מבוא למערכות לומדות- תרגיל 1

ישי לביא- תז: 209713031

חלק תאורטי

שאלה 1

בפורום נכתב שיש להוכיח רק למכפלה הפנימית הסטנדרטית. תהי $\langle,\rangle$ המכפלה הפנימית הסטנרטית מעל V ו- $\|\cdot\|$ הנורמה המושרית ממנה, ו- A מטריצה אורתוגונלית המתאימה להעתקה הלינארית $T: V \to W$ מעל הבסיס הסטנדרטי. נשים לב כי: $x = \sum_{i=1}^{n} x_i e_i$, אזי מתקיים מתכונות המכפלה הפנימית הסטנדרטית:

$$\|Ax\|^2 = \langle Ax, Ax \rangle = (Ax)^T \cdot Ax = x^T A^T A x$$

אבל A אורתוגונלית, לכן $A^T A = I$, ומכאן:

$$\|Ax\|^2 = x^T A^T A x = x^T I x = x^T x = \langle x, x \rangle = \|x\|^2$$

מחיוביות הנורמה נוכל להוציא שורש משני האגפים ולקבל את הדרוש.

שאלה 2

נתחיל מלחשב את $A^T A$:

$$A^T A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 2 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 2 \\ 0 & 2 & -2 \\ 2 & -2 & 4 \end{pmatrix}$$

נזכור כי $A^T A = V \Sigma^T \Sigma V^T$, כאשר האגף הימני הוא פירוק ה-EVD של $A^T A$. בשביל למצוא את הפירוק, נלכסן כפי שראינו בתרגול ע"י פתרון המשוואה: $\det(A^T A - tI) = 0$. נציב:

$$det \begin{pmatrix} 2-t & 0 & 2 \\ 0 & 2-t & -2 \\ 2 & -2 & 4-t \end{pmatrix} = (2-t)\big((2-t)(4-t) - 4\big) - 4(2-t) =$$

$$(2-t)(8 - 6t + t^2 - 8) = t(2-t)(t-6)$$

וקיבלנו כי הערכים העצמיים הם 0,2,6, כלומר קיבלנו כי:

$$\Sigma^T \Sigma = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

נמצא את הוקטורים העצמיים כדי למצוא את V, ע"י פתרון המשוואה $(A^T A - \lambda I)v = 0$:

$$(A^T A - 0I)v = 0 \to \begin{pmatrix} 2 & 0 & 2 \\ 0 & 2 & -2 \\ 2 & -2 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0$$

ע"י דירוג מקבלים:

$$\begin{pmatrix} 2 & 0 & 2 \\ 0 & 2 & -2 \\ 2 & -2 & 4 \end{pmatrix} \to \begin{pmatrix} 2 & 0 & 2 \\ 0 & 2 & -2 \\ 2 & 0 & 2 \end{pmatrix} \to \begin{pmatrix} 2 & 0 & 2 \\ 0 & 2 & -2 \\ 0 & 0 & 0 \end{pmatrix}$$

ולכן נקבל את הפתרון:

$$2y-2z=0\rightarrow y=z,\qquad 2x+2z=0\rightarrow -z$$

ולכן פתרון אפשרי יהיה $(1,-1,-1)$. נמשיך:

$$(A^TA-2I)v=0\rightarrow\begin{pmatrix}0&0&2\\0&0&-2\\2&-2&2\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}=0$$

נפתור ונקבל: $2z=0\rightarrow z=0$, ומהשורה השלישית:

$$2x-2y+2z=0\rightarrow x=y$$

לכן פתרון אפשרי יהיה $(1,1,0)$. נמשיך:

$$(A^TA-6I)v=0\rightarrow\begin{pmatrix}-4&0&2\\0&-4&-2\\2&-2&-2\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}=0$$

ע"י פעולות שורה נקבל:

$$\begin{pmatrix}-4&0&2\\0&-4&-2\\2&-2&-2\end{pmatrix}\rightarrow\begin{pmatrix}-4&0&2\\0&-4&-2\\-4&4&4\end{pmatrix}\rightarrow\begin{pmatrix}-4&0&2\\0&-4&-2\\-4&0&2\end{pmatrix}\rightarrow\begin{pmatrix}4&0&-2\\0&4&2\\0&0&0\end{pmatrix}$$

ולכן קיבלנו:

$$4y+2z=0\rightarrow z=-2y,\qquad 4x-2z=0\rightarrow z=2x$$

לכן אם נקבע $z=2$ נקבל פתרון $(1,-1,2)$. ננרמל את הוקטורים ונקבל:

$$V=\begin{pmatrix}\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{6}}&\frac{1}{\sqrt{3}}\\\frac{1}{\sqrt{2}}&\frac{-1}{\sqrt{6}}&-\frac{1}{\sqrt{3}}\\0&\frac{2}{\sqrt{6}}&\frac{-1}{\sqrt{3}}\end{pmatrix}$$

וכן אנו יודעים כי Σ "אלכסונית" ומהצורה 3×2, ואנו יודעים כי $\Sigma^T\Sigma$ מכילה את ריבועי ערכי האלכסון, לכן:

$$\Sigma=\begin{pmatrix}\sqrt{2}&0&0\\0&\sqrt{6}&0\end{pmatrix}$$

כעת נמצא את U ע"י המשוואה $AV=U\Sigma$. אנו יודעים כי U מגודל 2x2, לכן:

$$\begin{pmatrix}1&1&0\\1&-1&2\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{6}}&\frac{1}{\sqrt{3}}\\\frac{1}{\sqrt{2}}&\frac{-1}{\sqrt{6}}&-\frac{1}{\sqrt{3}}\\0&\frac{2}{\sqrt{6}}&\frac{-1}{\sqrt{3}}\end{pmatrix}=\begin{pmatrix}a&b\\c&d\end{pmatrix}\begin{pmatrix}\sqrt{2}&0&0\\0&\sqrt{6}&0\end{pmatrix}\rightarrow$$

$$\begin{pmatrix}\sqrt{2}&0&0\\0&\sqrt{6}&0\end{pmatrix}=\begin{pmatrix}\sqrt{2}a&\sqrt{6}b&0\\\sqrt{2}c&\sqrt{6}d&0\end{pmatrix}$$

ולכן בוודאי $U=\begin{pmatrix}0&1\\1&0\end{pmatrix}$, ומכאן שמצאנו את U,Σ,V כדרוש.

שאלה 3

נתחיל מלמצוא נוסחה כללית ל-b_k כתלות ב-k:

$$b_k = \frac{C_0 b_{k-1}}{\|C_0 b_{k-1}\|} = \frac{\frac{C_0^2 b_{k-2}}{\|C_0 b_{k-2}\|}}{\left\|\frac{C_0^2 b_{k-2}}{\|C_0 b_{k-2}\|}\right\|} = \frac{C_0^2 b_{k-2}}{\|C_0^2 b_{k-2}\|} \dots = \frac{C_0^k b_0}{\|C_0^k b_0\|}$$

ראינו בלינארית כי אם λ ע"ע של מטריצה B עם הוקטור העצמי v, אזי לכל $n \in \mathbb{N}$ המספר λ^n הוא ע"ע של B^n עבור הו"ע v. לכן נקבל כי:

$$C_0^k b_0 = \sum_{i=1}^{n} a_i C_0^k v_i = \sum_{i=1}^{n} a_i \lambda_i^k v_i$$

כמו כן, ו"ע של מטריצה סימטרית הם אורתוגונליים, ונתון לנו כי הם מנורמלים, לכן:

$$\left\|C_0^k b_0\right\|^2 = \left\langle \sum_{i=1}^{n} a_i \lambda_i^k v_i, \sum_{i=1}^{n} a_i \lambda_i^k v_i \right\rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \lambda_i^k \lambda_j^k \langle v_i, v_j\rangle$$

אבל $\{v_1, \dots, v_n\}$ אורתונורמליים, לכן:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \lambda_i^k \lambda_j^k \langle v_i, v_j\rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \lambda_i^k \lambda_j^k \delta_{ij} = \sum_{i=1}^{n} \left(a_i \lambda_i^k\right)^2$$

נזכור כי $\lambda_1 > \lambda_2 \geq \cdots \geq \lambda_n$, וכן כפי שראינו בתרגול, C_0 חיובית למחצה (שכן היא ניתנת להצגה בתור $A^T A$), ולכן כל הע"ע הנ"ל אי שליליים. מכאן (החלוקה חוקית שכן $a_1, \lambda_1 \neq 0$):

$$\frac{\sqrt{\sum_{i=1}^{n}\left(a_i \lambda_i^k\right)^2}}{|a_1| \lambda_1^k} = \sqrt{\sum_{i=1}^{n}\left(\frac{a_i}{|a_1|} \cdot \left(\frac{\lambda_i}{\lambda_1}\right)^k\right)^2}$$

אבל לכל $i > 1$, $\lim_{k\to\infty}\left(\frac{\lambda_i}{\lambda_1}\right)^k = 0$, לכן נקבל:

$$\lim_{k\to\infty}\sqrt{\sum_{i=1}^{n}\left(\frac{a_i}{|a_1|} \cdot \left(\frac{\lambda_i}{\lambda_1}\right)^k\right)^2} = \sqrt{1} = 1$$

ולכן גם $\lim_{k\to\infty}\frac{|a_1|\lambda_1^k v_1}{\sqrt{\sum_{i=1}^{n}\left(a_i\lambda_i^k\right)^2}} = 1$, ולכן $\lim_{k\to\infty}\frac{a_1\lambda_1^k v_1}{\sqrt{\sum_{i=1}^{n}\left(a_i\lambda_i^k\right)^2}} = \pm 1$. כעת עבור $1 < i \leq n$ מתקיים:

$$\frac{a_i \lambda_i^k}{\sqrt{\sum_{i=1}^{n}\left(a_i\lambda_i^k\right)^2}} \leq \frac{a_i \lambda_i^k}{\sqrt{a_1^2 \lambda_1^{2k}}} = \frac{a_i \lambda_i^k}{|a_1|\lambda_1^k} = \frac{a_i}{|a_1|} \cdot \left(\frac{\lambda_i}{\lambda_1}\right)^k \to_{k\to\infty} 0$$

ומכאן שגם $\lim_{k\to\infty}\frac{a_i\lambda_i^k}{\sqrt{\sum_{i=1}^{n}\left(a_i\lambda_i^k\right)^2}} = 0$. סך הכל קיבלנו:

$$\lim_{k\to\infty} b_k = \lim_{k\to\infty}\frac{\sum_{i=1}^{n} a_i\lambda_i^k v_i}{\sqrt{\sum_{i=1}^{n}\left(a_i\lambda_i^k\right)^2}} = \sum_{i=1}^{n}\lim_{k\to\infty}\frac{a_i\lambda_i^k v_i}{\sqrt{\sum_{i=1}^{n}\left(a_i\lambda_i^k\right)^2}} = \pm v_1 + 0 v_2 + \cdots + v_n = \pm v_1$$

שאלה 4

ננסה לפרק את הפעולה לפעולות פשוטות יותר. מכיוון שהפונקציה היא $\mathbb{R}^n \to \mathbb{R}^n$, היעקוביאן הוא מטריצה מגודל $n \times n$. נחשב את הכניסה ה-i,j ביעקוביאן, כלומר את $\frac{\partial f_i(\sigma)}{\partial x_j}$. מהגדרת כפל מטריצות:

$$f(\sigma)_i = \sum_{k,l,p\in[n]} U_{ik} diag(\sigma)_{kl} U_{lp}^T x_p$$

אבל עבור $k \neq l$ נקבל כי $diag(\sigma)_{kl} = 0$, לכן נקבל:

$$f(\sigma)_i = \sum_{l,p\in[n]} U_{il} diag(\sigma)_{ll} U_{lp}^T x_p = \sum_{l,p\in[n]} U_{il} \sigma_l U_{lp}^T x_p$$

כעת בשביל למצוא את הכניסה ה-j, נגזור את הביטוי האחרון לפי σ_j. עבור $l \neq j$ נקבל 0, ועבור $l = j$:

$$\frac{\partial f_i(\sigma)}{\partial \sigma_j} = \sum_{p=1}^{n} U_{ij} U_{jp}^T x_p = \sum_{p=1}^{n} U_{ij} U_{jp}^T x_p = U_{ij} \cdot \sum_{p=1}^{n} U_{jp}^T x_p = U_{ij} \cdot (U^T x)_j$$

כלומר קיבלנו כי:

$$J_\sigma(f) = U \cdot diag(U^T x)$$

שאלה 5

נחשב את הגרדיאנט של הפונקציה $h(\sigma) = \frac{1}{2}\|f(\sigma) - y\|^2$. נגדיר:

$$g: \mathbb{R}^n \to \mathbb{R}^n, \qquad g(\sigma) = f(\sigma) - y$$

כאשר $y \in \mathbb{R}^n$ קבוע, וכן:

$$w: \mathbb{R}^n \to \mathbb{R}, \qquad w(\sigma) = \frac{1}{2}\|\sigma\|^2$$

נציב לב כי $J_\sigma(g) = J_\sigma(f)$ שכן הן נבדלות זו מזו בחיבור בקבוע. נחשב את הפלט של w מפורשות:

$$w(\sigma) = \frac{1}{2}\|\sigma\|^2 = \frac{1}{2}\langle \sigma, \sigma \rangle = \frac{1}{2}\sum_{i=1}^{n} \sigma_i^2$$

ולכן נקבל ע"י גזירה כי:

$$J_\sigma(w) = (\sigma_1, \dots \sigma_n) = \sigma$$

ומכלל השרשרת נקבל:

$$J_\sigma(h) = J_{g(\sigma)}(w) \circ J_\sigma(g) = g(\sigma) \circ J_\sigma(f) = (f(\sigma) - y) \circ \left(U \cdot diag(U^T x)\right)$$

וסיימנו.

<u>שאלה 6</u>

נחשב את הכניסה ה-i,j, כלומר את $\frac{\partial S(x)_i}{\partial x_j}$:

אם $i=j$:

$$\frac{\partial S(\boldsymbol{x})_j}{\partial x_j}=\frac{\partial}{\partial x_j}\frac{e^{x_j}}{\sum_{l=1}^{k}e^{x_l}}=\frac{e^{x_j}\sum_{l=1}^{k}e^{x_l}-e^{x_j}e^{x_j}}{\left(\sum_{l=1}^{k}e^{x_l}\right)^2}=e^{x_j}\cdot\frac{\sum_{l\neq j}^{k}e^{x_l}}{\left(\sum_{l=1}^{k}e^{x_l}\right)^2}$$

אחרת:

$$\frac{\partial S(\boldsymbol{x})_i}{\partial x_j}=\frac{\partial}{\partial x_j}\frac{e^{x_i}}{\sum_{l=1}^{k}e^{x_l}}=\frac{0-e^{x_i}\cdot e^{x_j}}{\left(\sum_{l=1}^{k}e^{x_l}\right)^2}=-\frac{e^{x_i+x_j}}{\left(\sum_{l=1}^{k}e^{x_l}\right)^2}$$

<u>שאלה 7</u>

נחשב ראשית נגזרות ראשונות של $f(x,y)=x^3-5xy-y^5$:

$$\frac{\partial f}{\partial x}(x,y)=3x^2-5y,\qquad \frac{\partial f}{\partial y}(x,y)=-5x-5y^4$$

ונחשב כעת נגזרות שניות:

$$\frac{\partial^2 f}{\partial^2 x}(x,y)=6x,\qquad \frac{\partial^2 f}{\partial^2 y}(x,y)=-20y^3,\qquad \frac{\partial^2 f}{\partial y\partial x}=-5=\frac{\partial^2 f}{\partial x\partial y}$$

ולכן קיבלנו:

$$H(f)=\begin{pmatrix}6x & -5\\ -5 & -20y^3\end{pmatrix}$$

<u>שאלה 8</u>

יהי $n\in\mathbb{N}$, ו-$\varepsilon>0$. נרצה לחשב את $P(|\hat{\mu}_n-\mu|\geq\varepsilon)$. מאי שוויון מרקוב נקבל:

$$P(|\hat{\mu}_n-\mu|\geq\varepsilon)\leq\frac{\mathbb{E}[|\hat{\mu}_n-\mu|]}{\varepsilon}$$

ניזכר כי הפונקציה $f(x)=x^2$ היא פונקציה <u>קמורה</u>, לכן מאי שוויון ינסן נקבל כי:

$$\mathbb{E}^2[|\hat{\mu}_n-\mu|]\leq\mathbb{E}[(\hat{\mu}_n-\mu)^2]$$

אבל הוכחנו בהרצאה כי $\mathbb{E}[\hat{\mu}_n]=\mu$, לכן:

$$\mathbb{E}^2[|\hat{\mu}_n-\mu|]\leq\mathbb{E}[(\hat{\mu}_n-\mu)^2]=\mathbb{E}[(\hat{\mu}_n-\mathbb{E}[\hat{\mu}_n])^2]=Var[\hat{\mu}_n]=Var\left[\frac{1}{n}\sum_{i=1}^{n}x_i\right]=\frac{1}{n^2}\cdot Var\left[\sum_{i=1}^{n}x_i\right]$$

אבל נתונה לנו אי תלות, לכן:

$$\frac{1}{n^2}\cdot Var\left[\sum_{i=1}^{n}x_i\right]=\frac{1}{n^2}\sum_{i=1}^{n}Var[x_i]=\frac{1}{n^2}\cdot n\cdot\sigma^2=\frac{\sigma^2}{n}$$

ולכן נסיק כי:

$$\mathbb{E}^2[|\hat{\mu}_n-\mu|]\le\frac{\sigma^2}{n}\to\mathbb{E}[|\hat{\mu}_n-\mu|]\le\frac{\sigma}{\sqrt{n}}$$

נציב באי השוויון ונקבל:

$$P(|\hat{\mu}_n-\mu|\ge\varepsilon)\le\frac{\mathbb{E}[|\hat{\mu}_n-\mu|]}{\varepsilon}\le\frac{\sigma}{\varepsilon\sqrt{n}}\to_{n\to\infty}0$$

כנדרש.

<u>שאלה 9</u>

נסמן $\theta=\{\mu,\Sigma\}$, וב-f_θ את פונקציית הצפיפות של $\boldsymbol{X}$ לפי הפרמטר θ. מהגדרת ה-$likelihood$ מתקיים לכל $i\in[n]$:

$$L(\theta|\boldsymbol{x}_i)=f_\theta(\boldsymbol{x_i})=\frac{1}{\sqrt{(2\pi)^d|\Sigma|}}\exp\left\{-\frac{1}{2}(\boldsymbol{x_i}-\mu)^T\Sigma^{-1}(\boldsymbol{x_i}-\mu)\right\}$$

אבל $\{\boldsymbol{x_1},\dots,\boldsymbol{x_m}\}$ בלתי תלויים, לכן:

$$\begin{aligned}L(\theta|\boldsymbol{x_1},\dots,\boldsymbol{x_m})=f_\theta(\boldsymbol{x_1},\dots,\boldsymbol{x_m})&=\prod_{i=1}^{m}f_\theta(\boldsymbol{x_i})\\&=\frac{1}{\left(\sqrt{(2\pi)^d|\Sigma|}\right)^m}\exp\left\{\sum_{i=1}^{m}-\frac{1}{2}(\boldsymbol{x_i}-\mu)^T\Sigma^{-1}(\boldsymbol{x_i}-\mu)\right\}\end{aligned}$$

נפעיל ln ומחוקי לוגריתמים נקבל:

$$ln\big(L(\theta|\boldsymbol{x_1},\dots,\boldsymbol{x_m})\big)=-\frac{m}{2}\cdot\ln\big((2\pi)^d|\Sigma|\big)-\frac{1}{2}\cdot\sum_{i=1}^{m}(\boldsymbol{x_i}-\mu)^T\Sigma^{-1}(\boldsymbol{x_i}-\mu)=$$

$$-\frac{md}{2}\ln(2\pi)-\frac{m}{2}\ln(|\Sigma|)-\frac{1}{2}\cdot\sum_{i=1}^{m}(\boldsymbol{x_i}-\mu)^T\Sigma^{-1}(\boldsymbol{x_i}-\mu)$$

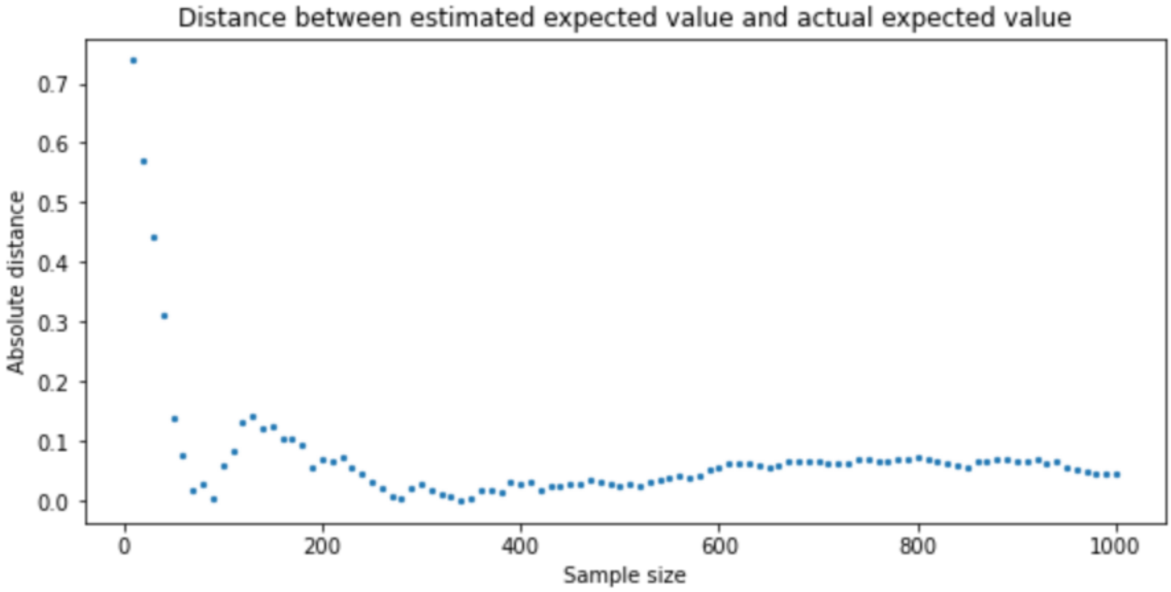
כדרוש.

חלק פרקטי

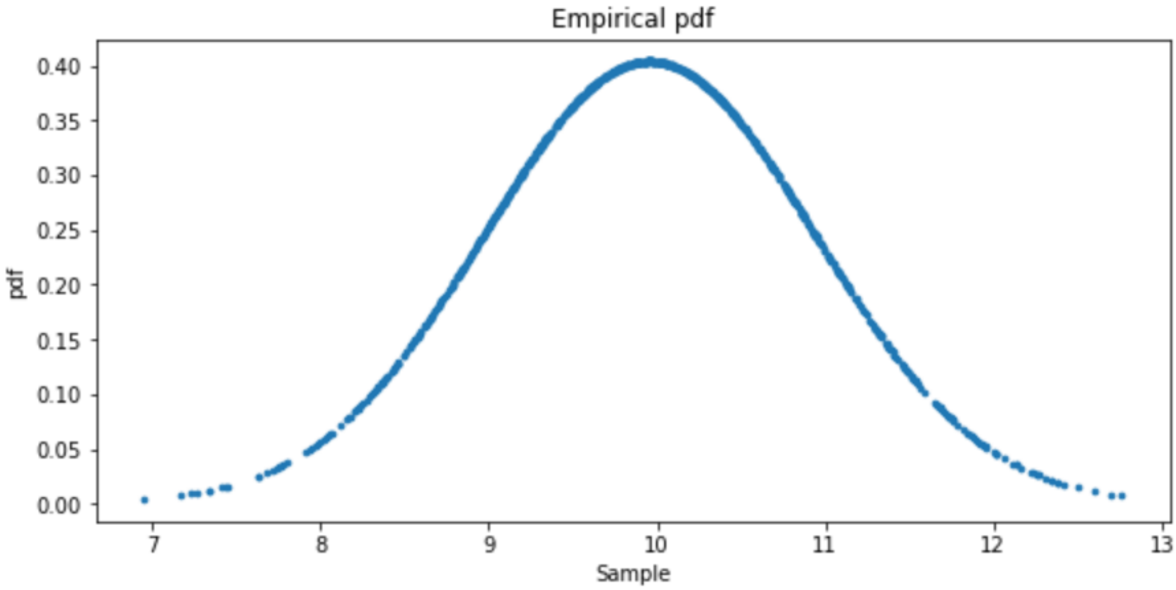
שאלה 1

הערך שקיבלתי עבור התוחלת הוא 9.954743292509804 ועבור השונות 0.9752096659781323.

שאלה 2



שאלה 3

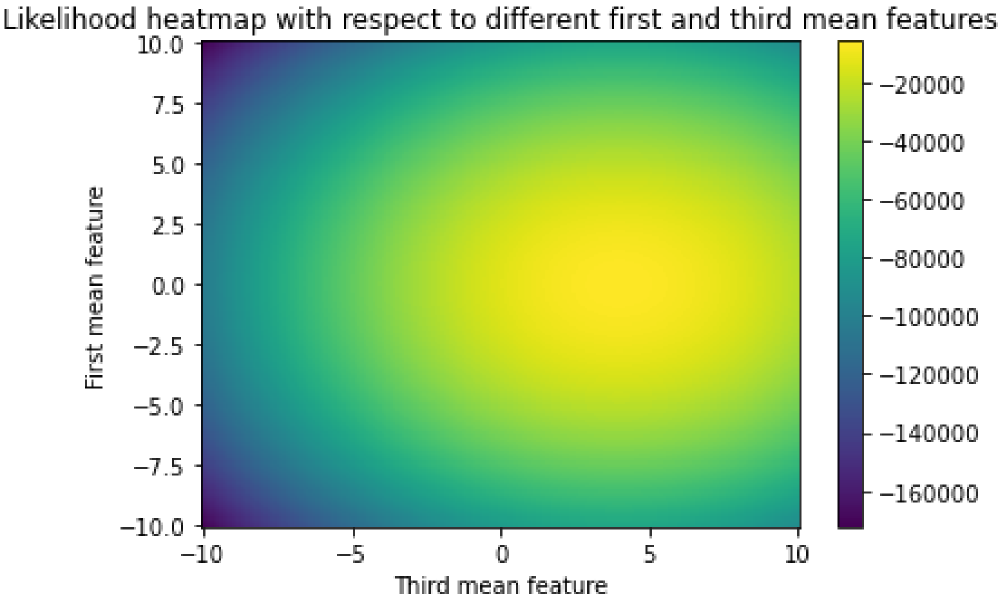


שאלה 4

צילום מסך של וקטור התוחלת ומטריצות השונות המשותפת שקיבלתי (לפי הסדר):

```
[-0.02282878 -0.04313959  3.9932571  -0.02038981]
[[ 0.91667608  0.16634444 -0.03027563  0.46288271]
 [ 0.16634444  1.9741828  -0.00587789  0.04557631]
 [-0.03027563 -0.00587789  0.97960271 -0.02036686]
 [ 0.46288271  0.04557631 -0.02036686  0.9725373 ]]
```

שאלה 5



ניתן ללמוד מהגרף שמבחן היחס נראות מקבל ערך מקסימלי כאשר הפיצ'ר הראשון הוא בערך 0, והפיצ'ר השלישי הוא בערך 4. כלומר, בהינתן שהפיצ'ר השני והרביעי הם 0 בתוחלת, ניתן לומר שהתוחלת המתאימה ביותר לתצפיות היא הוקטור $[0,0,4,0]$, וזהו אכן הוקטור תוחלת בו השתמשנו כאשר דגמנו את התצפיות.

שאלה 6

הערך שקיבלתי עבור הפיצ'ר הראשון הוא -0.05, ועבור הפיצ'ר השני 3.97. אכן מאוד קרוב לערכים המקוריים של וקטור התוחלת.