מבוא לתורת המידע הקוונטית

רוני נהבו - אלגוריתמים

אלגברה לינארית:

יהי A מ"ל אוניטרי.

אזי נוכיח כי לכל x מתקיים $\|Ax\| = \|x\|$

הוכחה: נעבור בראיה ונקבל.

יהי $\beta = (b_1, \dots, b_n)$ בסיס אורתונormלי ל-V

אזי ראינו כי מ"ל המקיים אוניטריות ניתן לכתוב

$$A = \sum_{i=1}^{n} b_i \otimes b_i =$$

כמו כן $\|x\| = \sqrt{\langle x|x\rangle}$

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

תכונות מכפלה סקלרית.

מבקרים לינארית ב ראינו כי התכונה Q מ"ל אוניטרי הינה המשמרת שטח.

נוסף לכך נוכיח זאת גם בצורה יותר, מכיוון שמטריצת A הינה בסיס אורתונormלי אזי

$$(A^T A)_{ij} = u_i^T u_j = \langle u_i | u_j \rangle$$

$\Leftarrow$ מכיוון ש $u_1, \dots, u_n$ בסיס אורתונormלי מתקיים כי

$$\langle u_i | u_j \rangle = \delta_{ij}$$

אם $i = j$ אחרת 0, כלומר בדיוק.

$$\Leftarrow \quad A^T A = I_n$$

ולכן $= x^T A^T A x = x^T I x = x^T x = \langle x | x \rangle =$

$= \|x\|^2$

$\Leftarrow$ ולכן לפי הגדרת נורמה $\|Ax\|^2 = \|x\|^2$

$\Downarrow$

$\|Ax\| = \|x\|$

כנדרש!

ב) תהי $A = \begin{pmatrix} 1 & 1 & 0 \\ 2 & 1 & -1 \end{pmatrix}$ ותסמן את ה-SVD של A ע"י U, Σ, V^T כפי שראינו בהרצאה.
חשבו ותסמנו ע"י $A^T A$

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

נחשב ע"פ:

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

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

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

$$= \dots = -\lambda^3 + 8\lambda^2 - 12\lambda = -\lambda(\lambda-2)(\lambda-6) = 0$$

$\lambda = 6, 2, 0$ // ⇐

⇐ ערכים ע"

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

<u>וקטורים</u> עצ:

$$A^T A \cdot v = 0 \Rightarrow \begin{cases} v_1 + v_3 \\ v_2 - v_3 \\ v_1 - v_2 + 2v_3 \end{cases} = 0$$

$$v_1 = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$$ ⇐

$$(A - 6I) \cdot v = 0 \Rightarrow \begin{pmatrix} -4 & 0 & 2 \\ 0 & -4 & -2 \\ 2 & -2 & -2 \end{pmatrix} v$$

$$\Rightarrow v_2 = \begin{bmatrix} 1/2 \\ -1/2 \\ 1 \end{bmatrix}$$

$$(A - 2I) v = 0 \Rightarrow \begin{pmatrix} 0 & 0 & 2 \\ 0 & 0 & -2 \\ 2 & -2 & 2 \end{pmatrix} v \Rightarrow$$

$$v_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

⇐ <u>נורמלים</u>:

$$\|v_1\| = 3$$
$$\|v_2\| = 3/2 \Rightarrow V = \begin{bmatrix} -1/\sqrt{3} & \frac{\sqrt{2}}{2\sqrt{3}} & 1/\sqrt{2} \\ 1/\sqrt{3} & -\frac{\sqrt{2}}{2\sqrt{3}} & 1/\sqrt{2} \\ 1/\sqrt{3} & \sqrt{2}/\sqrt{3} & 0 \end{bmatrix}$$
$$\|v_3\| = 2$$

⇐ לא <u>נורמלים</u>!

כעת נחשב את U לפי הנוסחה שבא מן הוקטורים

הינו

$$u_i = \frac{1}{\sigma_i} A \cdot v_i$$

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

איפוס כיוון אבר מקביל.

$$u_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 2 \end{bmatrix} \begin{bmatrix} 1/\sqrt{6} \\ -1/\sqrt{6} \\ \sqrt{2}/3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

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

וסה"כ קיבלנו!

$$A = U \Sigma V^T = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \sqrt{6} & 0 & 0 \\ 0 & \sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} -1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{6} & -1/\sqrt{6} & \sqrt{2}/3 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \end{bmatrix}$$

$\blacksquare$

ג) יהי A מ"ש כלשהי ו- $C_0 = A^T A$, יהי $\lambda_1 \geq \dots \geq \lambda_n$ הע"ע של C_0 עם ו"ע $v_1, \dots, v_n$ בהתאמה. (v_i אורתונורמלי)
יהי b_k וקטור המוגדר באופן הבא:

$$b_0 = \sum_{i=1}^{n} a_i v_i \quad , \quad b_{k+1} = \frac{C_0 b_k}{\|C_0 b_k\|}$$

ויהי אורתונורמלי.

הוכיח $a_1 \neq 0$ אז $\lim_{k\to\infty} b_k = \pm v_1$

הוכחה: נפרק את ה הנוסחה:

$$b_{k+1} = \frac{C_0 b_k}{\|C_0 b_k\|} = \frac{C_0 \cdot \frac{C_0 b_{k-1}}{\|C_0 b_{k-1}\|}}{\left\| C_0 \cdot \frac{C_0 b_{k-1}}{\|C_0 b_{k-1}\|} \right\|} =$$

$$= \frac{\frac{C_0^2 b_{k-1}}{\cancel{\|C_0 b_{k-1}\|}}}{\frac{\|C_0^2 b_{k-1}\|}{\cancel{\|C_0 b_{k-1}\|}}} = \frac{C_0^2 b_{k-1}}{\|C_0^2 b_{k-1}\|} = \dots =$$

תר נירמל ע"מ שהם יהיו שווים.

$$= \frac{C_0^k b_0}{\|C_0^k b_0\|}$$

תיצב באופן דקורסיבי?

יהי $b_0 = \sum_{i=1}^{n} a_i v_i$ אז ע"ז אבר. בס"ס אורתונורמלי.

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

הגדרת ל"ע.

$$= a_1 \lambda_1^k \left(v_1 + \frac{a_2}{a_1} \left(\frac{\lambda_2}{\lambda_1}\right)^k v_2 + \dots + \frac{a_n}{a_1} \left(\frac{\lambda_n}{\lambda_1}\right)^k v_n \right)$$

$\Leftarrow$ כאן נראה שזה כי $\left(\frac{\lambda_i}{\lambda_1}\right)^k$ שאיפה אפסון לינדון

כי הע"ע עסוריות בסדר יורד.

$\Leftarrow$ כאשר נשאיף את $k \to \infty$ נקבל כי הגבול הוא ש.ס.

$\Leftarrow$

$$a_1 \lambda^k v_1 + 0 \dots + 0 = a_1 \lambda^k v_1$$

ובגלל אותו דבר תדר בעזרת הנורמה ונקבל

$$= \lim_{k \to \infty} \frac{a_1 \lambda^k v_1}{\underbrace{\|a_1 \lambda^k v_1\|}_{\text{דומה}}} \overset{\text{הגדרת נורמה}}{=} \lim_{k \to \infty} \frac{\overset{\pm 1}{\cancel{a_1}} \lambda^k v_1}{\|v_1\| \cancel{|a_1 \lambda^k|}} =$$

$$= \lim_{k \to \infty} \pm \frac{v_1}{\underbrace{\|v_1\|}_{1}} = \pm v_1$$

גבול נורמלי

כנדרש.

$\blacksquare$

multivariable calculus:

4) יהי $x \in \mathbb{R}^n$ וקבוע קבוע ו- $U \in \mathbb{R}^{n \times n}$ גם אורתוגונלי.
נחשב את יעקוביאן של הפונקציה $f: \mathbb{R}^n \to \mathbb{R}^n$

$$f(\sigma) = U \cdot \mathrm{diag}(\sigma) U^T \cdot x$$

כאשר $\mathrm{diag}(\sigma) = \begin{cases} \sigma_i & i=j \\ 0 & \text{אחרת} \end{cases}$

סרטוטין: ראשית עלינו לחשב את המטריצה ניסה עלינו כי $J_x(f) \in \mathbb{R}^{n \times n}$

נרצה עלינו לחשוב על לבדוק את תוצאת הכפל f_i
אנחנו כמה פעמים עצור עם ושלוה אחר, הרי ברור כי
כל שורה ב-J היא שאלת הגרדיאנט עם הפונקציה f_i

$$f(\sigma) = U \,\mathrm{diag}(\sigma) U^T x =$$

$$U \cdot \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} U^T \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

$U = a_{ij} = a_{ji}$

$$u = U^T \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{n} a_{ij} x_j \end{pmatrix} =$$

u סקלר

$$A\,\mathrm{diag}(\sigma) = \begin{bmatrix} a_{1,1} & \cdots & a_{1,k} \\ \vdots & \ddots & \vdots \\ a_{k,1} & \cdots & a_{k,k} \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{bmatrix} =$$

$$= \begin{bmatrix} \sigma_1 a_{1,1} & \cdots & \sigma_1 a_{k,k} \\ \vdots & \ddots & \vdots \\ \sigma_k a_{k,1} & \cdots & \sigma_k a_{k,k} \end{bmatrix} \begin{bmatrix} \vdots \\ \sum_{j=1}^{k} a_{i,j} x_j \\ \vdots \end{bmatrix} =$$

$$= \begin{bmatrix} | \\ \sum_{t=1}^{k} \sigma_j a_{j,t} \cdot \sum_{i=1}^{n} a_{t,i} x_i \\ | \end{bmatrix} =$$

$$= \begin{bmatrix} | \\ \sigma_j \sum_{t=1}^{k} \sum_{i=1}^{n} a_{j,t} a_{t,i} x_i \\ | \end{bmatrix}$$ מה <u>אורך</u> ה.

ירדה נתון שגזרה:

$$\frac{df_p}{d\sigma_l} = \begin{cases} \sum_{t=1}^{k} \sum_{i=1}^{n} a_{l,t} \cdot a_{t,i} x_i & p = l \\ 0 & p \neq l \end{cases}$$

קל היאקובין היה גם אלכסוני (בגלל כי כל פונק' סקלר ארכ) ולכן

$$J_\sigma f = \begin{bmatrix} \sum_{t=1}^{k}\sum_{i=1}^{n} a_{1,t}a_{t,i}x_i & 0 & \cdots & 0 \\ 0 & \ddots & & 0 \\ 0 & & \ddots & 0 \\ 0 & \cdots & 0 & \sum_{t=1}^{k}\sum_{i=1}^{n} a_{n,t}a_{t,i}x_i \end{bmatrix}$$

ג) נגדיר כפלת השגיאה ונ לדיסק את הבדיקה מה

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

נגדיר:

$$g(\sigma) = f(\sigma) - y$$

$$h(\sigma) = \frac{1}{2}\|g(\sigma)\|^2$$

נגדיר בכיוונים שרבינו בהגדלת ונקט

$$\nabla h(\sigma) = \nabla_{g(\sigma)} h \cdot \nabla_\sigma g =$$

ראינו כי ה גרדיאנט מה הפונק $\frac{1}{2}\|g(\sigma)\|$ הינו $g(\sigma)^T$

ולכן:

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

6) נגזור את הנגזרת של פונ' softmax

כמראה לעיל כי, בנקודה מ $i=j$:

$$\frac{dS_i}{dx_j} = S_i(1-S_j)$$

כעת נרצה לגזור בנקודה בו $i \neq j$ לפי הגזרת את הנגזרת:

$$\frac{dS_i}{dx_j} = \frac{d}{dx_j} \cdot \frac{e^{x_i}}{\sum_{k=1}^{d} e^{x_k}}$$

ונגזור מחדש:

0 מה קבוע

$$\frac{(\cancel{e^{x_i}})' \cdot \sum_{k=1}^{d} e^{x_k} - (e^{x_i}) \cdot \left(\sum_{k=1}^{d} e^{x_k}\right)'}{\left(\sum_{k=1}^{d} e^{x_k}\right)^2} =$$

לוגיקה בנגזרת

?

$$= \frac{-e^{x_i} \cdot \sum_{k=1}^{d} (e^{x_k})'}{\left(\sum_{k=1}^{d} e^{x_k}\right)^2} = \frac{-e^{x_i} \cdot e^{x_j}}{\left(\sum_{k=1}^{d} e^{x_k}\right)^2} =$$

פ מה שלא x_j אכפ

$$= -S_i \cdot S_j$$

$\Leftarrow$

$$J_{z(j,i)} f = s_i \left(\delta_{i,j} - s_j \right)$$

וכלל לב:

אלו $j=i$ אז תקבל $s_i(1-s_i)$ כפי שרבנו

בקרבת, אחרת $-s_i s_j = s_i(0-s_j)$ כנדרש.

7) תהי $f(x,y) = x^3 - 5xy - y^2$ נחשב את הסיאן של

f.

בנקודה זה נצפה שהסיאן יהיה מטריצה 2×2

$$H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial^2 x} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} =$$

$$\frac{\partial^2 f}{\partial^2 x} = \left(3x^2 - 5y\right)' = 6x$$

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

נשר אינו

הכונה זו לגזרה בקיום $\mathcal{C}^2$.

$$\frac{\partial^2 f}{\partial^2 y} = \left(-5y^4 - 5x\right)' = -20y^3$$

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

$\Leftarrow$

Estimation Theory:

תהי $X_1, X_2, \dots$ דגימות מן התפלגות P על תחלת ושונות סופיות.

נראה כי האומד לתוחלת $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$ הוא קונסיסטנטי.

הגדרה: נתבונן בהגדרה θ אומד קונסיסטנטי

$$\lim_{n\to\infty} P(|\theta_n - \theta| > \varepsilon) = 0 \quad \forall \varepsilon > 0$$

כלומר נצביע בהסתברות גבוהה ונמוכה ש"ג שכיחות קרובה ל"ג.

יהי סכום ויהי $n \in \mathbb{N}$,

הקבוע אלפי אוניברסלי

$$\theta_n = \hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} x_i \quad , \quad \hat{\mu} \sim P$$

ונבדוק:

$$P(|\hat{\mu} - \hat{\mu}_n| > \varepsilon)$$

נחפש חסמים בש"ג שכיחות ודרך שכיחות מהצורה

$$|X - E(X)|$$

ונניח שכבר כי בהינתן $\mu = E(x_i)$ עבור כל i, מבוים ב- P x_i מתקיים:

$$E(\hat{\mu}_n) = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}\cdot \mu n = \mu$$

$\Leftarrow$ נוכל לחשב את ש"ג שכיחות ומכן

$$P(|\hat{\mu} - \hat{\mu}_n| > \varepsilon) < \frac{\mathrm{Var}(\hat{\mu}_n)}{\varepsilon^2} =$$

$$= \frac{\mathrm{Var}\left(\frac{1}{k}\sum_{i=1}^{k} x_i\right)}{\varepsilon^2} = \frac{\frac{1}{k^2}\cdot k\,\mathrm{Var}(x_i)}{\varepsilon^2} =$$

סכום שונות
ב"ת.

$$= \frac{1}{k}\cdot\frac{\mathrm{Var}(x_i)}{\varepsilon^2}$$

כלומר נקבל בהתאם לאבסולוטי ולשונות בגשפס הסנדוויץ' ולקבל:

$$\lim_{k\to\infty} \mathbb{P}\left(|\hat{\mu}-\hat{\mu}_k| > \varepsilon\right) < \frac{1}{k}\underbrace{\frac{\mathrm{Var}(x_i)}{\varepsilon^2}}_{\text{קבוע}}$$

$$\xrightarrow{k\to\infty} 0$$

$$\Rightarrow \lim_{k\to\infty} \mathbb{P}\left(|\hat{\mu}-\hat{\mu}_k| > \varepsilon\right) = 0$$

דרך $\hat{\mu}$ קונסיסטנטי לפי הגדרה כלומר!

דוגמה

אם יהיו $x_1,\dots,x_k \overset{iid}{\sim} \mathcal{N}(\mu,\Sigma)$ אז הריבוע באר
$\mu \in \mathbb{R}^d$ וקיור הקובאריאנס ו- Σ גם הסווגות.
ומכך שהוא פונקציה לוגריתם-לייקליהוד עבור $\mathcal{N}(\mu,\Sigma)$.
ויכבר כי אבור מקרה רב-מימדי נוראלי מתקבל:

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

$\forall i \in [k]$

כאן הוא בדיוק הדבר שבו כ $\Theta = \{\mu, \Sigma\}$

$$L(\Theta, x_1, \dots, x_n) = \prod_{i=1}^{n} f(x_i)$$

נזכר:

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

כאן נפעיל $\log$ על הביטוי:

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

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

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

$$= \frac{n}{2}\left(\log\left(\frac{1}{(2\pi)^d}\right) + \log\left(\frac{1}{|\Sigma|}\right)\right) - \dots =$$

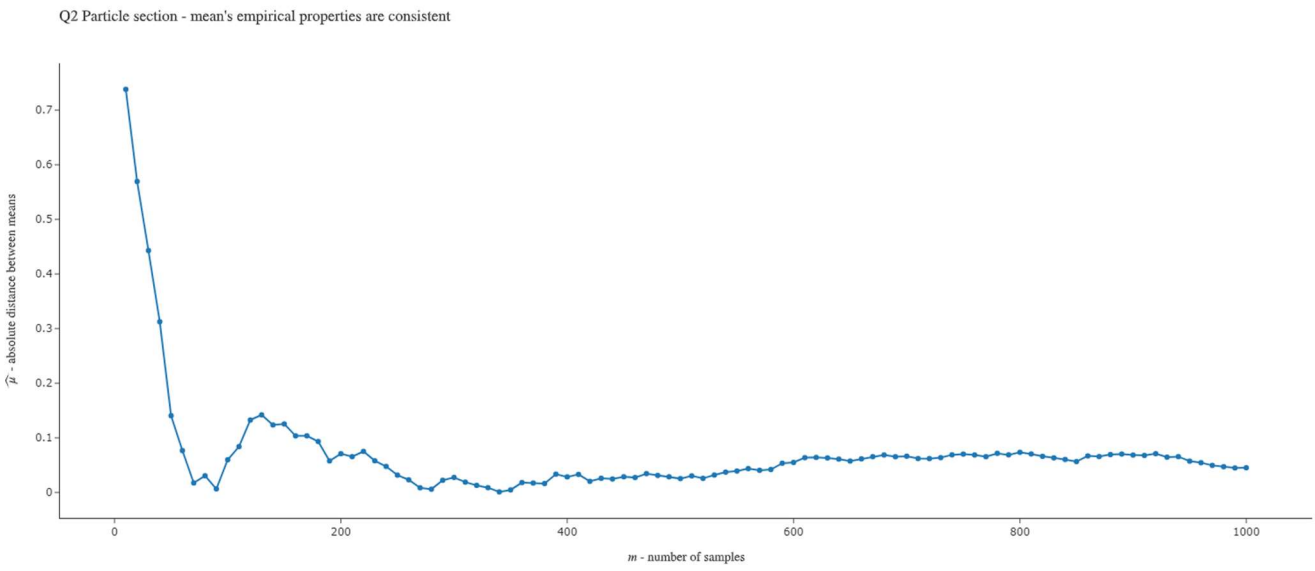
$$= \frac{n}{2}\left(-d\log(2\pi) - \log(|\Sigma|)\right) - \frac{1}{2} \sum \dots =$$

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

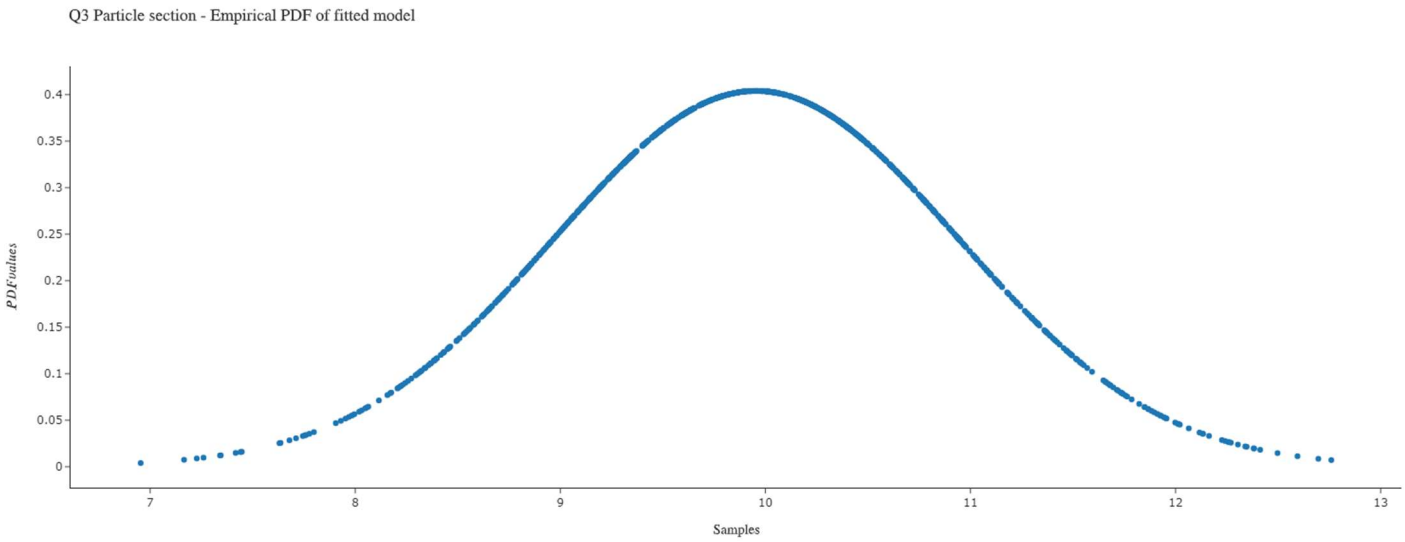
מבוא למערכות לומדות 67577
רועי זהבי 208648154

פתרונות לחלק המעשי - תרגיל 1

1. לאחר הרצת התוכנית יודפס למסך (9.954743292509804, 0.9752096659781339).
2.

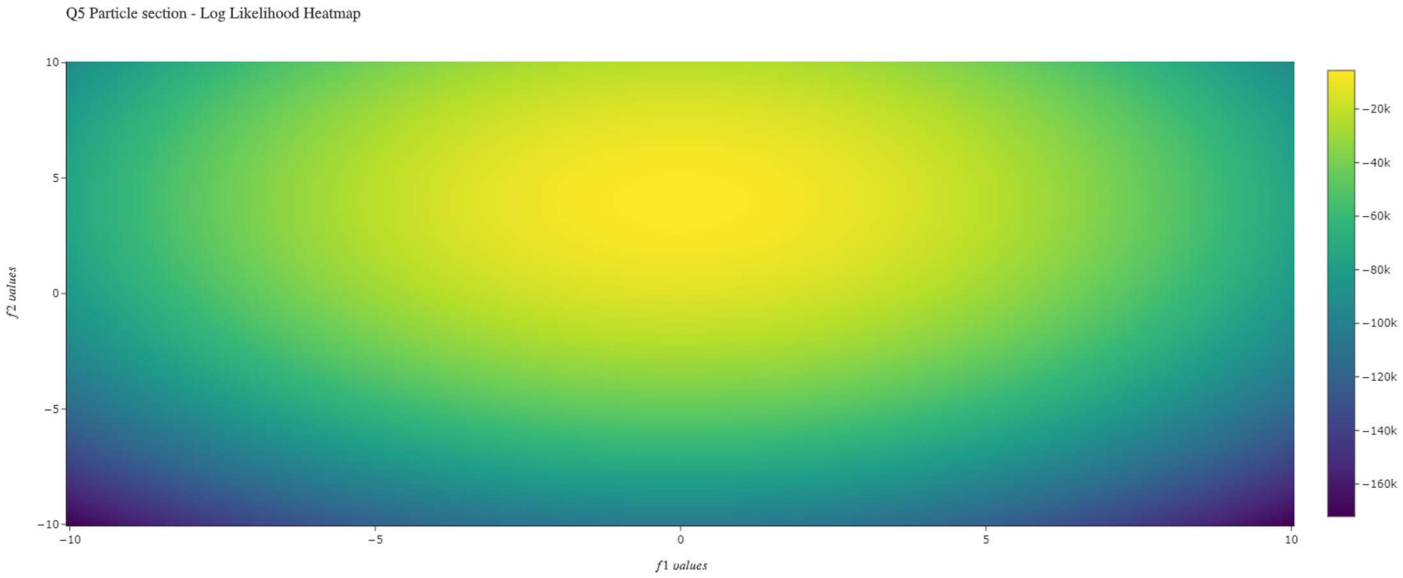


3. בגרף זה נצפה לראות את תצורת הפעמון שאנו מצפים מהתפלגות נורמלית חד-ממדית, הרי אנו מחשבים עבור כל דגימה את ערך הצפיפות תחת שימוש באומדים unbiased כפי שראינו בהרצאה. לכן, נצפה לקבל פלט שעבור כמות דגימות השואפת לאינסוף מתכנס לערכי הצפיפות המקוריים.



ניתן לראות כי זהו אכן המצב בגרף זה, כאשר עבור 1000 דגימות קיבלנו ערכי צפיפות ששואפים לפונקציית הצפיפות המקורית של ההתפלגות הנורמלית.

4. מודפס למסך כנדרש ע״י התוכנית.
5. נתבונן במפת החום המופקת ע״י הנ״ל.



נוכל ללמוד מפלט כי אילו היינו מסתכלים על תמונת הערכים ברב מימד הינו מקבלים מעין פרבולואיד, כאשר נקודת המקסימום שלו הינו באזור $f1 = 3\sim, f2 = 0\sim$. נרצה להשתמש במסקנה זו כדי למצוא את הקלטים שממקסמים את ערכה של פונקצית הנראות ולהשתמש באלו ע״מ לחקות אחרי התפלגות שאנו מחפשים, כפי שהערכת MLE מאפשרת כפי שראינו בהרצאה.

6. ערכים שיוחזרו הינם $f_1 = -0.050$ ו $f_2 = 3.970$. נשים לב כי אלו ערכים שמתקרבים לערכי התוחלת המקוריים, 0 ו-4 בהתאמה עבור קורדינאטות אלו, כפי שנצפה.