מערכות לומדות

לינאריות:

1) יהי W טרנס' לינארית ו-A המטריצה המתאימה. נניח A מטריצה אורתוגונלית,

כך שמתקיים $A^*A = \bar{A}^tA = A^tA = I$

$$\|Ax\| = \langle Ax, Ax\rangle^{\frac{1}{2}} \overset{*}{=} (A\langle x, Ax\rangle)^{\frac{1}{2}} \overset{*}{=} (\bar{A}^tA\langle x, x\rangle)^{\frac{1}{2}}$$

$$= \sqrt{\langle x, x\rangle} = \|x\|$$

* תכונות מכפלה פנימית.

2)

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 2 \end{pmatrix} \quad A^T = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 2 \end{pmatrix} \quad A\cdot A^T = \begin{pmatrix} 2 & 0 \\ 0 & 6 \end{pmatrix}, \quad A^T\cdot A = \begin{pmatrix} 2 & 0 & 2 \\ 0 & 2 & -2 \\ 2 & -2 & 4 \end{pmatrix}$$

$$|AA^T - \lambda I| = \begin{vmatrix} 2-\lambda & 0 \\ 0 & 6-\lambda \end{vmatrix} = (2-\lambda)(6-\lambda) = 0$$

$$\Rightarrow \lambda_1 = 2,\ \lambda_2 = 6$$

$$\lambda_1 = 2: \begin{pmatrix} 0 & 0 \\ 0 & 6 \end{pmatrix} \rightarrow u_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \lambda_2 = 6: \begin{pmatrix} -4 & 0 \\ 0 & 0 \end{pmatrix} \rightarrow u_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

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

$$|A^TA - \lambda I| = \begin{vmatrix} 2-\lambda & 0 & 2 \\ 0 & 2-\lambda & -2 \\ 2 & -2 & 4-\lambda \end{vmatrix} = (2-\lambda)[(2-\lambda)(4-\lambda) - 4] + 2[-2(2-\lambda)] = 0$$

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

$$\Rightarrow (2-\lambda)^2(4-\lambda) - 16 + 8\lambda = (4 - 4\lambda + \lambda^2)(4-\lambda) - 16 + 8\lambda$$

$$= 16 - 4\lambda - 16\lambda + 4\lambda^2 + 4\lambda^2 - \lambda^3 - 16 + 8\lambda = -\lambda^3 + 8\lambda^2 - 12\lambda = 0$$

$$\rightarrow -\lambda(\lambda-6)(\lambda-2) = 0 \quad \lambda_1 = 0,\ \lambda_2 = 6,\ \lambda_3 = 2$$

$$\lambda_1: \begin{pmatrix} 2 & 0 & 2 \\ 0 & 2 & -2 \\ 2 & -2 & 4 \end{pmatrix} \rightarrow V_1 = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} \quad \lambda_2: \begin{pmatrix} -4 & 0 & 2 \\ 0 & -4 & -2 \\ 2 & -2 & -2 \end{pmatrix} \rightarrow V_2 = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} \quad \lambda_3: \begin{pmatrix} 0 & 0 & 2 \\ 0 & 0 & -2 \\ 2 & -2 & 2 \end{pmatrix} \rightarrow V_3 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

$$\Rightarrow V = \begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 2 & 0 \end{pmatrix} \xrightarrow{\text{נרמול}} \begin{pmatrix} -1/\sqrt{3} & 1/\sqrt{6} & 1/\sqrt{2} \\ 1/\sqrt{3} & -1/\sqrt{6} & 1/\sqrt{2} \\ 1/\sqrt{3} & 2/\sqrt{6} & 0 \end{pmatrix}$$

$$AV = \begin{pmatrix} 0 & 0 & \sqrt{2} \\ 0 & \sqrt{6} & 0 \end{pmatrix} = U\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \cdot \Sigma \Rightarrow \Sigma = \begin{pmatrix} 0 & \sqrt{2} & 0 \\ \sqrt{6} & 0 & 0 \end{pmatrix}$$

③

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

נשים לב כי עפ"י הנתונים עבור V סימטרית הם V ו- λ מהלכסון שלה

הוא ע"ע של c_0 מתקיים $C_0^k = V\lambda^k V^{-1} \longleftarrow C_0 = V\lambda V^{-1}$

ונקבל:

$$b_k = \frac{V\lambda^k V^{-1}(\sum a_i v_i)}{\|V\lambda^k V^{-1}(\sum_{i=1}^n a_i v_i)\|} = \frac{\sum_{i=1}^n a_i \lambda^k v_i}{\|\sum_{i=1}^n a_i \lambda^k v_i\|} = \frac{a_1\lambda_1^k\left(v_1 + \frac{a_2}{a_1}\left(\frac{\lambda_2}{\lambda_1}\right)^k v_2 + \dots\right)}{\|a_1\lambda_1^k\left(v_1 + \frac{a_2}{a_1}\left(\frac{\lambda_2}{\lambda_1}\right)^k v_2 + \dots\right)\|}$$

$$= \frac{a_1\lambda_1^k}{\|a_1\lambda_1^k\|} \cdot \frac{v_1 + \frac{a_2}{a_1}\left(\frac{\lambda_2}{\lambda_1}\right)^k + \dots}{\|v_1 + \frac{a_2}{a_1}\left(\frac{\lambda_2}{\lambda_1}\right)^k + \dots\|} \xrightarrow[k\to\infty]{} \frac{a_1\lambda_1^k}{\|a_1\lambda_1^k\|} \cdot \frac{v_1}{\|v_1\|} \overset{*}{=} \pm 1 \cdot \frac{v_1}{\|v_1\|} = \pm v_1$$

1 כי v_1 מנורמל

④

$$f(\sigma) = U\,\mathrm{diag}(\sigma)\,U^T \vec{x} = \left(\sum_{i=0}^n u_i \sigma_i u_i^T\right)\vec{x}$$

$$\Rightarrow \frac{\partial f_j}{\partial \sigma_i} = (u_i u_i^T \vec{x})_j \quad \Rightarrow$$

השורה ה-j

המטריצה ה-k של היעקוביאן היא $\langle u_i, x\rangle u_i$

⑤ עפ"י הגדרה -

$$h(\sigma) = \frac{1}{2}\|f(\sigma) - y\|^2 \qquad \nabla_\sigma(h\circ f) = J_\sigma(h\circ f)^T$$

$$J_\sigma(h\circ f) = J_{f(\sigma)}(h) \cdot J_\sigma(f)$$

$J_\sigma(f)$ - מסעיף קודם

$$J_{f(\sigma)}(h) = (f(\sigma) - y)$$

$$\Rightarrow \nabla_\sigma(h\circ f) = (f(\sigma) - y)^T \cdot \nabla_\sigma(f)$$

⑥

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

$$\frac{\partial S_j}{\partial x_i} = \frac{\frac{\partial e^{x_j}}{\partial x_i} \cdot \sum_{l=1}^k e^{x_l} - \frac{\partial \sum e^{x_l}}{\partial x_i} \cdot e^{x_j}}{(\sum_{l=1}^k e^{x_l})^2} = \begin{cases} -e^{x_i} e^{x_j} / (\sum_{l=1}^k e^{x_l})^2 & i \neq j \\ \frac{e^{x_i} \cdot \sum e^{x_l} - e^{2x_i}}{(\sum_{l=1}^k e^{x_l})^2} & i = j \end{cases}$$

נשים לב כי קיבלנו:

$$\boxed{\frac{\partial S_j}{\partial x_i} = \begin{cases} -S_i \cdot S_j & i \neq j \\ S_j(1 - S_j) & i = j \end{cases}}$$

$$f(x,y) = x^3 - 5xy - y^5 \rightarrow \nabla f = \begin{pmatrix} 3x^2 - 5y \\ -5y^4 - 5x \end{pmatrix}$$ (7)

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

$$\hat{u}_n = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad E[\hat{u}_n] = \frac{1}{n}E[\Sigma x_i] = \frac{1}{n}\sum_{i=1}^{n} E[X] = u$$ (8)

לפי א' שוויון מרקוב:

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

$$Var(|\hat{u}_n - u|) = E[|\hat{u}_n - u|^2] - E^2[|\hat{u}_n - u|] \geq 0$$

$$\Rightarrow E^2[|\hat{u}_n - u|] \leq E[|\hat{u}_n - u|^2] = E[(\hat{u}_n - u)^2] = E[(\hat{u}_n - E[\hat{u}_n])^2] = Var(\hat{u}_n)$$

$$= Var\left(\frac{1}{n}\Sigma x_i\right) = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}$$

$$\Rightarrow E[|\hat{u}_n - u|] \leq \frac{\sigma}{\sqrt{n}} \Rightarrow 0 \leq P(|\hat{u}_n - u| \geq \varepsilon) \leq \frac{\sigma}{\varepsilon\sqrt{n}}$$

$$\Rightarrow \lim_{n\to\infty} P(|\hat{u}_n - u| \geq \varepsilon) \leq \lim_{n\to\infty}\left(\frac{\sigma}{\varepsilon\sqrt{n}}\right) = 0$$

$$\Rightarrow \lim_{n\to\infty}(P(|\hat{u}_n - u| \geq \varepsilon)) = 0 \quad \blacksquare$$

$$\mathcal{L}(\theta|\bar{x}_i) = f_\theta(\bar{x}_i) = \frac{1}{(2\pi)^d|\Sigma|} \cdot e^{-\frac{1}{2}\cdot(\bar{x}_i - \bar{u})^T \Sigma^{-1}\cdot(\bar{x} - \bar{u})}$$ (9)

$$\Rightarrow \mathcal{L}(\theta|x_1 \ldots x_m) = \prod_{i=1}^{m} \frac{1}{(2\pi)^d|\Sigma|} \cdot e^{-\frac{1}{2}(\bar{x}_i - \bar{u})^T \Sigma^{-1}(\bar{x}_i - \bar{u})} = \frac{1}{[(2\pi)^d|\Sigma|]^m} \cdot e^{-\frac{1}{2}\sum_i^m (x_i - u)^T \Sigma^{-1}(x_i - u)}$$

$$\log(\mathcal{L}) = \log\left[\frac{1}{[(2\pi)^d|\Sigma|]^m} \cdot e^{-\frac{1}{2}\Sigma(x_i - u)^T \Sigma^{-1}(x_i - u)}\right]$$

$$= \log([(2\pi)^d|\Sigma|]^{-m}) + \left(-\frac{1}{2}\sum_{i=1}^{m}(\bar{x}_i - \bar{u})^T \Sigma^{-1}(\bar{x}_i - \bar{u})\right)$$

$$= -m \cdot \log((2\pi)^d|\Sigma|) - \frac{1}{2}\sum_{i=1}^{m}(\bar{x}_i - \bar{u})^T \Sigma^{-1}(\bar{x}_i - \bar{u})$$

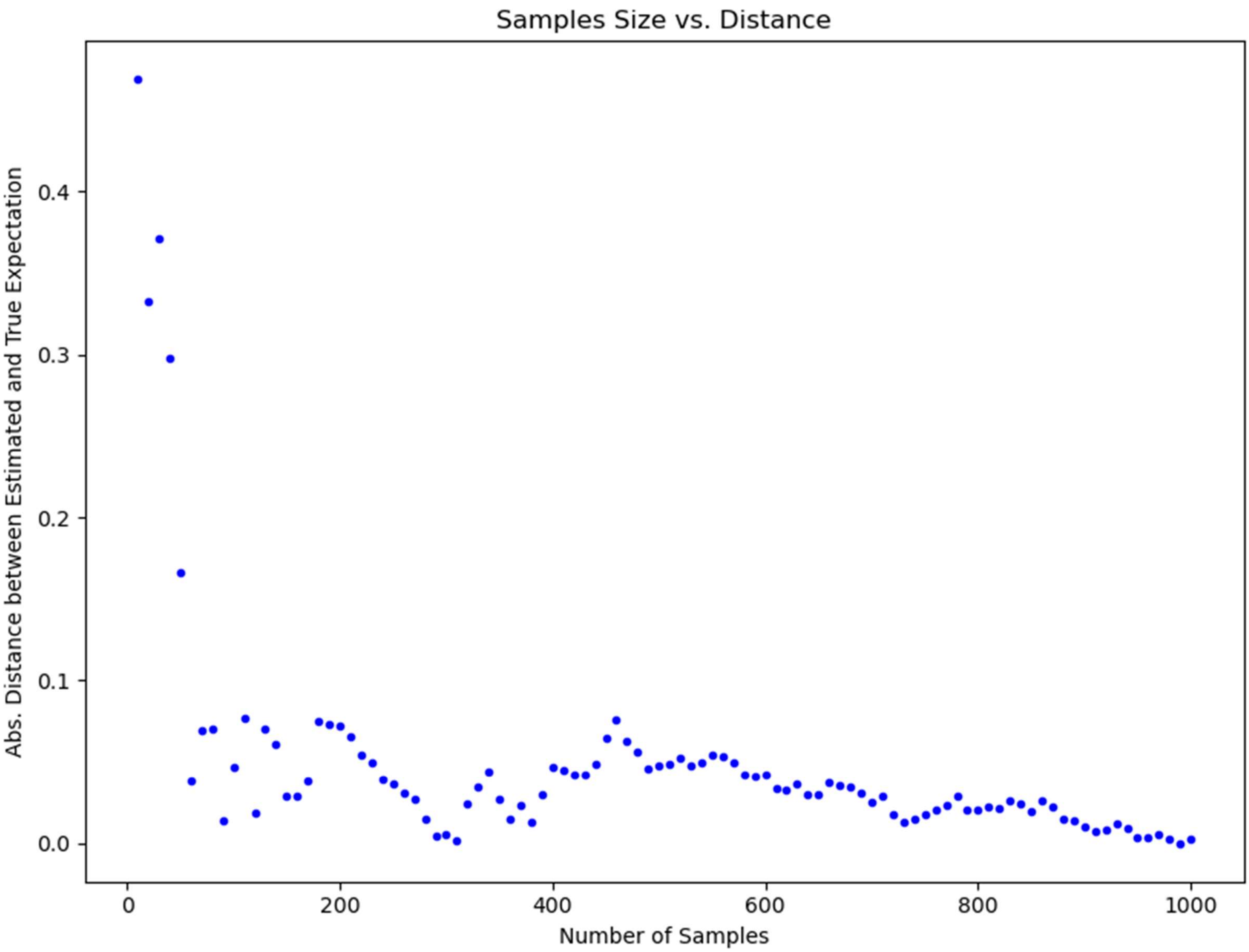
$$= -md\log(2\pi) - m\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{m}(\bar{x}_i - \bar{u})^T \cdot \Sigma^{-1}(\bar{x}_i - \bar{u})$$

<u>חלק מעשי</u>

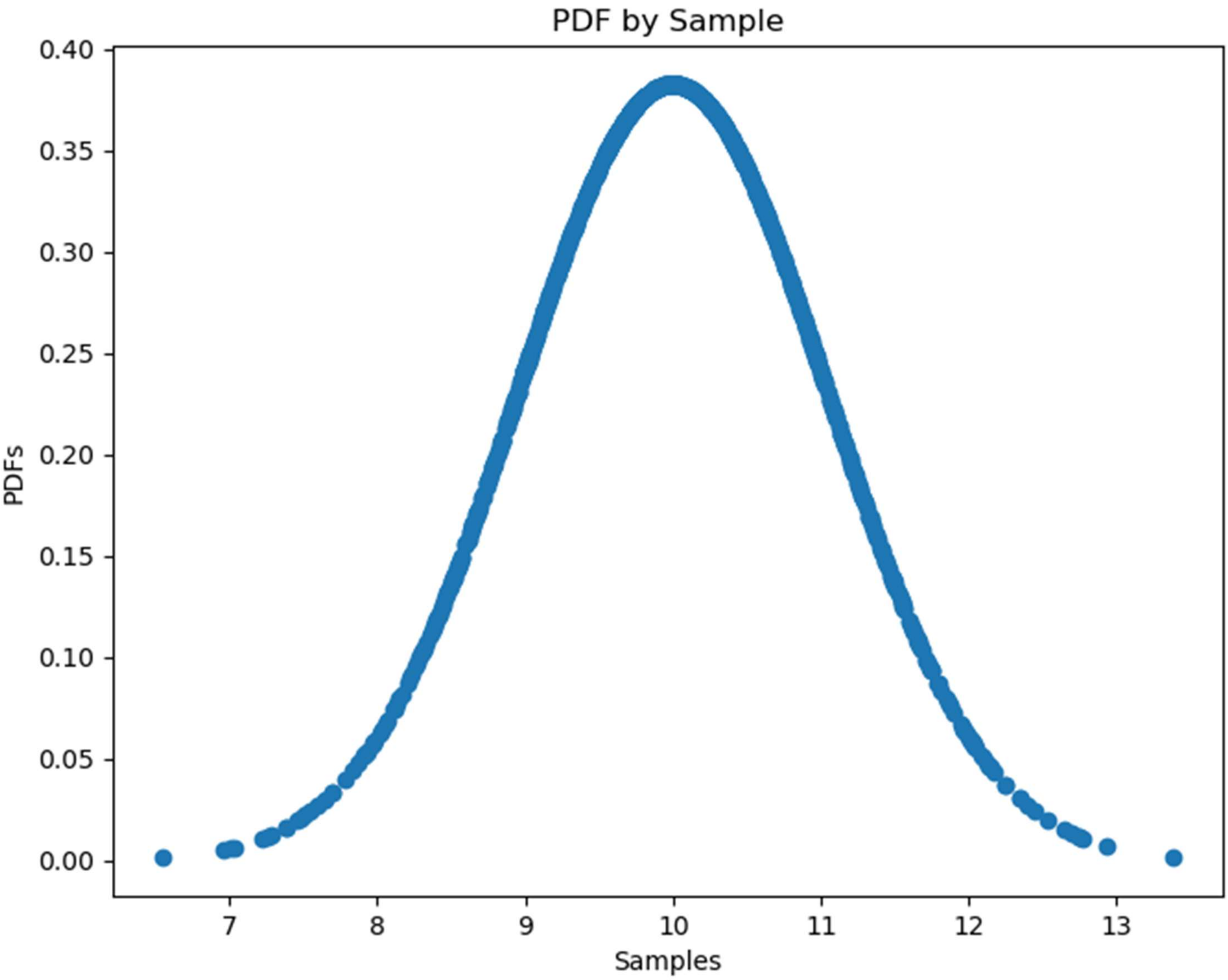
1) תוחלת ושונות שהותאמו (רלוונטי להרצה ספציפית)

(9.997053997632051, 1.087897120940502)

2) גרף שהתקבל עבור התאמות שונות:



3) דגמנו משתנים ממרחב עם התפלגות גאוסיאנית ולכן נצפה שפונקציית הצפיפות תיראה כמו פעמון גאוס – מקבלת ערכים גבוהים סביב התוחלת וברוחב המתאים לשונות. הגרף שהתקבל:



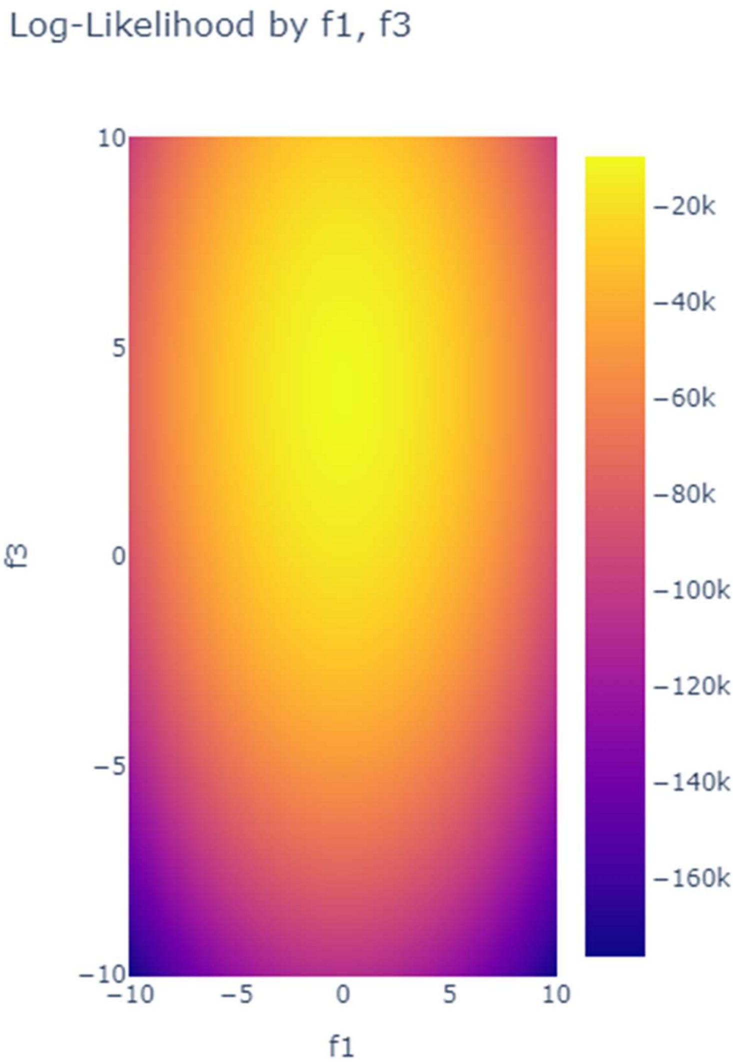
4) תוחלת:

[-0.008687   -0.03455066 4.00620038 0.0079784]

שונות משותפת:

```
[[ 0.92618336  0.12987336 -0.0233973   0.46758105]
 [ 0.12987336  1.85799249 -0.0417188  -0.07496864]
 [-0.0233973  -0.0417188   1.03177586  0.01623343]
 [ 0.46758105 -0.07496864  0.01623343  0.97866918]]
```

5) נצפה שאזורים עם ערכי f1, f3 עם סבירות (Likelihood) גבוהה יותר יהיו "חמים" יותר, דהיינו סביב f1=0, f3=4 (הערכים האמיתיים) נצפה לראות צבעים חמים וככל שמתרחקים מהערכים האלה הצבעים יתקררו. גרף:



f1: -0.05 , f3: 3.97 **(6**