\left\{\mathbf{w}_n \in \mathbb{R}^p\right\}$ הם בת"ס.
2. איברי וקטור הסיגנל $\mathbf{s} \triangleq \left[s_1,\ldots,s_p\right]^T$ הם בת"ס במשותף עם וריאנס זהה $\sigma_{\mathbf{s}}^2$.
3. איברי וקטור הרעש $\mathbf{w} \triangleq \left[w_1,\ldots,w_p\right]^T$ הם חסרי קורלציה עם תוחלת 0 ווריאנס זהה $\sigma_{\mathbf{w}}^2$.
4. תחת שתי ההיפותזות מתקיים $E\left[\left\|\mathbf{x}\right\|^8\right] < \infty$.

מעוניינים לגלות את האות באמצעות מבחני הנורמליות של Mardia.

א. (2 נק') נתון כי עבור סף החלטה t_α ומספר סופי N_α של תצפיות, הסתברות גילוי השווא (PFA) של Mardia's-skewness-test היא α. מצאו את ה- PFA עבור t_α ו-N_α כאשר תהליך הרעש מוכפל בפקטור c.

ב. (2 נק') נתון כי עבור סף החלטה t_β ומספר סופי N_β של תצפיות, הסתברות הגילוי (PD) של Mardia's-Kurtotis-test היא β. מצאו את ה- PD עבור t_β ו-N_β כאשר תהליכי הסיגנל והרעש מוכפלים בפקטור c.

ג. (3 נק') מצאו את סף ההחלטה t שעבורו תתקבל PFA אסימפטוטית ($N \to \infty$) של 0.01 כאשר משתמשים ב- Mardia's skewness test ו-$p = 10$.

ד. (3 נק') מצאו את סף ההחלטה t שעבורו תתקבל PFA אסימפטוטית ($N \to \infty$) של 0.01 כאשר משתמשים ב-Mardia's kurtosis test ו-$p = 10$.

ה. (5 נק') מצאו תנאי מספיק על היחס $r \triangleq \frac{1}{p\sigma_{\mathbf{s}}^{4}}\sum_{k=1}^{p}\mathrm{E}\left[s_k^4\right]$ שעבורו Mardia’s kurtosis test הוא קונסיסטנטי.

<u>שאלה 6: מבחן לחוסר תלות-סטטיסטית (20 נק')</u>

יהי $(\mathbf{x},\mathbf{y})^N \triangleq \left\{(\mathbf{x}_n,\mathbf{y}_n)\right\}_{n=1}^N$ סט של דגימות i.i.d מהפילוג המשותף $P_{\mathbf{x},\mathbf{y}}$ של שני וקטורים אקראיים גאוסיים במשותף $\mathbf{x}\in\mathbb{R}^p$ ו- $\mathbf{y}\in\mathbb{R}^q$, $p\le q$. מעוניינים לבחון את ההשערה ש- $\mathbf{x}$ ו- $\mathbf{y}$ הם בת"ס (היפותזה H_0) אל מול ההשערה הנגדית (היפותזה H_1). לשם כך, מגדירים את ה-test-statistic הבא $T\left[(\mathbf{x},\mathbf{y})^N\right] \triangleq \eta\sum_{k=1}^{p}\log\left(1-\hat{\rho}_k^2\right)$, כאשר $\eta \triangleq \frac{p+q+3}{2}-N$, ו- $\left\{\hat{\rho}_k\right\}_{k=1}^{p}$ הם שיערוכים של מקדמי הקורלציה הקנוניים, אשר מתקבלים באמצעות משערכי ה-Maximum-Likelihood של הקוואריאנסים והקרוס-קוואריאנס. נסמן משערכים אלה באמצעות $\hat{\Sigma}_{\mathbf{x}}$, $\hat{\Sigma}_{\mathbf{y}}$ ו- $\hat{\Sigma}_{\mathbf{xy}}$ ונניח כי $\hat{\Sigma}_{\mathbf{x}}$ ו- $\hat{\Sigma}_{\mathbf{y}}$ הפיכות. בחינת ההשערות מתבצעת באמצעות השוואת $T\left[(\mathbf{x},\mathbf{y})^N\right]$ לסף $t \triangleq c\times p\times q$, כאשר c הוא קבוע חיובי גדול מ-1. אם $T\left[(\mathbf{x},\mathbf{y})^N\right]\le t$, מחליטים H_0 וההיפך עבור H_1 . מצאו ערך מספרי של c שעבורו תחת היפותיזה H_0 תמיד מתקיים:

$$\lim_{N\to\infty}\Pr\left[T\left[(\mathbf{x},\mathbf{y})^N\right]>t\right]\le 0.01$$

נוסחאות עזר

:Markov אי-שיוויון

יהי X משתנה אקראי אי-שלילי עם תוחלת סופית ויהי α סקלר חיובי. אזי מתקיים:

$$\boxed{\Pr\left[X \geq \alpha\right] \leq \frac{\mathrm{E}\left[X\right]}{\alpha}}$$

משפט עזר לשאלה 4:

תהי $v(\mathbf{x})$ פונקציה אינטגרבילית מעל $\mathbb{R}^p$ כך שמתקיים $\int_{\mathbb{R}^p} v(\mathbf{r})d\mathbf{r} = a$. נגדיר $v_t(\mathbf{r}) \triangleq \frac{1}{t^p} v\left(\frac{\mathbf{r}}{t}\right)$

כאשר $t > 0$. אם $g(\mathbf{r})$ היא פונקציה חסומה וגזירה ברציפות מעל $\mathbb{R}^p$ אזי לכל $\mathbf{r} \in \mathbb{R}^p$ מתקיים:

$$\lim_{t \to 0} (v_t * g)(\mathbf{r}) = a g(\mathbf{r})$$

כאשר ההתכנסות היא יוניפורמית ב- $\mathbb{R}^p$ ו- $(v_t * g)(\mathbf{r}) \triangleq \int_{\mathbb{R}^p} v_t(\mathbf{r}-\mathbf{y}) g(\mathbf{y}) d\mathbf{y}$.

מומנט רביעי של משתנה אקראי גאוסי עם תוחלת 0 ווריאנס σ^2 :

$$E\left[X^4\right] = 3\sigma^4$$

תוחלת של משתנה אקראי מפולג chi-squared עם p דרגות חופש:

$$E\left[X\right] = p$$