IML תרגיל 1

2.1.1 (1) A אורתוגונלית $\Leftarrow$ $AA^T = A^TA = I_n$

שים לב, $\|Ax\| = \|x\| \iff \|Ax\|^2 = \|x\|^2$

כמו כן, ראינו בלינארית 2 כי אחת התכונות של מט'

אורתוגונלית היא ש- $\langle Av, Aw\rangle = \langle v, w\rangle \quad \forall v,w \in V$

לכן נקבל

$$\|Av\|^2 = \langle Av, Av\rangle = \langle v, v\rangle = \|v\|^2$$

כלומר, $\|Av\| = \|v\|$ כנדרש.

(2) תחילה, נחשב את A^TA:

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 2 \\ 0 & 2 & -2 \\ 2 & -2 & 4 \end{bmatrix}$$

כעת, נמצא את הע"ע של A^TA:

פיתוח לפי
שורה 1

$$\det\left(A^TA-\lambda I_n\right)=\left|\begin{bmatrix}2-\lambda & 0 & 2\\ 0 & 2-\lambda & -2\\ 2 & -2 & 4-\lambda\end{bmatrix}\right|=(2-\lambda)\left|\begin{bmatrix}2-\lambda & -2\\ -2 & 4-\lambda\end{bmatrix}\right|+2\begin{vmatrix}0 & 2-\lambda\\ 2 & -2\end{vmatrix}=$$

$$=(2-\lambda)\Big((2-\lambda)(4-\lambda)-(-2\cdot-2)\Big)+2\Big(-2(2-\lambda)\Big)=$$

$$=(2-\lambda)\left(8-2\lambda-4\lambda+\lambda^2-4\right)+2\left(-4+2\lambda\right)=(2-\lambda)\left(\lambda^2-6\lambda+4\right)+$$

$$+2\left(2\lambda-4\right)=2\lambda^2-12\lambda+\underline{8}-\lambda^3+6\lambda^2-4\lambda+4\lambda-8=$$

$$=-\lambda^3+8\lambda^2-12\lambda$$

לאחר פתרון של המשוואה נקבל

$\lambda_1=6$ $\lambda_3=0$

$\lambda_2=2$

כעת נמצא בסיס של ו"ע :

- עבור $\lambda=6$:

$$\begin{bmatrix}2 & 0 & 2\\ 0 & 2 & -2\\ 2 & -2 & 4\end{bmatrix}\begin{pmatrix}x\\ y\\ z\end{pmatrix}=\begin{pmatrix}6x\\ 6y\\ 6z\end{pmatrix}$$

$$\begin{cases} ①\ 2x + 2z = 6x \\ ②\ 2y - 2z = 6y \\ ③\ 2x - 2y + 4z = 6z \end{cases} \Longrightarrow \begin{matrix} ①\ 2z = 4x \\ ②\ -2z = 4y \\ ③\ 2x - 2y = 2z \end{matrix} \Longrightarrow \begin{matrix} ①\ z = 2x \\ ②\ z = -2y \\ ③\ x - y = z \end{matrix}$$

$$\Longrightarrow ①+②\quad 2x = -2y \longrightarrow x = -y$$

לכן, עבור $\lambda = 6$ ע"ע הוא $\begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$

עבור $\lambda = 2$

$$\begin{bmatrix} 2 & 0 & 2 \\ 0 & 2 & -2 \\ 2 & -2 & 4 \end{bmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2x \\ 2y \\ 2z \end{pmatrix}$$

$$\begin{cases} ①\ 2x + 2z = 2x \\ ②\ 2y - 2z = 2y \\ ③\ 2x - 2y + 4z = 2z \end{cases} \Longrightarrow \begin{cases} 2z = 0 \\ -2z = 0 \\ 2x - 2y = -2z \end{cases} \Longrightarrow \begin{matrix} z = 0 \\ x = y \end{matrix}$$

לכן עבור $\lambda = 2$ ע"ע הוא $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$

עבור $\lambda=0$:

$$\begin{bmatrix} 2 & 0 & 2 \\ 0 & 2 & -2 \\ 2 & -2 & 4 \end{bmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{array}{l} ① \\ ② \\ ③ \end{array} \begin{cases} 2x+2z=0 \\ 2y-2z=0 \\ 2x-2y+4z=0 \end{cases} \Longrightarrow \begin{cases} x=-z \\ y=z \\ x-y=-2z \end{cases} \Longrightarrow \begin{array}{l} x=-z \\ y=z \end{array}$$

לכן $\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$ הוא ו"ע עבור $\lambda=0$

כלומר, קיבלנו $\left\{ \underset{v_1}{\begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}}, \underset{v_2}{\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}}, \underset{v_3}{\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}} \right\}$ בסיס של ו"ע

כעת נמצא G אחרי נרמול בסיס לכל וקטור עצמי:

$$\frac{v_1}{\|v_1\|} = \frac{1}{\sqrt{1+1+4}} v_1 = \frac{1}{\sqrt{6}} v_1 \qquad \frac{v_3}{\|v_3\|} = \frac{1}{\sqrt{1+1+1}} v_3 = \frac{1}{\sqrt{3}} v_3$$

$$\frac{v_2}{\|v_2\|} = \frac{1}{\sqrt{1+1}} v_2 = \frac{1}{\sqrt{2}} v_2$$

כלומר קיבלנו

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

כמובן, ערכי האלכסון של Σ הם השורשים של הע״ע של AA^T, והוא מגודל $m \times n$ לכן,

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

כדי נמצא את המשוואה:

$$AV = U\Sigma$$

$$\Updownarrow$$

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

$$\Updownarrow$$

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

$$\Updownarrow$$

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

(3) יהיו

$$u = \begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix} \qquad V = \begin{pmatrix} V_1 \\ \vdots \\ V_n \end{pmatrix}$$

כמובן המימדים שלהם כללי:

$$A = V \otimes u = \begin{bmatrix} V_1 u_1 & V_1 u_2 & \dots & V_1 u_m \\ V_2 u_1 & V_2 u_2 & \dots & V_2 u_m \\ \vdots & \vdots & \ddots & \vdots \\ V_n u_1 & V_n u_2 & \dots & V_n u_m \end{bmatrix}$$

נסמן $c_1 \dots c_m$ עמודות A, ונשים לב כי מתקיים

$$c_j = u_j \cdot V$$

כש u_j הוא הכניסה ה-j של u

כלומר, כל עמודה ב-A היא כפולה של V בסקלר.

לכן, לכל $1\le i,j\le m$ עבורם $u_i, u_j \neq 0$, נקבל

$$c_i - \frac{u_i}{u_j}c_j = u_iV - \frac{u_i}{u_j}\cdot\left(u_jV\right) = u_iV - u_iV = 0$$

וכיוון ש $\frac{u_i}{u_j}\neq 0$ אזי קיבלנו צ"ל לא טריוויאלי המתאפס

לכן העמודות c_i, c_j תלויות, כלומר הן תלויות לינארית.

כיוון שבחרנו שרירותי i,j כלשהם (שבהם u_i, u_j לא אפס) כל

מכאן (בהכרח) אזי כל העמודות ב A הן תלויות ולכן $rk(A)=1$

כנדרש ∎

④ צ"ל: לכל $X\in\mathbb{R}^n$ כך ש $X=\sum_{i=1}^{n} a_iu_i$ מתקיים:

$$a_i = \langle X, u_i\rangle \iff a_i = \left\langle \sum_{j=1}^{n} a_ju_j, u_i\right\rangle \iff$$

$$\iff a_i = \langle a_1u_1, u_i\rangle + \langle a_2u_2, u_i\rangle + \dots + \langle a_nu_n, u_i\rangle \iff$$

לינאריות

$$\iff a_i = a_1\langle u_1, u_i\rangle + a_2\langle u_2, u_i\rangle + \dots + a_n\langle u_n, u_i\rangle \quad (*)$$

כי הבסיס אורתונורמלי, ולכן לכל $i\neq j$

מתקיים $\langle u_i, u_j\rangle = 0$ ומהגדרת אורתונורמליות מתקיים

$\langle u_i, u_i\rangle = 1$. לכן אם נציב ב- (*):

$$a_i = a_1\langle u_1, u_i\rangle + a_2\langle u_2, u_i\rangle + \dots + a_n\langle u_n, u_i\rangle \iff$$

$$\iff a_i = a_i\langle u_i, u_i\rangle \iff a_i = a_i$$

כלומר, הוכחנו ש $a_i = \langle x, u_i\rangle \iff a_i = a_i$

אז נכון תמיד כנדרש. $\blacksquare$

$$\operatorname{diag}(\sigma) = \begin{bmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_n \end{bmatrix}$$

S (2.1.2

$$U \operatorname{diag}(\sigma) U^T x =$$

$$= \begin{pmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ u_{n1} & \dots & & u_{nn} \end{pmatrix} \begin{pmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_n \end{pmatrix} \begin{pmatrix} u_{11} & \dots & & u_{n1} \\ \vdots & u_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ & & & u_{nn} \\ u_{1n} & \dots & & \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ \vdots \\ x_n \end{pmatrix}$$

$$= \begin{pmatrix} u_{11}\sigma_1 & u_{12}\sigma_2 & \dots \\ u_{21}\sigma_1 & u_{22}\sigma_2 & \dots \\ \vdots & & \ddots \\ u_{n1}\sigma_1 & & u_{nn}\sigma_n \end{pmatrix} \begin{pmatrix} u_{11} & \dots & u_{n1} \\ \vdots & u_{22} & \\ \vdots & & \\ u_{1n} & & u_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ \vdots \\ x_n \end{pmatrix} =$$

$$= \begin{pmatrix} \sum_{i=1}^{n} u_{1i} u_{1i} \sigma_i & \cdots & \\ \sum_{i=1}^{n} u_{1i} u_{1i} \sigma_i & \ddots & \\ \vdots & & \ddots \\ \sum_{i=1}^{n} u_{ni} u_{ni} \sigma_i & \cdots & \sum_{i=1}^{n} u_{ni} u_{ni} \sigma_n \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

כלומר, נשים לב שלמעשה קיים:

$$f(\sigma) = \sum_{i=1}^{n} \sigma_i u_i u_i^T x =$$

עמודה ה-i ב-U

$$= \begin{pmatrix} (u_{11} u_{11} + \ldots)\sigma_1 + (\ldots)\sigma_2 + \ldots + (\ldots)\sigma_n \\ (u_{22} u_{11} + \ldots)\sigma_1 + \ldots \quad \ldots + (\ldots)\sigma_n \\ \\ \\ \end{pmatrix}$$

לכן, נגזור לפי ש

$$\left(f(\sigma)\right)_i = \sum_{j=1}^{n} \left(u_j u_j^T x\right)_i \cdot \sigma_j$$

כלומר, אם נגזור את f_i לפי σ_j נקבל:

$$\frac{\partial f_i(\sigma)}{\partial \sigma_j} = \left(u_j u_j^T x\right)_i$$

לכן זה נכון לכל $1\le i,j\le n$ ונקבל את מה
שרצינו.

$$h(\sigma) = \frac{1}{2}\|f(\sigma)-y\|^2 \qquad 6$$

$$h(\sigma) = \frac{1}{2}\left(\langle f(\sigma)-y, f(\sigma)-y\rangle\right) = \frac{1}{2}\left(\langle f(\sigma), f(\sigma)\rangle - 2\langle f(\sigma), y\rangle +\right.$$

$$\left. + \langle y, y\rangle\right) = \frac{1}{2}\left(\|f(\sigma)\|^2 - 2y^T f(\sigma) + \|y\|^2\right) = \frac{1}{2}\|f(\sigma)\|^2 - y^T f(\sigma) + \|y\|^2$$

נסמן $h_1(x) = \frac{1}{2}\|x\|^2$, $h_2(x) = y^T x$ ונקבל:

$$h_1 \circ f(\sigma) = \frac{1}{2}\|f(\sigma)\|^2 \Longrightarrow (h_1 \circ f)'(\sigma) \underset{\text{כלל השרשרת}}{=} f(\sigma)^T \cdot J_\sigma(f)$$

$$h_2 \circ f(\sigma) = y^T f(\sigma) \Longrightarrow (h_2 \circ f)'(\sigma) \underset{\text{כלל השרשרת}}{=} y^T \cdot J_\sigma(f)$$

לכן,

$$h(\sigma)' = f(\sigma)^T \cdot J_\sigma(f) - y^T J_\sigma(f) = \left(f(\sigma)-y\right)^T J_\sigma(f)$$

$$\Longrightarrow \quad \nabla h(\sigma)' = \left(\left(f(\sigma)-y\right)^T J_\sigma(f)\right)$$

7

$$S(x)=\begin{pmatrix} \dfrac{e^{x_1}}{\sum_{i=1}^{n} e^{x_i}} \\ \vdots \\ \dfrac{e^{n}}{\sum_{i=1}^{n} e^{x_i}} \end{pmatrix}$$

נגזור קואורדינטה: קואורדינטה של הערכים ונבדוק לפי מקרים:

אם $k=j$ כש k נגזור כמנה:

$$S'(x)_j=\frac{e^{x_j}\cdot\sum_{i=1}^{n} e^{x_i}-e^{x_j}e^{x_j}}{\left(\sum_{i=1}^{n} e^{x_i}\right)^2}=\frac{e^{x_j}\left(\left(\sum_{i=1}^{n} e^{x_i}\right)-e^{x_j}\right)}{\left(\sum_{i=1}^{n} e^{x_i}\right)^2}$$

אם $k\neq j$:

$$S(x)_j=e^{x_j}\cdot\left(\sum_{i=1}^{n} e^{x_i}\right)^{-1}\Rightarrow$$

$$\Rightarrow S'(x)_j=-e^{x_k}e^{x_j}\left(\sum_{i=1}^{n} e^{x_i}\right)^{-2}=\boxed{-\frac{e^{x_k}e^{x_j}}{\left(\sum_{i=1}^{n} e^{x_i}\right)^2}}$$

$$f(x,y) = x^3 - 5xy - y^5$$ 8

$$f'_x(x,y) = 3x^2 - 5y$$

$$f'_y(x,y) = -5x - 5y^4$$

$$\Longrightarrow \nabla f = \begin{pmatrix} 3x^2 - 5y \\ 5x - 5y^4 \end{pmatrix}$$

$$f''_{x,x}(x,y) = 6x \qquad f''_{y,x} = -5$$

$$f''_{x,y}(x,y) = -5 \qquad f''_{y,y} = -20y^3$$

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

$\forall u, v \quad \forall \alpha \in (0,1)$

$\alpha u + (1-\alpha) v \in C$

2.1.3) 9) יהיו $x,y\in C$, אזי $x,y\in\bigcap_{i\in I}C_i$

ובפרט גם שייכים ל-C_i לכל i. כיוון ש-C_i קבוצה קמורה,

מתקיים ש $\alpha x+(1-\alpha)y\in C_i$ לכל i, לכן

$\alpha x+(1-\alpha)y\in\bigcap_{i\in I}C_i=C$. לסיום הראינו ש-C קבוצה קמורה,

כנדרש.

10) יהיו $x_1,y_1\in C_1$, $x_2,y_2\in C_2$. אנו יודעים כי

$\forall i\in\{1,2\}\ \forall\alpha\in[0,1]\ \alpha x_i+(1-\alpha)y_i\in C_i$ לכן,

$$\alpha(\underset{\cap}{x_1+x_2})+(1-\alpha)(y_1+y_2)=\alpha x_1+\alpha x_2+(1-\alpha)y_1+(1-\alpha)y_2=$$

$$=\underset{\underset{C_1}{\cap}}{\left(\alpha x_1+(1-\alpha)y_1\right)}+\underset{\underset{C_2}{\cap}}{\left(\alpha x_2+(1-\alpha)y_2\right)}$$

לסיום, הראינו שלכל $\underset{x_1+x_2}{\overset{\parallel}{X}}\in C_1+C_2$, $\underset{y_1+y_2}{\overset{\parallel}{Y}}\in C_1+C_2$ מתקיים

$$\alpha X+(1-\alpha)Y\in C_1+C_2$$

iii) יהיו $x,y\in C$, נראה ש- λC היא קמורה:

$$x,y\in C \Rightarrow \lambda x,\lambda y\in \lambda C$$

אזי

$$\alpha\lambda x+(1-\alpha)\lambda y=\lambda\underbrace{\left(\alpha x+(1-\alpha)y\right)}_{\in C}$$

נסמן $t=\left(\alpha x+(1-\alpha)y\right)\in C$ (כי C קמורה) ונקבל כי:

$$\alpha\lambda x+(1-\alpha)\lambda y=\lambda t$$

סיימנו, הראינו כי λC קמורה גם היא.

2.2) (12) צ"ל: $\forall \varepsilon>0 \quad \lim_{n\to\infty} \mathbb{P}\left(|\hat{\mu}_n - \mu| > \varepsilon\right) = 0$

נשתמש בא"ש צ'בישב:

$$\mathbb{P}\left(|\hat{\mu}_n - \mu_n| \geq \varepsilon\right) \leq \frac{Var(\hat{\mu}_n)}{\varepsilon^2}$$

נסמן ב-σ^2 את $Var(X_i)$ ונחשב את $Var(\hat{\mu}_n)$:

$$Var(\hat{\mu}_n) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right) =$$

$$= \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$$

לכן:

$$\mathbb{P}\left(|\hat{\mu}_n - \mu_n| \geq \varepsilon\right) \leq \frac{\frac{\sigma^2}{n}}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$$

נשים לב כי לכל n ההסתברות הזו תמיד גדולה שווה 0,

לכן $0 \leq \mathbb{P}\left(|\hat{\mu}_n - \mu_n| \geq \varepsilon\right)$

וכיוון ש $\frac{\sigma^2}{\varepsilon^2}$ הוא קבוע, מ- סנדוויץ' זה נובע

$$\lim_{n\to\infty} \frac{\sigma^2}{n\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2} \cdot \lim_{n\to\infty} \frac{1}{n} = 0$$

לכן המומנט המשוער מתכנס?

$$\lim_{n\to\infty} \mathbb{P}\left(|\hat{\mu}_n - \mu_n| \geq \varepsilon \right) = 0$$

כנראה.

(13) נשים לב, שכפל בקבועות חיוביות לא משנה נקודות קיצון וגם

הפונקציה המונוטונית עולה, כמו log, שומרת על הסדר

לכן נקודת המקסימום שלנו היא גם של ה log-likelihood

ולכן נרצה למצוא מקסימום של הביטוי:

יהיו $x_1, \dots, x_n \overset{iid}{\sim} N(\mu, \Sigma)$:

$$L(\mu, \Sigma | x_1 \dots x_n) = f_{\mu,\Sigma}(x_1 \dots x_n) = \prod_{i=1}^{n} f_{\mu,\Sigma}(x_i) =$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{(2\pi)^d \cdot |\Sigma|}} \exp\left(-\frac{1}{2}(x_i-\mu)^T \Sigma^{-1} (x_i-\mu)\right) =$$

כמכפלה:

$$= \frac{1}{\left((2\pi)^d |\Sigma|\right)^{\frac{n}{2}}} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} (x_i-\mu)^T \Sigma^{-1} (x_i-\mu)\right)$$

נפעיל את ה log:

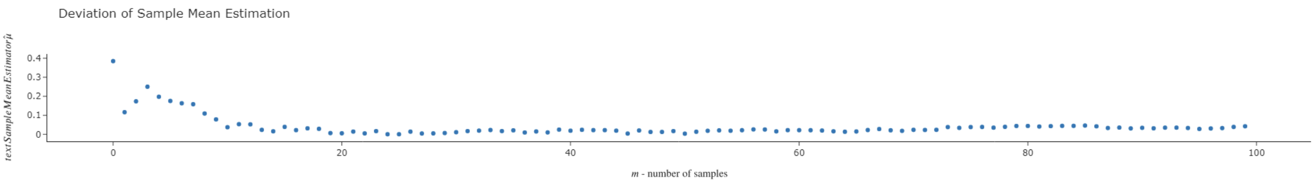
$$\log(L(\mu, \Sigma | x_1 \dots x_n)) = \log\left(\frac{1}{\left((2\pi)^d |\Sigma|\right)^{\frac{n}{2}}}\right) + \log\left(\exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)\right)\right) =$$

$$= \frac{n}{2}\left(\log(1) - \log\left((2\pi)^2 |\Sigma|\right)\right) - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu) =$$

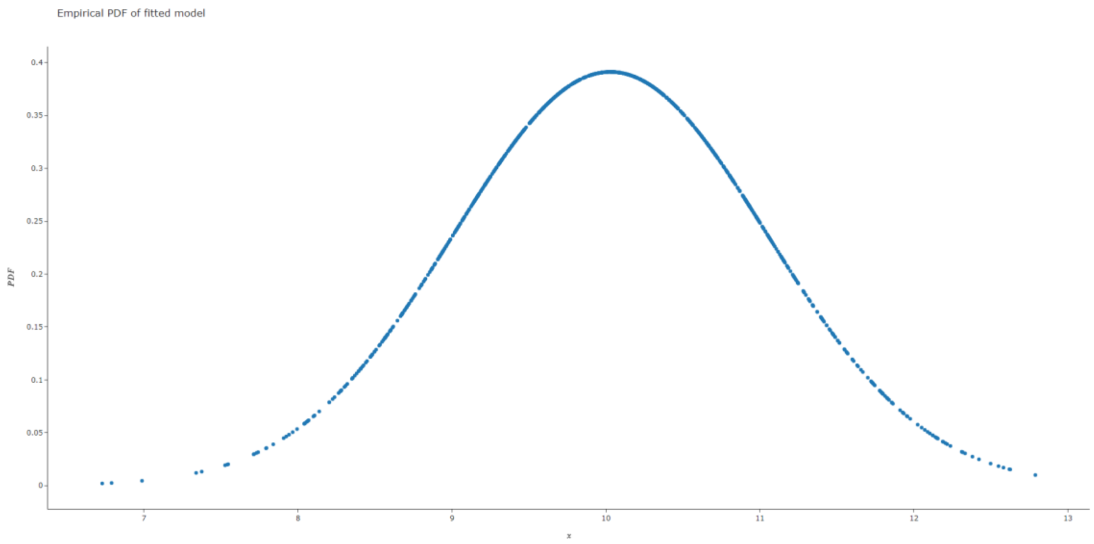
$$= -\frac{n}{2}\left(\log(2\pi)^2 + \log(|\Sigma|)\right) - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)$$

(1.0111270628421916 , 9.957785999684369) (1

(2



(3



(4

$\mu =$ [-0.04339347 -0.06314678 3.97132241 0.02595522]

$\Sigma =$

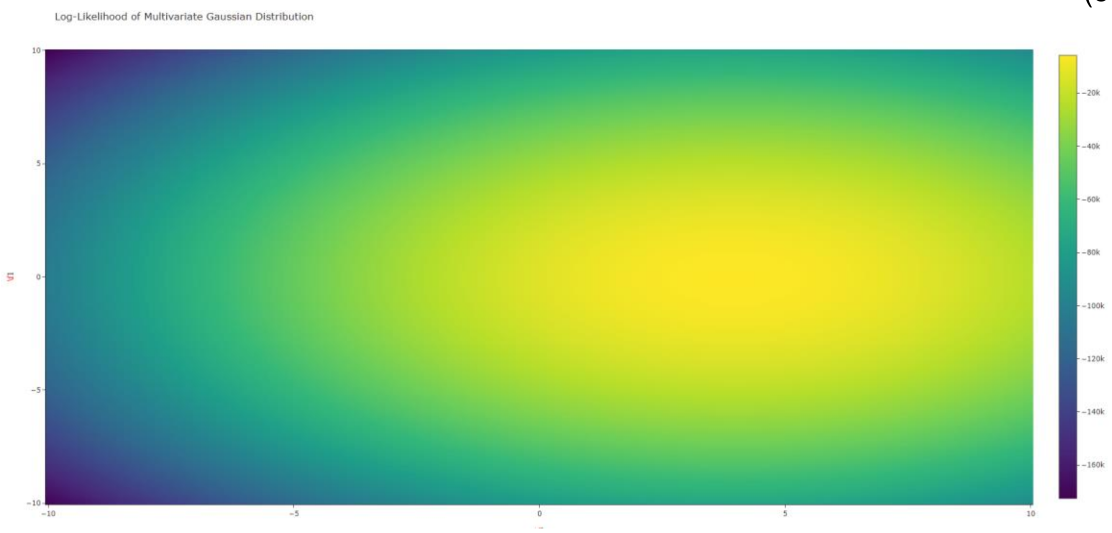
[[0.96227096 0.23840962 0.01354167 0.47540567]

[0.23840962 1.96171282 0.01872518 0.02383669]

[0.01354167 0.01872518 0.97800904 0.04238053]

[0.47540567 0.02383669 0.04238053 0.95344068]]

(5


Log-Likelihood of Multivariate Gaussian Distribution
10
5
0
−5
−10
−10
−5
0
5
10
−20k
−40k
−60k
−80k
−100k
−120k
−140k
−160k

(-0.05, 3.97) (6