# Exercise 1 IML

Eyal Perets 209541903

Based on Recitation 1

1. Prove that orthogonal matrices are isometric transformations. That is, let $T : V \mapsto W$ be some linear transformation and $A$ the corresponding matrix. Show that if $A$ is an orthogonal matrix then $\forall x \in V \;\; ||Ax|| = ||x||$.

1. לפי הגדרה, עבור מטריצה $A$ אורתוגונלית $A^t A = I$ $\quad \forall x \in V$

$$\|Ax\|^2 = (Ax)^t \cdot Ax = x^t A^t A x = x^t x = \|x\|^2$$

$$\|Ax\| = \|x\| \quad (\text{כי } \|\cdot\| \geq 0)$$

2. Calculate the SVD of the following matrix $A$. That is, find the matrices $U, \Sigma, V^\top$ where $U, V$ are orthogonal matrices and $\Sigma$ diagonal.

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

Recall, that to find the SVD of $A$ we can calculate $A^\top A$ to deduce $V, \Sigma$ and then calculate $AA^\top$ to deduce $U$. Equivalently, once we deduced $V, \Sigma$ we can fine $U$ using the equality $AV = U\Sigma$.

2.

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

$$AA^T = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix}$$

$$A = U \Sigma V$$

$$AA^T = U \Sigma V V^t \Sigma^t U^T = U \Sigma \Sigma^T U^T$$

כאשר $\Sigma\Sigma^T$ זו מטריצה אלכסונית, שבאלכסונה של הערכים העצמיים

בהתאם,
כדי נחשב את $U$, שהיא מטריצת הו"ע של $AA^t$

הערכי של $AA^t$ הינם 2 ו-6 (שהם ערכים עצמיים).

וק' עצמי ל- $\lambda=2$

$$Av=\lambda v$$

$$\begin{pmatrix}2&0\\0&6\end{pmatrix}\begin{pmatrix}v_1\\v_2\end{pmatrix}=\begin{pmatrix}2v_1\\2v_2\end{pmatrix}$$

$$\begin{pmatrix}0\\4v_2\end{pmatrix}=0 \qquad v_1=\begin{pmatrix}1\\0\end{pmatrix}$$

$$v_2=0$$
$$v_1=1$$

וק' עצמי ל- $\lambda=6$ הינו

$$Av=\lambda v$$

$$\begin{pmatrix}2&0\\0&6\end{pmatrix}\begin{pmatrix}v_1\\v_2\end{pmatrix}=\begin{pmatrix}6v_1\\6v_2\end{pmatrix}$$

$$\begin{pmatrix}-4v_1\\0v_2\end{pmatrix}=0 \qquad v_1=0,\ v_2=1 \qquad V_2=\begin{pmatrix}0\\1\end{pmatrix}$$

ולכן

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

מציאת המטריצה הנדרשת - $V$:

$$A^tA=\begin{bmatrix}2&0&2\\0&2&-2\\2&-2&4\end{bmatrix}=V^t\Sigma^t\Sigma V$$

$V$ תהיה המטריצה המורכבת מו"ע של $A^tA$

כמו כן נמצא ע"ע של $A^tA$

$$\det(A^tA-\lambda I)=\det\begin{bmatrix}2-\lambda&0&2\\0&2-\lambda&-2\\2&-2&4-\lambda\end{bmatrix}=$$

$$(2-\lambda)\cdot((2-\lambda)(4-\lambda)-4)+2\cdot(-2(2-\lambda))=-\lambda^3+8\lambda^2-12\lambda=0$$

$$\lambda=0 \qquad \lambda=2 \qquad \lambda=6$$

וק עצמי של 0:

$$Av=\lambda v$$

$$\begin{pmatrix}2&0&2\\0&2&-2\\2&-2&4\end{pmatrix}\begin{pmatrix}v_1\\v_2\\v_3\end{pmatrix}=0 \Rightarrow \begin{pmatrix}2v_1+2v_3\\2v_2-2v_3\\2v_1-2v_2+4v_3\end{pmatrix}=0$$

$V_1 = -V_3$

$V_2 = V_3$

$\Downarrow$

$\bar{V}_1 = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} \quad V_1 = \begin{pmatrix} \frac{-1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}$

ו"ע עבור ע"ע 2:

$$\begin{pmatrix} 2V_1 + 2V_3 \\ 2V_2 - 2V_3 \\ 2V_1 - 2V_2 + 4V_3 \end{pmatrix} - \begin{pmatrix} 2V_1 \\ 2V_2 \\ 2V_3 \end{pmatrix} = 0$$

$$\begin{pmatrix} 2V_3 \\ -2V_3 \\ 2V_1 - 2V_2 + 2V_3 \end{pmatrix} = 0$$

$V_3 = 0$

$V_1 = V_2 \Rightarrow \quad \bar{V}_2 = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \\ 0 \end{pmatrix}$

ו"ע עבור ע"ע 6:

$$\begin{pmatrix} 2V_1 + 2V_3 \\ 2V_2 - 2V_3 \\ 2V_1 - 2V_2 + 2V_3 \end{pmatrix} - \begin{pmatrix} 6V_1 \\ 6V_2 \\ 6V_3 \end{pmatrix} = \begin{pmatrix} 2V_3 - 4V_1 \\ -4V_2 - 2V_3 \\ 2V_1 - 2V_2 - 4V_3 \end{pmatrix} = 0$$

$\bar{V}_3 = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} \quad V_3 = \begin{pmatrix} \frac{1}{\sqrt{6}} \\ \frac{-1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{pmatrix}$

$\Downarrow$

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

$\Sigma$ כאמור היא מטריצה אלכסונית שבאלכסונה ערכים הסינגולרים

של $A$, שהם $\sigma_1 = 2$, $\sigma_2 = 6$, ולכן (נשלים כפל מטריצות)

$\Sigma = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 6 & 0 \end{pmatrix}$ , $A_{2\times3} = (U_{2\times2} \cdot \Sigma_{2\times3} V_{3\times3})$

3. Show that the outer product of two vectors $\mathbf{v} \in \mathbb{R}^n, \mathbf{u} \in \mathbb{R}^m$, which is denoted by $\mathbf{v} \otimes \mathbf{u}$ or $\mathbf{v} \cdot \mathbf{u}^\top$ is a matrix $A \in \mathbb{R}^{n \times m}$ with $rank(A) = 1$. That is, show that all rows (or columns) in $A$ are linearly dependent.

3.

$v \in \mathbb{R}^n, \ u \in \mathbb{R}^m$

$$v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \qquad u^t = (u_1, \dots, u_m)$$

$$v \cdot u^T = \begin{bmatrix} v_1 u_1 & v_1 u_2 & \cdots & v_1 u_m \\ \vdots & & & \\ v_n u_1 & \cdots & \cdots & v_n u_m \end{bmatrix}$$

כעת כל שורה $i$ נחלק ב- $v_i$ ונקבל:
(כאשר $v_i \neq 0$ כמובן)

$$= \begin{bmatrix} u_1 & u_2 & \cdots & u_m \\ u_1 & u_2 & \cdots & u_m \\ \vdots & & & \\ u_1 & u_2 & \cdots & u_m \end{bmatrix}$$

(נחסר את השורה הראשונה מכל שאר השורות)

$$= \begin{bmatrix} u_1 & \cdots & \cdots & u_m \\ 0 & & & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix} \rightarrow$$ מטריצה מדרגה 1

ולכן לכל $v, u$ המכפלה החיצונית $v \otimes u$ הינה בדרגה 1.

4. Show that for any otrthonormal basis $(\mathbf{u}_1, \dots, \mathbf{u}_n)$ and any aribtrary vector $\mathbf{x} \in \mathbb{R}_n$ such that $\mathbf{x} = \sum_{i=1}^{n} a_i \cdot \mathbf{u}_i$, it holds that $a_i = \langle \mathbf{x}, \mathbf{u}_i \rangle$ for any $i \in [1, n]$. That is, show that the i'th coefficient of representing $\mathbf{x}$ in the basis $(\mathbf{u}_1, \dots, \mathbf{u}_n)$, is the inner product between $\mathbf{x}$ and $\mathbf{u}_i$.

4.

$$X = \sum_{i=1}^{n} a_i u_i \qquad X_j = \left(\sum_{i=1}^{n} a_i u_i\right)_j = a_j u_j$$

$$\langle X, u_i \rangle = \sum_{i=1}^{n} X_j u_i = \sum_{i=1}^{n} a_j u_j u_i$$

והרי $1 = u_j u_i$ אם $j = i$ ו- 0 עבור $j \neq i$, ולכן

$$= a_i u_i u_i = a_i \cdot \|u_i\|^2 \overset{\text{אורתונורמלי}}{=} a_i$$

5. Let $x \in \mathbb{R}^n$ be a fixed vector and $U \in \mathbb{R}^{n\times n}$ a fixed orthogonal matrix. Calculate the Jacobian of the function $f : \mathbb{R}^n \to \mathbb{R}^n$:

$$f(\sigma) = U \cdot \operatorname{diag}(\sigma) U^\top x$$

Where $\operatorname{diag}(\sigma)$ is an $n \times n$ matrix where

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

0.5

$$f(\sigma) = U \cdot \operatorname{diag}(\sigma) U^T x \qquad \operatorname{diag}(\sigma) = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ \vdots & \sigma_2 & & \vdots \\ 0 & \cdots & & \sigma_n \end{pmatrix}$$

$$\begin{pmatrix} U_{11}\sigma_1 & U_{12}\sigma_2 & \cdots & U_{1n}\sigma_n \\ \vdots & & & \\ U_{n1}\sigma_1 & \cdots & & U_{nn}\sigma_n \end{pmatrix} \cdot U^T x \qquad U^T x = \begin{pmatrix} U_{11} & U_{21} & \cdots & U_{n1} \\ \vdots & & & \vdots \\ U_{1n} & \cdots & & U_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^{n} U_{i1} x_i \\ \vdots \\ \sum_{i=1}^{n} U_{in} x_i \end{pmatrix}$$

$$U \operatorname{diag} \sigma U^T x = \begin{pmatrix} U_{11}\sigma_1 & U_{12}\sigma_2 & \cdots & U_{1n}\sigma_n \\ \vdots & & & \\ U_{n1}\sigma_1 & \cdots & & U_{nn}\sigma_n \end{pmatrix} \begin{pmatrix} \sum_{i=1}^{n} U_{i1} x_i \\ \vdots \\ \sum_{i=1}^{n} U_{in} x_i \end{pmatrix} =$$

$$f(\sigma) = \begin{bmatrix} \sum_{i=1}^{n} U_{1i}\sigma_i \sum_{j=1}^{n} U_{ji} x_j \\ \vdots \\ \sum_{i=1}^{n} U_{ni}\sigma_i \sum_{j=1}^{n} U_{ji} x_j \end{bmatrix}$$

$$f_m = \sum_{i=1}^{n} U_{mi}\sigma_i \sum_{j=1}^{n} U_{ji} x_j$$

$$\frac{\partial f_m}{\partial \sigma_1} = U_{m1} \sum_{j=1}^{n} U_{j1} x_j$$

$$J_\sigma(f) = \begin{bmatrix} U_{11}\sum_{j=1}^{n} U_{j1} x_j & \cdots & U_{n1}\sum_{j=1}^{n} U_{jn} x_j \\ \vdots & & \vdots \\ U_{1n}\sum_{j=1}^{n} U_{jn} x_j & \cdots & U_{nn}\sum_{j=1}^{n} U_{jn} x_j \end{bmatrix}$$

0.5

6. Use the chain rule to calculate the gradient of $h(\sigma) = \frac{1}{2}\|f(\sigma) - y\|^2$

.6

לפי כלל השרשרת: $(f \circ g)' = (f' \circ g) \cdot g'$

נגדיר $g(x) = \frac{1}{2}\|x\|^2$ ו $l(x) = f(\sigma) - y$

$$h(\sigma) = g(l(\sigma))$$

$$h'(\sigma) = g'(l(\sigma)) \cdot l'(\sigma) \qquad g'(\sigma) = \sigma$$

$$l'(x) = f'(\sigma)$$

לכן

$$h'(\sigma) = f'(\sigma) \cdot (f(\sigma) - y)$$

$$= f'(\sigma) \cdot f(\sigma) - y \cdot f'(\sigma)$$

$$\nabla h(\sigma) = h'(\sigma)^T$$

7. Calculate the Jacobian of the softmax function $S : \mathbb{R}^d \to [0,1]^k$

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

.7

נגזור ראשון את המקרה בו $i = j$, כמו שאמרנו ש-$i \neq j$:

$$(S_i)^j = \frac{0\left(\sum_k e^{x_k}\right) - e^{x_i} e^{x_j}}{\left(\sum_k e^{x_k}\right)^2} = -\frac{e^{x_i}}{\sum_{l=1}^{k} e^{x_l}} \cdot \frac{e^{x_j}}{\sum_{l=1}^{k} e^{x_l}} = -S_i S_j$$

(מהגזירה)

עבור $i = j$ נקבל הנגזרת $(S_i)^j = S_i \cdot (1 - S_i)$

ולכן היעקוביאן הינו:

$$J_x(S) = \operatorname{diag}(S) - SS^T$$

8. Let $f : \mathbb{R}^d \to \mathbb{R}$ be defined as $f(x,y) = x^3 - 5xy - y^5$. Calculate the Hessian of $f$.

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

$\partial_x f = 3x^2 - 5y$

$\partial_y f = -5x - 5y^4$

$\partial_{xx} f = 6x$

$\partial_{yy} f = -20y^3$

$\partial_{xy} f = -5$

$$\mathrm{Hess}\, f = \begin{bmatrix} 6x & -5 \\ -5 & -20y^3 \end{bmatrix}$$

### 2.1.3 convexity

Based on Recitation 2

9. Prove that the intersection $C := \bigcap_{i \in I} C_i$ for $\{C_i : i \in I\}$ a collection of convex sets is convex.
10. Prove that the vector sum $C_1 + C_2 := \{c_1 + c_2 : c_1 \in C_1, c_2 \in C_2\}$ of two convex sets is convex.
11. Prove that the set $\lambda C := \{\lambda c : c \in C\}$ is convex, for any convex set $C$, and every scalar $\lambda$.

יהי $u,v \in C$, לכן $\forall i \quad u,v \in C_i$ .9

מכך ש $C_i$ היא קמורה, לכן, לכל $\alpha \in [0,1]$

מתקיים $\alpha v + (1-\alpha)u \in C_i$ לכל $i \in [n]$, ולכן

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

כלומר $C$ קמורה כנדרש.

יהי $u,v \in C_1 + C_2$, ולכן $u = u_1 + u_2$ כך ש .10

$u_1 \in C_1$ ו $u_2 \in C_2$ ו $v = v_1 + v_2$ כך ש $v_1 \in C_1, v_2 \in C_2$.

לכל $\alpha \in [0,1]$ מתקיים

$\alpha v + (1-\alpha)u = \alpha(v_1 + v_2) + (1-\alpha)(u_1 + u_2)$

$= \alpha v_1 + (1-\alpha)u_1 + \alpha v_2 + (1-\alpha)u_2$

ולכן מקמירות $\alpha v_1 + (1-\alpha)u_1 \in C_1$

ו $\alpha v_2 + (1-\alpha)u_2 \in C_2$ ולכן $\alpha v + (1-\alpha)u \in C_1 \cap C_2$

.11 יהי $v, u \in \lambda C$, אזי קיימים $x, y \in C$

כך ש $v = \lambda x$ ו $u = \lambda y$

לכל $\alpha \in [0,1]$ מתקיים

$$\alpha v + (1-\alpha)u = \alpha \cdot \lambda x + (1-\alpha)(\lambda y) = \lambda \cdot (\alpha x + (1-\alpha)y) \in \lambda C$$

12. Let $x_1, x_2, \ldots \overset{iid}{\sim} \mathcal{P}$ be a sample of infinity size drawn from some probability distribution function $\mathcal{P}$ with finite expectation and variance. Show that the sample mean estimator $\hat{\mu}_n = \frac{1}{n}\sum x_i$ calculated over the first $n$ samples is a *consistent estimator* (find the definition in the course book, page 14, Definition 1.1.10 under "Consistency"). Hint: for any given fixed value of $n \in \mathbb{N}$ bound from above the probability of deviating more than $\varepsilon$.

.12 $X_1, X_2, \ldots \sim P$

estimator $\hat{\mu}_n$ של פרמטר $\mu$ הוא consistent estimator אם הוא

מתכנס בהסתברות לערך $\theta$: $\forall \varepsilon > 0 \quad \lim_{n \to \infty} \mathbb{P}(|\hat{\mu}_n - \mu| > \varepsilon) = 0$

ננסה להוכיח שהוא מתכנס בהסתברות אז באמצעות צ'בישב.

נשים לב כי: $E(\hat{\mu}) = E\left(\frac{1}{m}\sum_{i=1}^{m} X_i\right) = \frac{1}{m}\sum_{i=1}^{m} E(X_i) = E(X) = \mu$

ושונות $\hat{\mu}_n$ היא: $\frac{1}{n^2}\sum Var(X_i) = \frac{\sigma^2}{n}$

$$\mathbb{P}(|\hat{\mu}_n - \mu| > \varepsilon) \leq \frac{\frac{\sigma^2}{n}}{\varepsilon} = \frac{\sigma^2}{n \cdot \varepsilon}$$

$$0 \leq \lim_{n \to \infty} \mathbb{P}(|\hat{\mu}_n - \mu| > \varepsilon) \leq \lim_{n \to \infty} \frac{\sigma^2}{n\varepsilon} = 0$$

ולכן לפי סנדוויץ' נקבל כי $\lim_{n \to \infty} \mathbb{P}(|\hat{\mu}_n - \mu| > \varepsilon) = 0$ ולכן estimator הוא

עקבי.

13. Let $\mathbf{x}_1,\dots,\mathbf{x}_m \overset{iid}{\sim} \mathcal{N}(\mu,\Sigma)$ be $m$ observations sampled i.i.d from a multivariate Gaussian with expectation of $\mu \in \mathbb{R}^d$ and a covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$. Provide an expression for the log-likelihood function of $\mathcal{N}(\mu,\Sigma)$. Develop the expression as much as you can. Hint: follow the approach used to derive the likelihood function for the univariate case.

13

פונ' likelihood ל $N(\mu,\Sigma)$ ידוע:

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

$$L(\theta|x_1,\dots,x_d) = f_\theta(x_1,\dots,x_m) = \prod_{i=1}^{m} f_\theta(x_i)$$

$$= \prod \frac{\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)}{\sqrt{(2\pi)^d|\Sigma|}}$$

log-likelihood ידוע:

$$\log\prod \frac{\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)}{\sqrt{(2\pi)^d|\Sigma|}} = \sum_{i=1}^{m}\log\frac{\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)}{\sqrt{(2\pi)^d|\Sigma|}}$$

$$= \sum_{i=1}^{m}\left(-\frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)\right) - \sum_{i=1}^{m}\log\sqrt{(2\pi)^d\cdot|\Sigma|} \quad (*)$$

דטרמיננטה

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

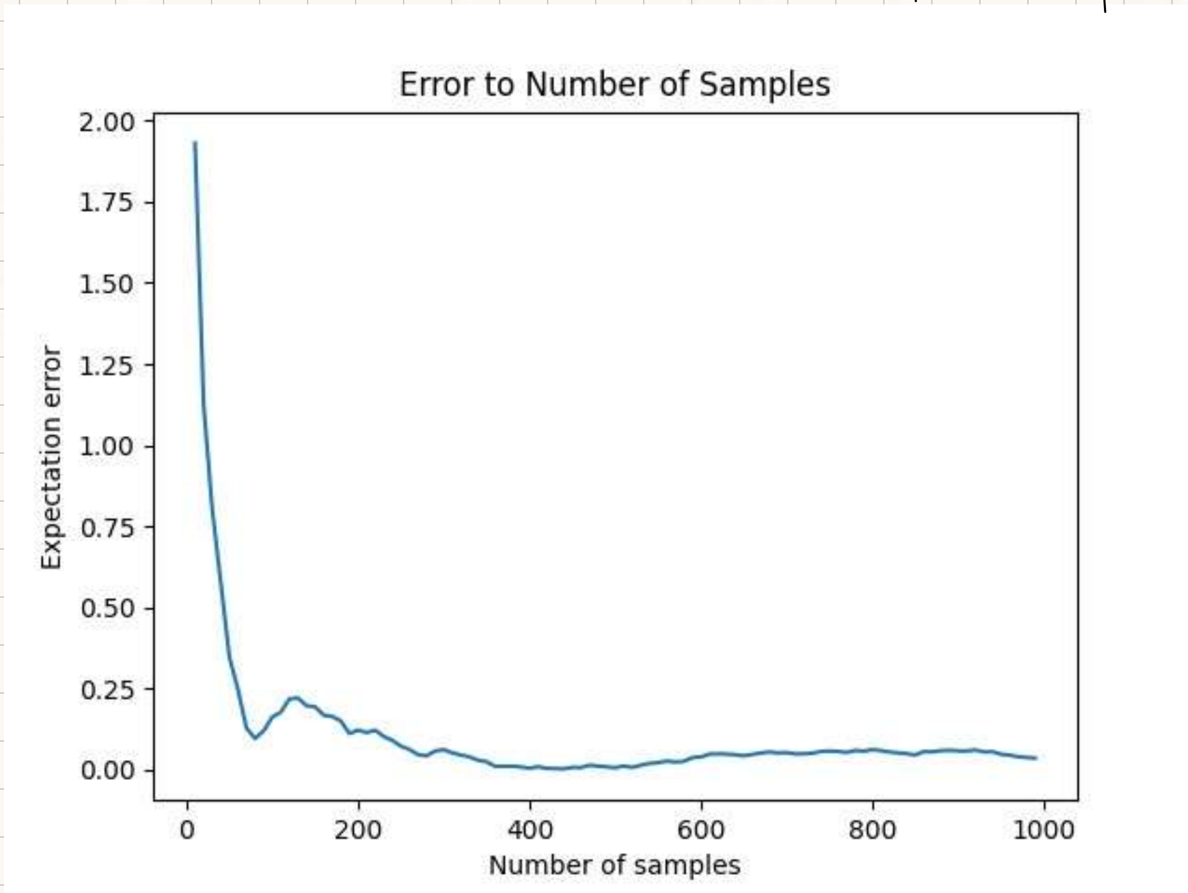
$$(*)\quad \log\sqrt{(2\pi)^d|\Sigma|} = \frac{1}{2}\log(2\pi)^d\cdot|\Sigma| = \frac{1}{2}\log(2\pi)^d + \frac{1}{2}\log|\Sigma|$$

$$\Downarrow$$

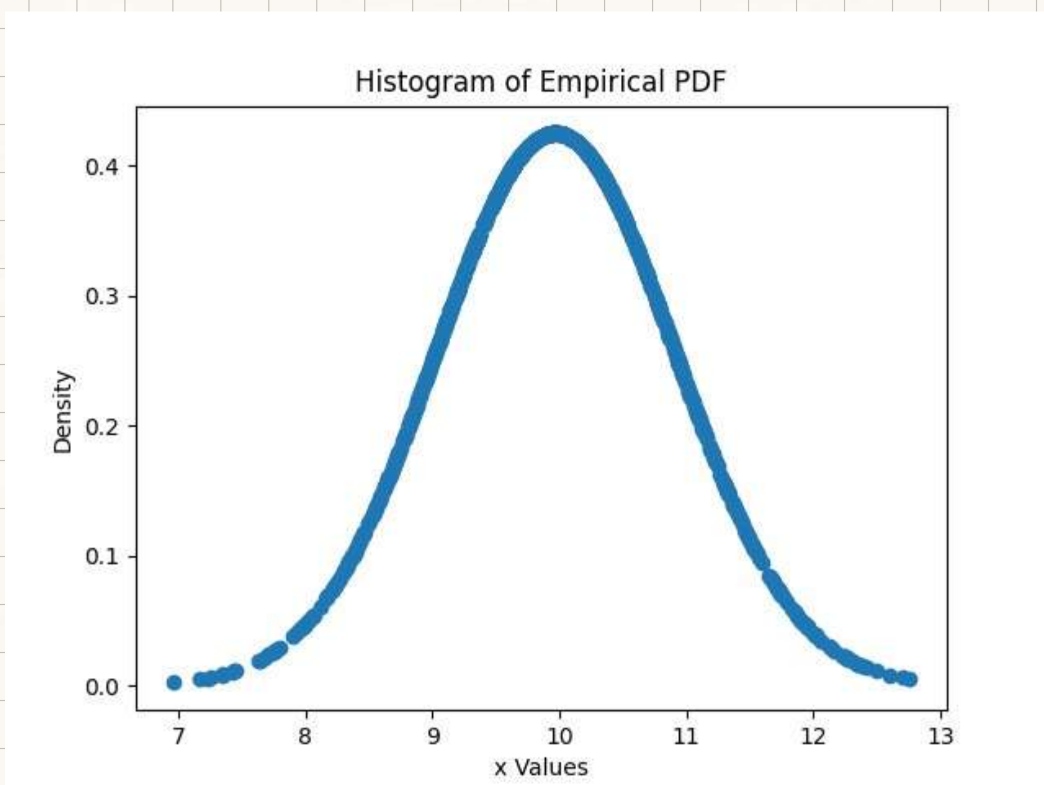
$$\sum_{i=1}^{m}\log\sqrt{(2\pi)^d|\Sigma|} = m\cdot\log\sqrt{(2\pi)^d\cdot|\Sigma|} = \frac{m}{2}\log\left((2\pi)^d\right) + \frac{m}{2}\log|\Sigma|$$

## החלק הפרקטי

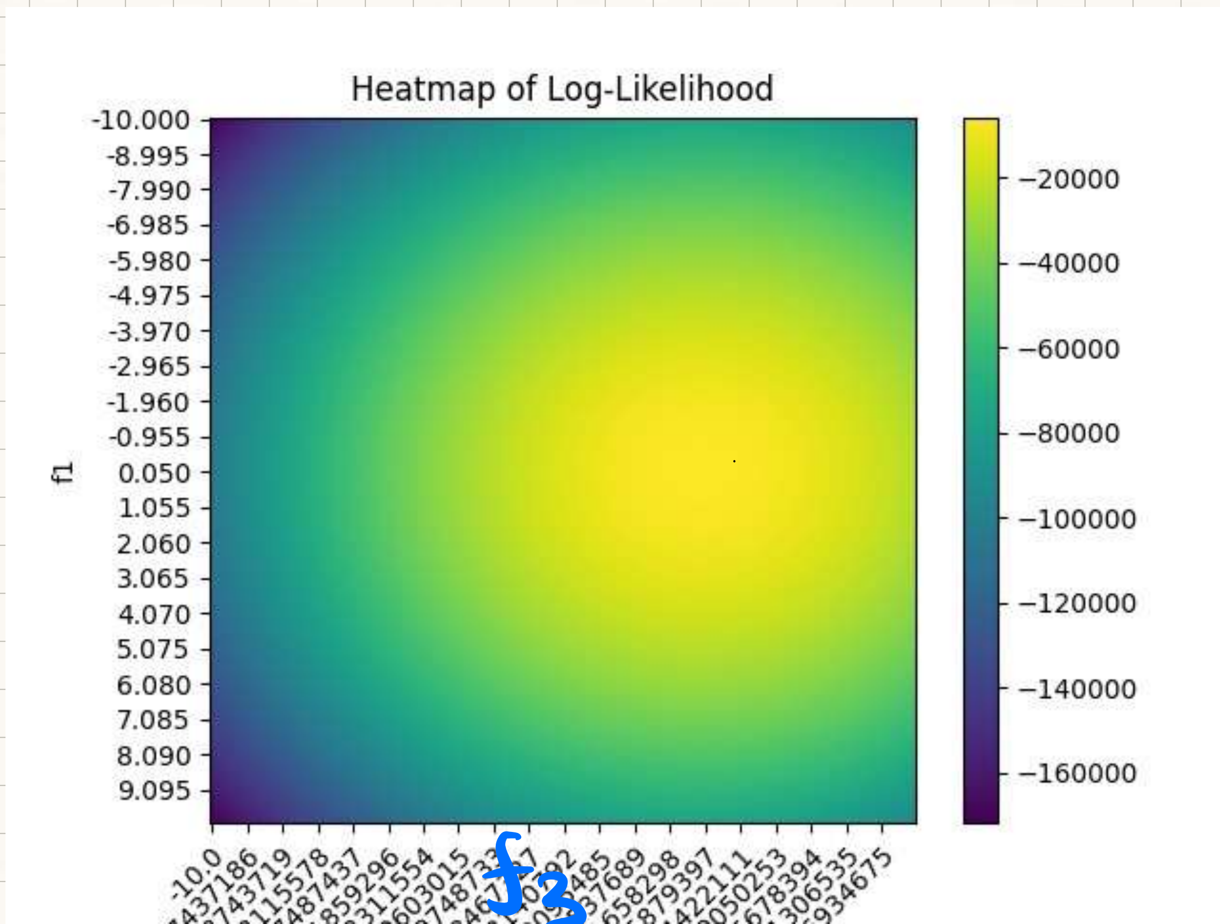
2.




3. אני מצפה לראות היסטוגרמה של התפלגות נורמלית
בעלת תוחלת 10 ותקן סטי' 1 (משמע השונות) ←



5. מהגרף ניתן לראות שככל ש-$f_1$ ו-$f_3$ מתקרבים לערכי $\mu$ האמיתי של הקבועים, כלומר ל- $\mu=(0,0,4,0)^T$, הסבירות (likelihood) מתמקסמת, כלומר הערכים מתקרבים לערכים האמיתיים.




המספרים קרסו את הכותרת של הגרף ולכן הוספתי אותה ידנית (את ה y-label)

6. ה-maximum log-likelihood מתקבל סביב הערכים $f_1=-0.058$
$f_3=3.9551$