מבוא למערכות לומדות - תרגיל 1

מגיש: יואב לוי 314963257

שאלה 1

$$v = (1,2,3,4)\,,\ \ w = (0,-1,1,2)$$

$$\text{The projection of } v \text{ on } w \text{ will be: } p = \frac{\langle v,w\rangle}{||w||^2}\cdot w = \frac{3}{2}\begin{bmatrix}0\\-1\\1\\2\end{bmatrix}$$

שאלה 2

$$v = (1,2,3,4)\,,\ \ w = (1,0,1,-1)$$

$$\text{The projection of } v \text{ on } w \text{ will be: } p = \frac{\langle (1,2,3,4)\,,(1,0,1,-1)\rangle}{||(1,0,1,-1)||^2}\cdot w = \vec{0}$$

שאלה 3

<u>הוכחה</u>: נתונים שני וקטורים $\vec{0} \neq v,w \in \mathbb{R}^m$.

נסמן ב־θ את הזווית בין שני הווקטורים שלעיל.

$(\Leftarrow)$ נניח כי $\langle v,w\rangle = 0$ ולפי הגדרה גם מתקיים כי $0 = \langle v,w\rangle = ||v||\cdot||w||\cdot\cos(\theta)$

מהנתון כי שני הווקטורים אינם וקטורי האפס אזי בהכרח $\theta = \pm\pi = \pm 90^\circ \iff \cos(\theta) = 0$, כנדרש.

$(\Rightarrow)$ נניח כי $\theta = \pm 90^\circ$, כעת $\langle v,w\rangle = ||v||\cdot||w||\cdot\cos(90) = 0$ וגם

$\langle v,w\rangle = ||v||\cdot||w||\cdot\cos(-90) = 0$, כנדרש.

■

שאלה 4

$$\|Ax\|_2^2 = \langle Ax|Ax\rangle \overset{\text{dot product definition}}{=} (Ax)^T Ax = \left(x^T A^T\right) Ax \overset{A^t=A^{-1}\ \text{Orthgonal matrix}}{=} x^T I x = \langle x|x\rangle = \|x\|_2^2$$

ולכן,

$$\|Ax\|_2 = \|x\|_2$$

$\blacksquare$

שאלה 5

$$A = UDV^T \Longrightarrow A^{-1} = VD^{-1}U^{-1}$$

בעזרת ה־SVD ניתן להכפיל את A או את A^{-1} בחזקה גדולה במהירות, מכיוון שלהעלות בחזקה מטריצה אלכסונית זה קל (האלכסון עולה בחזקה).

שאלה 6

נשים לב כי $C^TC = VD^2V^T$ מטריצה סימטרית,

$$M \equiv C^TC = \begin{pmatrix} 5 & 5 \\ -1 & 7 \end{pmatrix} \cdot \begin{pmatrix} 5 & 5 \\ -1 & 7 \end{pmatrix} = \begin{pmatrix} 26 & 18 \\ 18 & 74 \end{pmatrix}$$

ממשפט הפירוק הספקטרלי המטריצה הזו ניתנת ללכסון אורתוגונלי, לכן נלכסן אותה אורתוגונלית וכך נמצא את V, D^2.

- משווים את הפולינום האופייני לאפס ומוצאים את הפתרונות של λ.

$$\det(\lambda I - M) = \begin{pmatrix} \lambda - 26 & -18 \\ -18 & \lambda - 74 \end{pmatrix} = (\lambda - 26)\cdot(\lambda - 74) - 324 = 0 \iff \lambda^2 - 100\lambda + 1600 = 0$$

לכן ה־Eigenvalues (=ערכים עצמיים) הם: $\lambda_1 = 20,\ \lambda_2 = 80$. (python » w, v = numpy.linalg.eig([[26,18],[18,74]]))

- נמצא את הווקטורים העצמיים.

$$V_{20}:\ (20I - M)\,v = 0 \iff \begin{pmatrix} 20-26 & -18 \\ -18 & 20-74 \end{pmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \iff v_1 = \begin{bmatrix} -3 \\ 1 \end{bmatrix}$$

$$V_{80}:\ (80I - M)\,v = 0 \iff \begin{pmatrix} 80-26 & -18 \\ -18 & 80-74 \end{pmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \iff v_2 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

- נפעיל את אלגוריתם GS=Gram-schmidt למציאת בסיס אורתונורמלי ובכך נמצא את V (המטריצה המלכסנת)

ננרמל את הווקטור v_1

$$u_1 = \frac{v_1}{\|v_1\|_2} = \frac{\sqrt{10}}{10}\begin{bmatrix} -3 \\ 1 \end{bmatrix}$$

כעת,

$$b_2 = v_2 - \langle v_2, u_1 \rangle u_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \Longrightarrow u_2 = \frac{b_2}{\|b_2\|_2} = \frac{\sqrt{10}}{10} \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

לכן,

$$V = \begin{pmatrix} -\frac{3\sqrt{10}}{10} & \frac{\sqrt{10}}{10} \\ \frac{\sqrt{10}}{10} & \frac{3\sqrt{10}}{10} \end{pmatrix} \in M_{3,3}(\mathbb{R})$$

- כמובן שגם המטריצה האלכסונית D^2 הינה,

$$D^2 = \begin{pmatrix} 20 & 0 \\ 0 & 80 \end{pmatrix} \in M_{3,3}(\mathbb{R})$$

בלי הגבלת הכלליות (מכיוון שהפירוק SVD אינו יחיד) נקבל

$$D = \begin{pmatrix} 2\sqrt{5} & 0 \\ 0 & 4\sqrt{5} \end{pmatrix}$$

עכשיו מכיוון ש-$CV = UDV^{-1}V = UD \iff C = UDV^T$ אזי,

$$CV = \begin{pmatrix} 5 & 5 \\ -1 & 7 \end{pmatrix} \cdot \begin{pmatrix} -\frac{3\sqrt{10}}{10} & \frac{\sqrt{10}}{10} \\ \frac{\sqrt{10}}{10} & \frac{3\sqrt{10}}{10} \end{pmatrix} = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} \cdot \begin{pmatrix} 2\sqrt{5} & 0 \\ 0 & 4\sqrt{5} \end{pmatrix} = UD$$

$$\Updownarrow$$

$$\begin{pmatrix} -\sqrt{10} & 2\sqrt{10} \\ \sqrt{10} & 2\sqrt{10} \end{pmatrix} = \begin{pmatrix} 2\sqrt{5}a_{1,1} & 4\sqrt{5}a_{1,2} \\ 2\sqrt{5}a_{2,1} & 4\sqrt{5}a_{2,2} \end{pmatrix}$$

כלומר,

$$U = \begin{pmatrix} -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}$$

הערה: נשים לב כי המטריצה סימטרית (וההופכית שלה היא המטריצה המשוחלפת ולכן היא אוניטרית\אורתוגונלית)

ולבסוף הפירוק SVD יהיה:

$$C = \begin{pmatrix} -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \cdot \begin{pmatrix} 2\sqrt{5} & 0 \\ 0 & 4\sqrt{5} \end{pmatrix} \cdot \begin{pmatrix} -\frac{3\sqrt{10}}{10} & \frac{\sqrt{10}}{10} \\ \frac{\sqrt{10}}{10} & \frac{3\sqrt{10}}{10} \end{pmatrix} = \begin{pmatrix} 5 & 5 \\ -1 & 7 \end{pmatrix}$$

שאלה 7

יהיו $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{R}$ הערכים העצמיים של המטריצה $C_0 := A^T A$ ו־$v_1, v_2, \ldots, v_n$ הווקטורים העצמיים המתאמים להם.
נניח בלי הגבלת הכלליות כי $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$.
בהנחה ש־$\lambda_1 > \lambda_2$, אז נשים לב שמתקיים:

$$b_{k+1} = \frac{C_0 b_k}{\|C_0 b_k\|} = \frac{C_0^{k+1} b_0}{\left\|C_0^{k+1} b_0\right\|} \overset{b_k := \sum_{i=1}^n a_i v_i}{=} \frac{C_0^{k+1} \sum_{i=1}^n a_i v_i}{\left\|C_0^{k+1} \sum_{i=1}^n a_i v_i\right\|} \overset{v_1,\ldots,v_n \text{ corresponding eigenvectors}}{=} \frac{\lambda_1^{k+1} \sum_{i=1}^n \left(\frac{\lambda_i}{\lambda_1}\right)^{k+1} a_i v_i}{\left\|\lambda_1^{k+1} \sum_{i=1}^n \left(\frac{\lambda_i}{\lambda_1}\right)^{k+1} a_i v_i\right\|} =$$

ומכאן, ומהנתון כי λ_1 הוא הערך העצמי הדומיננטי (גדול) ביותר, נקבל כי:

$$\lim_{k \to \infty} b_k = \frac{\lambda_1^{k+1}}{\left\|\lambda_1^{k+1} v_1\right\|} v_1 = \frac{v_1}{\|v_1\|}$$

■

שאלה 8

$$f(\sigma) = \sum_{i=1}^n \sigma_i u_i u_i^T x \qquad \text{where } u_i \text{ is the i'th column of matrix U}$$

כעת, איבר בשורה ה־k ובעמודה ה־i במטריצת היעקוביאן תראה כך:

$$\frac{\partial f_k\left(\sigma_{\text{const}}\right)}{\partial \sigma_i} = \left[u_i u_i^T x\right]_k$$

שאלה 9

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

אגדיר את הפונקציות הבאות:

$$g(x) = \|x\|^2$$

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

נשים לב כי מתקיים $h(\sigma) = \frac{1}{2} g \bigcirc (w(\sigma))$.
כעת, לפי כלל השרשרת שראינו בתרגול על יעקוביאן (=לגרדיאנט בפונקציות מווקטור לסקלר)

$$J_\sigma\left(\frac{1}{2}g\bigcirc\left(w\left(\sigma\right)\right)\right)=\frac{1}{2}J_\sigma\left(g\bigcirc\left(w\left(\sigma\right)\right)\right)\overset{\text{chain rule}}{=}\frac{1}{2}\overbrace{J_{w(\sigma)}\left(g\right)}^{\blacktriangledown}\cdot\overbrace{J_\sigma\left(w\right)}^{\blacktriangle}=\frac{1}{2}\cdot\overbrace{2\left(f\left(\sigma\right)-y\right)^T}^{\blacktriangledown}\cdot\overbrace{\nabla f\left(\sigma\right)}^{\blacktriangle}$$

כלומר,

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

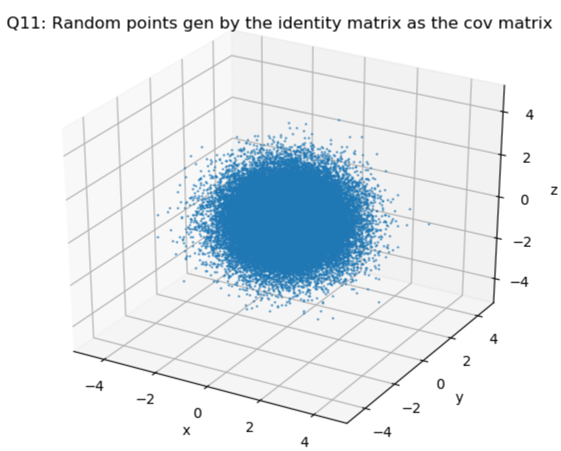
שאלה 10

חישוב היעקוביאן של פונקציית ה־softmax (נשים לב שהמשכתי את השלבים הראשונים שהופיעו בתרגול ולכן רוב המלל הוסר)

$$\frac{\partial g_i\left(z\right)}{\partial z_j}=g_i(z)\left(\delta_{ij}-g_j\left(z\right)\right)$$

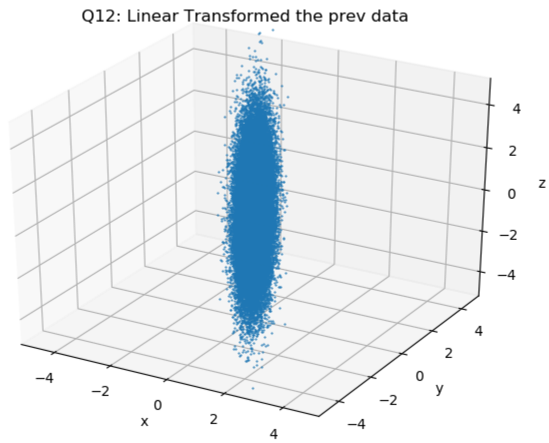
שאלה 11

<u>גרף הנתונים:</u>



שאלה 12

<u>גרף הנתונים לאחר הפעלת טרנספורמציה לינארית:</u>



<u>בנוסף מטריצת השונות המשותפת ("Covariance matrix") תהיה:</u>

צורה נומרית־

$$C=\begin{pmatrix}0.00995852 & 0.00030081 & -0.00073439\\ 0.00030081 & 0.25066825 & -0.00314623\\ -0.00073439 & -0.00314623 & 4.00677747\end{pmatrix}$$

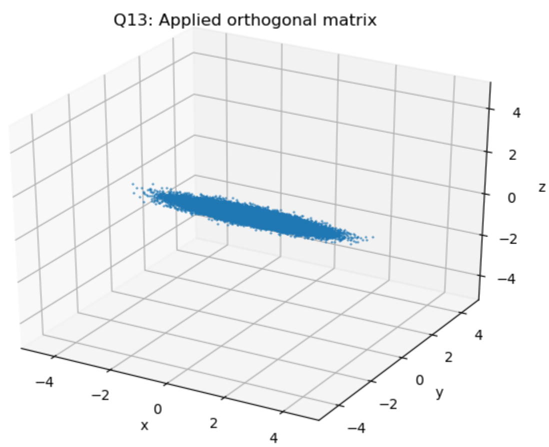
צורה אנליטית־

(כאשר דגמנו $50,000$ דגימות, והן מיוצגות במטריצה X ומטריצת השונות המשותפת המקורית היא $\hat{C}$)

$$C:=\frac{SX\left(SX\right)^{T}}{50000-1}=S\frac{XX^{T}}{49999}S^{T}=S\hat{C}S^{T}$$

שאלה 13

<u>גרף הנתונים לאחר הפעלת טרנספורמציה לינארית ואז הפעלת מטריצה אורתוגונלית:</u>



<u>מטריצת השונות המשותפת תהיה:</u>

צורה נומרית־

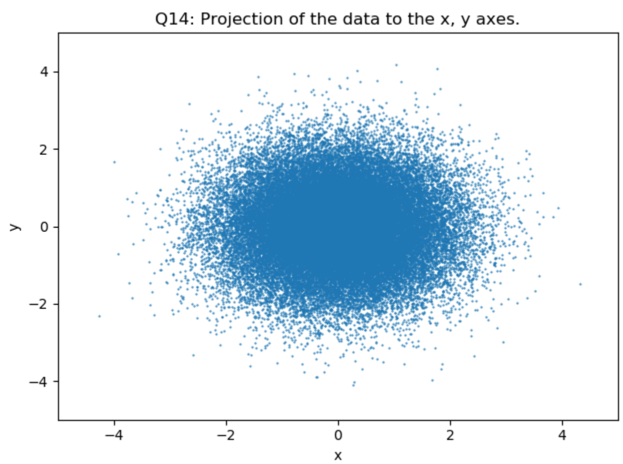
$$C = \begin{pmatrix} 0.24479533 & -0.02450719 & 0.05942625 \\ -0.02450719 & 2.64547029 & -1.84138481 \\ 0.05942625 & -1.84138481 & 1.30421308 \end{pmatrix}$$

צורה אנליטית־

$$RSCS^TR^T$$

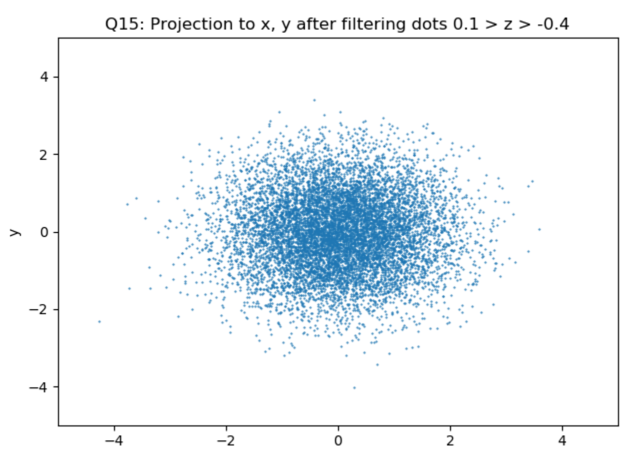
שאלה 14

<u>הגרף שמייצג את ההטלה של ה־Data על המישור x, y הוא:</u>



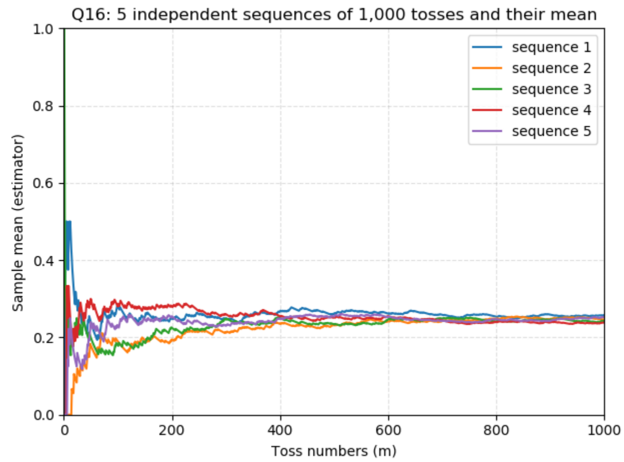
שאלה 15

<u>הגרף שמייצג את ההטלה של ה־Data שמקיים $0.1 > z > -0.4$ על המישור x, y :</u>



שאלה 16

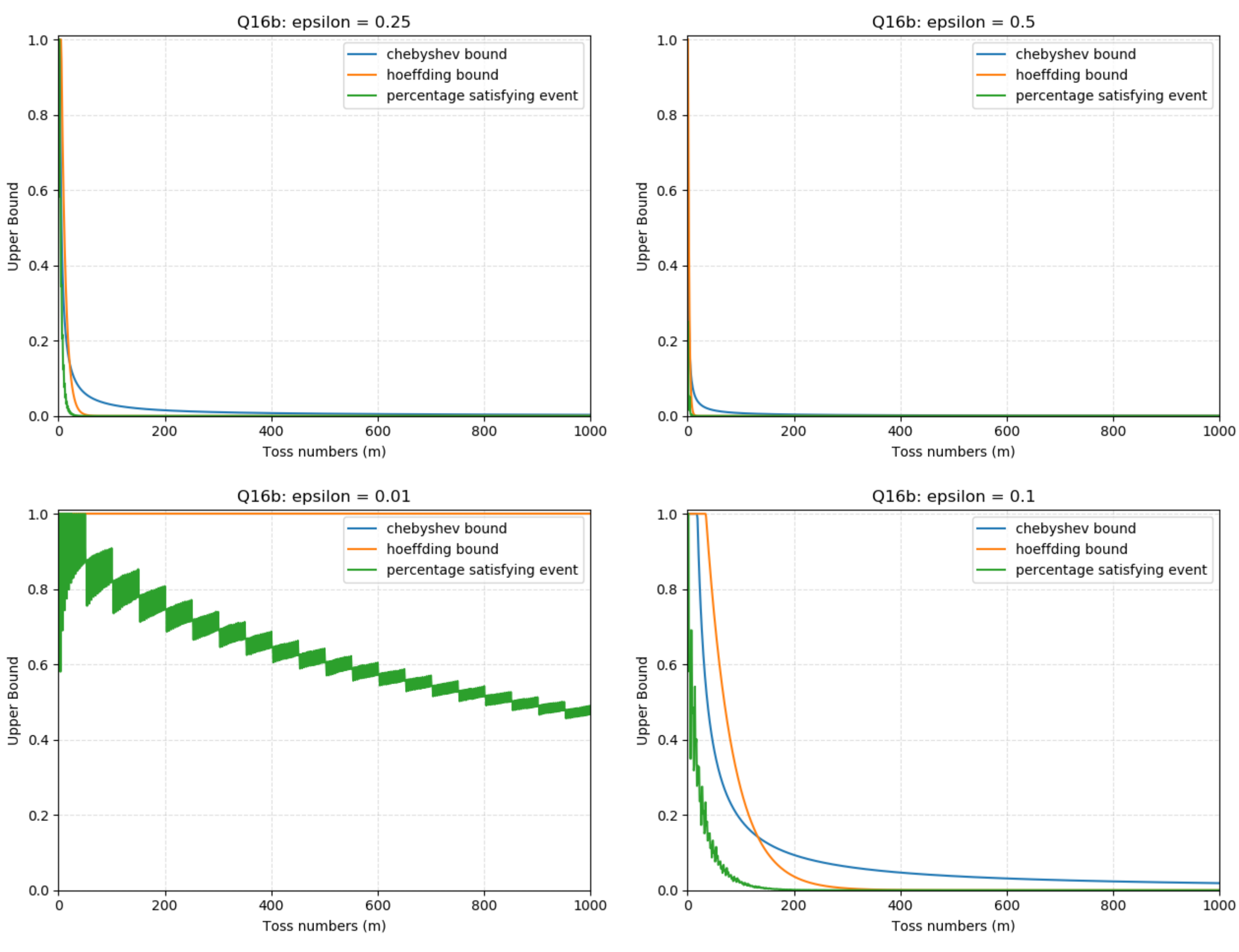
(A) אני מצפה לראות התכנסות לממוצא ככול שמספר ההטלות הולך וגדל (לפי החוק החלש של המספרים הגדולים), ואכן בגרף ניתן לראות זאת בבירור.

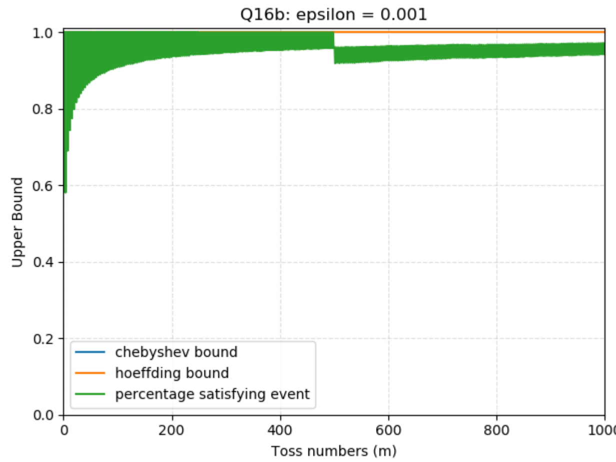


(B)+(C)

הסבר עבור סעיף c: אני מצפה לראות שהממוצע רחוק מהתוחלת בניסויים שלנו תרד ככול שמספר הזריקות עולה, כלומר ככל שאנחנו זורקים יותר כך המרחק שנתון בסעיף c יהיה קטן יותר וכך אחוז הניסויים שהיו רחוקים מאפסילון תקטן.

כמו כן כאשר אפסילון גדול יחסית, כך נגיע מהר יותר לאחוז 0, וההפך.




Q16b: epsilon = 0.001
1.0
0.8
0.6
0.4
0.2
0.0
Upper Bound
0
200
400
600
800
1000
Toss numbers (m)
chebyshev bound
hoeffding bound
percentage satisfying event