תרגיל 1 | IML (67329)

שם: רונאל חרדים | ת"ז:208917641

חלק I

$Theoretical:$

שאלה 1

תהי $T:V\Rightarrow W$ ה"ל ו A מטריצה תואמת, נוכיח כי אם A אורתוגונלית אזי $\|Ax\|=\|x\| \quad \forall x\in V$:

הוכחה:

אנו יודעים ש A אורתוגונלית, כלומר מרחב העמודות והשורות של המטריצה הן מרחב אורתונורמלי, כלומר הנורמה של הווקטורים שווה ל 1 לכן:

$$\|Ax\| = ||A||\cdot||x|| = I||x|| = ||x||$$

כנדרש.

שאלה 2

נחשב את הפירוק של $A=U\Sigma V^T$:

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 2 \end{bmatrix}$$

נחשב את המטריצות:

$$A^TA = V\Sigma^2V^T = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 2 \\ 0 & 2 & -2 \\ 2 & -2 & 4 \end{bmatrix}$$

$$det\left(A^TA-\lambda I\right)=\begin{vmatrix} 2-\lambda & 0 & 2 \\ 0 & 2-\lambda & -2 \\ 2 & -2 & 4-\lambda \end{vmatrix}=-\lambda^3+8\lambda^2-12\lambda=-\lambda\left(\lambda-2\right)\left(\lambda-6\right))\Rightarrow 0,2,6$$

כעת נחשב:

$$A^TA-0I=\begin{bmatrix} 2 & 0 & 2 \\ 0 & 2 & -2 \\ 2 & -2 & 4 \end{bmatrix}\Rightarrow eigenvector\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$

$$A^TA-2I=\begin{bmatrix} 0 & 0 & 2 \\ 0 & 0 & -2 \\ 2 & -2 & 2 \end{bmatrix}\Rightarrow eigenvector:\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

$$A^TA-6I=\begin{bmatrix} -4 & 0 & 2 \\ 0 & -4 & -2 \\ 2 & -2 & -2 \end{bmatrix}\Rightarrow eigenvector:\begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{bmatrix}$$

לכן V היא:

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

המטריצה Σ היא:

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

כעת נמצא את U:

$$u_1=\frac{1}{\sigma_1}\cdot\begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 2 \end{bmatrix}\cdot v_1=\frac{1}{\sqrt{6}}\cdot\begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 2 \end{bmatrix}\cdot\begin{bmatrix} \frac{\sqrt{6}}{6} \\ -\frac{\sqrt{6}}{\sqrt{6}} \\ \frac{\sqrt{3}}{3} \end{bmatrix}=\begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$u_2=\frac{1}{\sigma_2}\cdot\begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 2 \end{bmatrix}\cdot v_2=\frac{1}{\sqrt{2}}\cdot\begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 2 \end{bmatrix}\cdot\begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \\ 0 \end{bmatrix}=\begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

קיבלנו ש:

$$U=\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

לכן:

$$A=\begin{bmatrix}1&1&0\\1&-1&2\end{bmatrix}=\begin{bmatrix}0&1\\1&0\end{bmatrix}\begin{bmatrix}\sqrt{6}&0&0\\0&\sqrt{2}&0\end{bmatrix}\begin{bmatrix}\frac{\sqrt{6}}{6}&\frac{\sqrt{2}}{2}&-\frac{\sqrt{3}}{3}\\-\frac{\sqrt{6}}{6}&\frac{\sqrt{2}}{2}&\frac{\sqrt{3}}{3}\\\frac{\sqrt{6}}{3}&0&\frac{\sqrt{3}}{3}\end{bmatrix}^T$$

כנדרש.

שאלה 3

נוכיח את האלגוריתם $Power-Iteration$**:**

נוכיח כי מתקיים $\lim_{k\to\infty} b_k=\pm v_1$**:**

הווקטור v_1 הוא ווקטור עצמי של ע"ע λ ומקיים $Av_1=\lambda v_i$

$$\lim_{k\to\infty} b_k=C_0^k\cdot b_0=\lim_{k\to\infty} C_0^k\sum_{i=1}^{n}a_iv_i=\lim_{k\to\infty}\sum_{i=1}^{n}a_i\lambda_i^kv_i=\lim_{k\to\infty}\lambda_1^k\left[a_1v_1+\sum_{i=2}^{n}a_i\left(\frac{\lambda_i}{\lambda_1}\right)^kv_i\right]\Rightarrow_{k\to\infty}\lambda_1^k$$

מכיוון שמתקיים $\lambda_1>\lambda_2\geq\lambda_3...$ אזי מתקיים כי הסיגמא מתכנסת ל 0 כאשר $k\to\infty$ לכן מתקיים כי הסדרה תתכנס ל $\lambda_1 v$ כנדרש.

שאלה 4

נחשב את מטריצת יעקוביאן של הפונקציה הנתונה.

$$f(\sigma)=U\cdot diag(\sigma)U^\top x=\sum_{i=1}^{n}\sigma_i\boldsymbol{u}_i\boldsymbol{u}_i^\top\boldsymbol{x}$$

עבור

$$diag(\sigma)_{ij}=\begin{cases}\sigma_i & i=j\\0 & i\neq j\end{cases}$$

תחילה נחשב את הנגזרות החלקיות של הפונקציה: נשים לב כי עבור כל σ_i הנגזרת החלקית בנקודה ה i שווה ל $u_{i,i}^2x_i$, ובשאר הקורדינטות שווה ל 0.

לכן:

$$\frac{\partial f_j\left(\sigma_0\right)}{\partial\sigma_i}=\left[u_iu_i^\top\boldsymbol{x}\right]_j$$

ומתקיים כי השורה ה i **שווה ל** $u_i\cdot\langle u_i,\boldsymbol{x}\rangle$ **לכן המטריצה היא:**

$$J_\sigma(f):=Udiag\left(U^\top\boldsymbol{x}\right)$$

שאלה 5

נשתמש בכלל השרשרת ונחשב את הגרדיאנט:

$$h(\sigma) = \frac{1}{2}\|f(\sigma) - y\|^2$$

נגדיר: $g(x) = ||x||^2$**.** לפי כלל השרשת מתקיים $(f \circ g)' := (f' \circ g) \cdot g'$ץ

נחשב את הנגזרת של $g(x)$:

$$g(x)' = \nabla f(\sigma)$$

עבור מה שחישבנו בשאלה קודמת.

נחשב את הנגזרת של f ונקבל σ. בנוסף, בהרצאה הראינו כי הגרדינאט שווה לטרנספוז של המטריצה.

לכן סה"כ נקבל כי:

$$\nabla h = (\sigma)(f(\sigma) - y)^T \nabla f(\sigma)$$

כנדרש.

שאלה 6

נחשב את היעקוביאן של הפונקציה הבאה:

$$S(\mathbf{x})_j = \frac{e^{x_j}}{\sum_{l=1}^{k} e^{x_l}}$$

נסמן $g_i(a) = e^{a_i}$ ו $g_i(a) = e^{a_i}$

נקבל:

$$\frac{\partial S_i}{\partial a_j} = \frac{\partial}{\partial a_j} \frac{e^{a_i}}{\sum_{k=1}^{d} e^{a_k}} = \frac{\partial}{\partial a_j} \frac{g_i}{h}$$

נשים לב כי הנגזרת של h שווה ל

$$h' = \begin{cases} e^{a_j} & i = j \\ 0 & else \end{cases}$$

לכן מתקיים כי במקום ה j עבור $i = j$:

$$\frac{\partial}{\partial a_j} \frac{e^{a_i}}{\sum_{k=1}^{N} e^{a_k}} = \frac{e^{a_i}\left(\sum_{k=1}^{N} e^{a_k}\right) - e^{a_i} e^{a_j}}{\left(\sum_{k=1}^{N} e^{a_k}\right)^2} = \frac{e^{a_i}}{\left(\sum_{k=1}^{N} e^{a_k}\right)} \cdot \frac{\left(\sum_{k=1}^{N} e^{a_k}\right) - e^{a_j}}{\left(\sum_{k=1}^{N} e^{a_k}\right)} = S_i\left(1 - S_j\right)$$

ונקבל כי מטריצת יעקוביאן של הפונקציה S בקאורדינטה ה i, j מכילה את

$$S_i(x)\left(\delta_{ij} - S_j(x)\right)$$

כנדרש.

שאלה 7

נחשב את ההסיאן של הפונקציה $x^3 - 5xy - y^5$:

כפי שהזכרנו בתרגול, ההסיאן היא נגזרת של f פעמיים ־ פעם אחת לפי x ופעם שניה לפי y:

תחילה נגזור לפי x:

$$\frac{\partial f}{\partial x} = 3x^2 - 5x$$

כעת נגזור לפי y:

$$\frac{\partial f}{\partial y} = -5y - 5y^4$$

נגזור פעם שניה לפי x:

$$\frac{\partial^2 f}{\partial x^2} = 6x - 5$$

נגזור פעם שניה לפי y:

$$\frac{\partial^2 f}{\partial y^2} = -5 - 20y^3$$

קיבלנו כי המטריצה היא:

$$H_{i,j} = \begin{pmatrix} 6x & -5 \\ -5 & -20y^3 \end{pmatrix}$$

כנדרש.

שאלה 8

נתונות דגימות $x_1, x_2...$ ב"ת בעלות אותה התפלגות, עם תוחלת ושונות סופית. נראה כי אומד ממוצע המדגם $\hat{\mu}_n = \frac{1}{n}\sum x_i$ המחושב על n דגימות הוא קבוע.

כלומר צ"ל $\lim_{n\to\infty} \hat{\mu}_n = \mu$.

הוכחה:

ראינו בהרצאה כי מתקיים $\mathbb{E}[\hat{\mu}_n] = \mu$, לכן נותר להוכיח ש $var[\hat{\mu}_n] \to_{n\to\infty} 0$.

$$\lim_{n\to\infty} Var(\hat{\mu}_n) = \lim_{n\to\infty} Var\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \lim_{n\to\infty} \frac{1}{n^2} Var\left(\sum_{i=1}^{n} x_i\right) = \lim_{n\to\infty} \frac{1}{n^2} \cdot n \cdot \sigma^2 = \lim_{n\to\infty} \frac{\sigma^2}{n} \stackrel{1}{=} 0$$

כאשר 1 נובע מסופיות שהשונות.

כנדרש.

שאלה 9

נמצא את הדרוש:

בהרצאה הראנו כי מתקיים עבור התפלגות נורמלית:

$$X \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ . \\ . \\ \mu_d \end{bmatrix}, \begin{bmatrix} \sigma_1^1 & 0 & 0 & 0 \\ 0 & . & 0 & 0 \\ 0 & 0 & . & 0 \\ 0 & 0 & 0 & \sigma_d^d \end{bmatrix}\right)$$

וגם, עבור $x_i \in X$ מתקיים:

$$f\left(x_i\right) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2}\left(\frac{X_i-\mu_i}{\sigma_i}\right)^2\right)$$

הגדרנו שפונקציית הנראות שווה ל $\mathcal{C}(\boldsymbol{\theta} \mid x) := f_{\boldsymbol{\theta}}(x)$.

נחשב עבור X:

נתון שהמשתנים מתפלגים נורמלית וב"ת לכן ההסתברות שווה למכפלה, ונקבל:

$$\mathcal{C}(\boldsymbol{\theta} \mid x) := f_{\boldsymbol{\theta}}(X) = \prod_{i=1}^{m} f_{\boldsymbol{\theta}}(x_i) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2}\left(\frac{X_i-\mu_i}{\sigma_i}\right)^2\right) = \frac{m}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{m}\left(x_i-\mu\right)^2\right)$$

כעת נחשב את $log\ liklihood$:

$$ln\left[\frac{m}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{m}\left(x_i-\mu\right)^2\right)\right] = ln\left[\frac{m}{\sqrt{2\pi\sigma_i^2}}\right] + ln\left[\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{m}\left(x_i-\mu\right)^2\right)\right] =$$

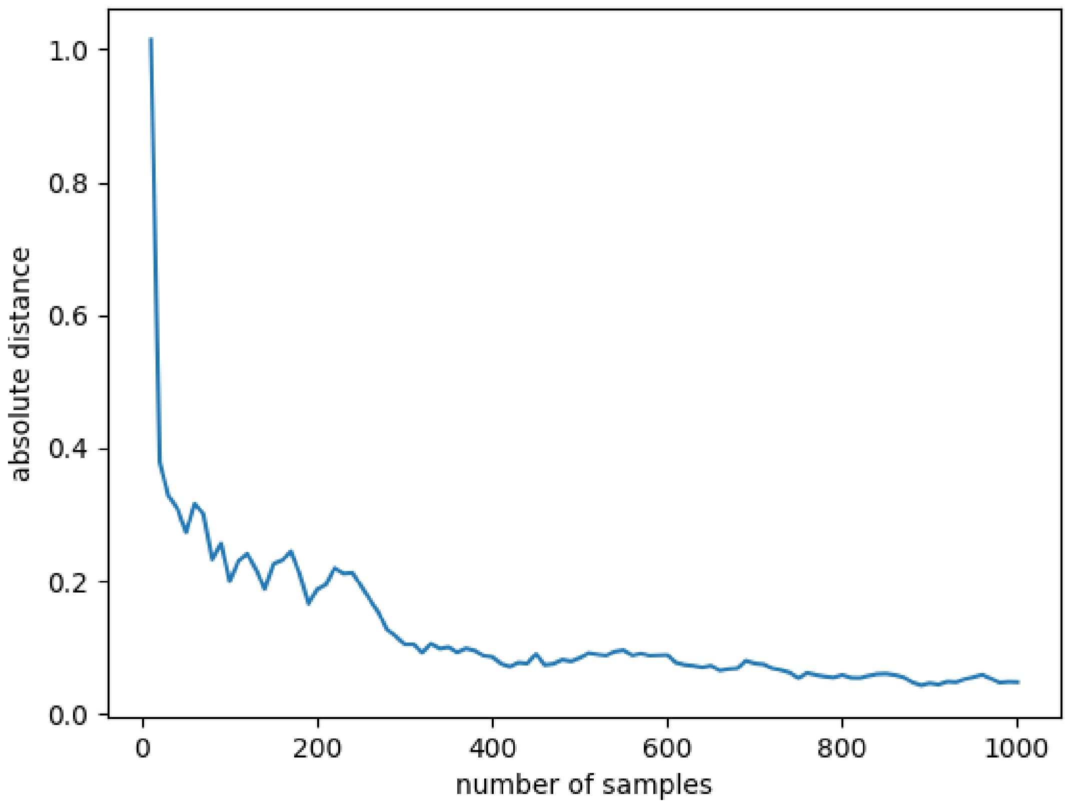
$$-\frac{m}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}\left(x_i-\mu\right)^2 = -\frac{m}{2}\ln(2\pi) - \frac{m}{2}\ln\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}\left(x_i-\mu\right)^2$$

חלק II

$Practical:$

שאלה 2

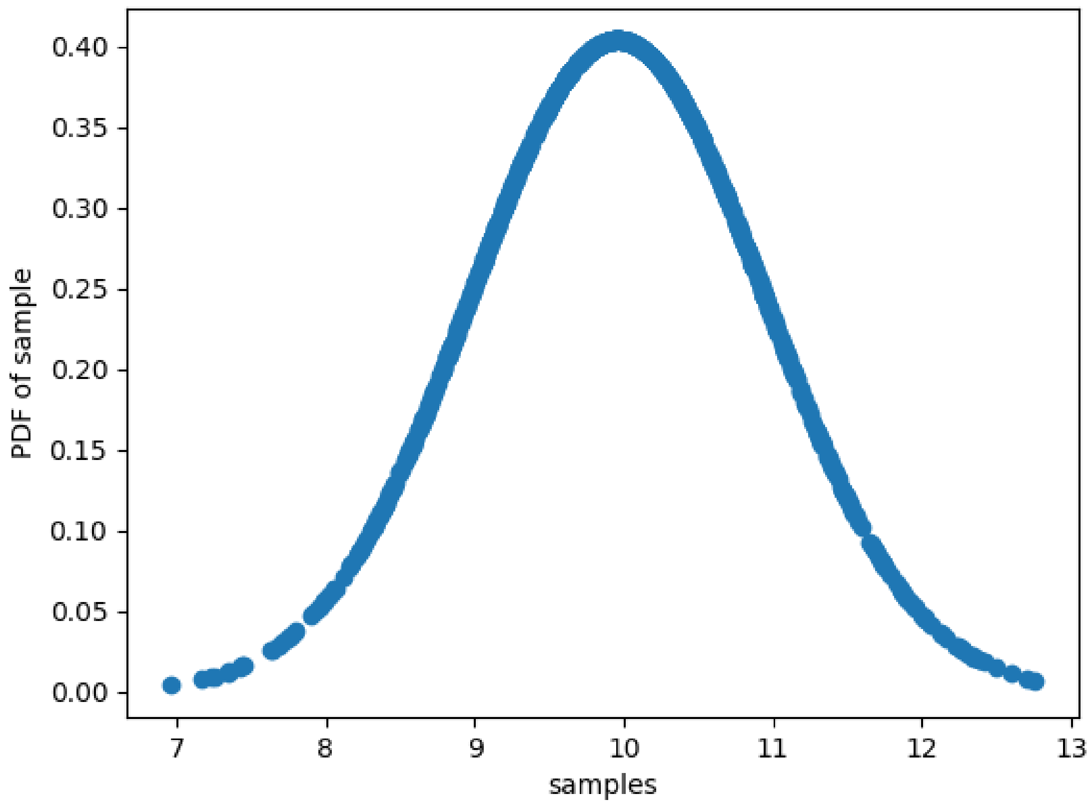
הגרף שקיבלנו:



שאלה 3

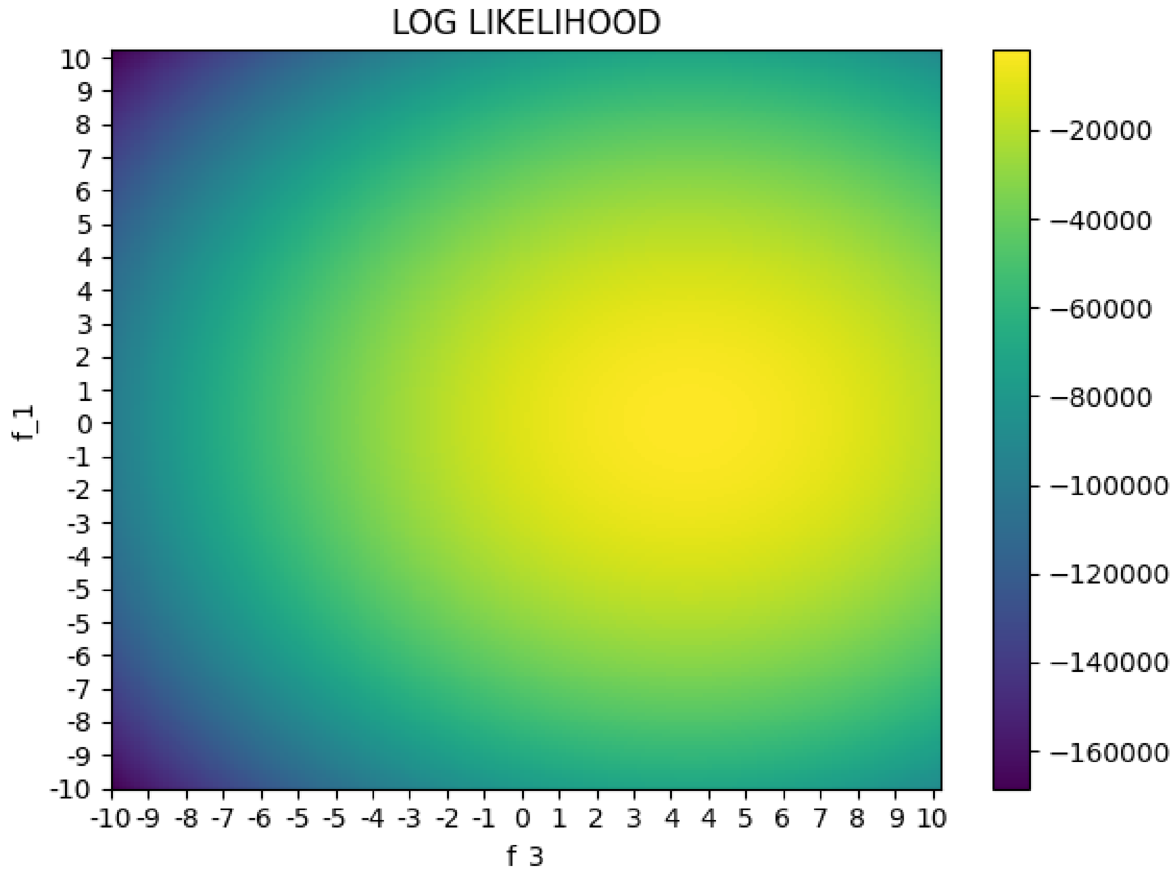
הגרף שקיבלנו.

ניתן להסיק מהגרף כי הערכים מתפלגים נורמלית סביב 10.



שאלה 5

גרף ה $heatmap$ הוא:



מהגרף ניתן לראות כי הערך המקסימלי מתקבל כאשר $f1 :\approx 0, f3 :\approx 4$.

שאלה 6

הערך המקסימלי מתקבל ע"י $f1 = -0.050, f3 = 3.970$