יובל דלוס 305211880

מבוא למערכות לומדות – תרגיל 1

Warm-up - Algebra Recap

1. Calculate the projection of $v = (1,2,3,4)$ on the vector $w = (0,-1,1,2)$.

1. נרצה להשתמש בנוסחה לחישוב הטלת וקטורים, כאשר v מוטל על w נקבל כי הווקטור המוטל יהיה:

$$\frac{<v,w>}{\| w \|^2}\cdot w = \frac{<(1,2,3,4),(0,-1,1,2)>}{\left(\sqrt{<(0,-1,1,2),(0,-1,1,2)>}\right)^2}\cdot(0,-1,1,2) = \frac{1\cdot 0+2\cdot(-1)+3\cdot 1+4\cdot 2}{\left(\sqrt{1+1+4}\right)^2}\cdot(0,-1,1,2) =$$

$$= \frac{9}{6}\cdot(0,-1,1,2) = 1\frac{1}{2}\cdot(0,-1,1,2) = \left(0,-\frac{3}{2},\frac{3}{2},3\right)$$

2. Calculate the projection of $v = (1,2,3,4)$ on the vector $w = (1,0,1,-1)$.

2. פעם נוספת נשתמש בנוסחה:

$$\frac{<v,w>}{\| w \|^2}\cdot w = \frac{<(1,2,3,4),(1,0,1,-1)>}{\|(1,0,1,-1)\|^2}\cdot(1,0,1,-1) = \frac{1+3-4}{3}\cdot(1,0,1,-1) = 0\cdot(1,0,1,-1) = (0,0,0,0)$$

3. Prove the angle between two non-zero vectors $v, w \in \mathbb{R}^m$ is ± 90 iff $\langle v, w\rangle = 0$.

3. נשתמש בזהות של $cos\theta$ ובפיתוח המלא של חישוב ווקטור הטלה:

$$\| v \|\, cos\theta\cdot\frac{u}{\| u \|} = \| v \|\cdot\frac{<v,u>}{\| v \|\cdot\| u \|}\cdot\frac{u}{\| u \|} = \frac{<v,u>}{\| u \|^2}\cdot u$$

$$cos\theta = \frac{<v,u>}{\| v \|\cdot\| u \|}$$

נתון לנו כי הזווית בין הווקטורים היא $90°$, לכן ניעזר בזהות של קוסינוס ונקבל:

$$\cos(\pm 90) = 0 = \frac{<v,u>}{\| v \|\cdot\| u \|} \Longrightarrow <v,u> = 0$$

מצד שני, נתון לנו כי $<v,u> = 0$, ושני הווקטורים שונים מ-0, לכן נקבל שווקטור ההטלה יהיה

$$\frac{<v,u>}{\| u \|^2}\cdot u = 0\cdot u = 0 \in \mathbb{R}^n$$

וזה קורה רק כשהווקטורים אנכים זה לזה.

יובל דלוס 305211880

4. Prove that Orthonormal matrices are isometric transformations. That is let $T : V \mapsto W$ be some linear transformation and A the corresponding matrix. Then if A is orthonormal then $\forall x \in V \quad ||Ax||_2 = ||x||_2$.

.4

$$\|x\|_2 \overset{\text{def}}{=} \left(\sum_{i=1}^{m} |x_i|^2\right)^{\frac{1}{2}} \overset{\text{def}}{=} \left(\sum_{i=1}^{m} < x_i|x_i >\right)^{\frac{1}{2}} = \left(\sum_{i=1}^{m} < x_i \mid I_M x_i >\right)^{\frac{1}{2}} \overset{\text{def}}{=} \left(\sum_{i=1}^{m} < x_i|A^T A x_i >\right)^{\frac{1}{2}} \overset{\text{Orthonormal}}{=}$$

$$= \left(\sum_{i=1}^{m} < Ax_i|Ax_i >\right)^{\frac{1}{2}} = \left(\sum_{i=1}^{m} |Ax_i|\right)^{\frac{1}{2}} = \|Ax\|_2$$

SVD

5. Assume A is invertible. Write a formula for the inverse of A using only the matrices U, D, V where UDV^T is an SVD decomposition of A. Many learning algorithm implementations require calculating the inverse of a matrix. Explain why knowing the SVD decomposition of matrix is usefull in this context.

5. נבחן ראשית מה עושה פירוק ה-SVD. כאשר אנו מפרקים מטריצה לשלוש מטריצות SVD אנו בעצם מפרקים אותה ל-3 פעולות שונות, סיבוב, מתיחה, וסיבוב חזרה לכיוון המקורי. לשם כך, בכדי לחשב את ההופכי של מטריצת A נרצה לבצע את הפעולות בסדר הפוך, אזי לסובב, לכווץ ולסובב חזרה, לשם כך נגדיר

$$A^{-1} = VD^{-1}U^T$$

כאשר

$$D^{-1} = \begin{bmatrix} \frac{1}{d_{11}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{d_{nn}} \end{bmatrix}$$

נוכיח את הטענה הזו :

$$AA^{-1} = (UDV^T)(VD^{-1}U^T) = UD(V^TV)D^{-1}U^T = U(DD^{-1})U^T = UU^T = I_n$$

כנדרש, כאשר נעזרנו בעובדה ש U ו- V הן מטריצות אורתוגונליות ולכן מקיימות $U^TU = I_n$.

לדעת את ה-SVD יכול מאוד להאיץ חישובים של מטריצות, בעוד חישוב של מטריצה הופכית כולל פעולות רבות ויכול לעלות $O(n^3)$, חישוב של כל אחת ממטריצות הטרנספוס וההופכית-מספרית של מטריצות ה-SVD יעלו $O(n)$ בלבד, כאשר n הוא אורך הווקטור הארוך ביותר במטריצה.

יובל דלוס 305211880

6. Find an SVD of

$$C = UDV^T = \begin{pmatrix} 5 & 5 \\ -1 & 7 \end{pmatrix}$$

I.e., find matrices U, D, V^T where U, V are orthogonal matrices and D is diagonal.

Do the following steps:

- Calculate $C^T C$.
- Deduce V and D (hint: use the *Eigenvalues Decomposition*; You can either use the function *numpy.linalg.eig* in python or refresh your memory and do it manually).
- Find U using the equality $CV = UD$.

.6

- $C^T = \begin{pmatrix} 5 & -1 \\ 5 & 7 \end{pmatrix}$

$$C^T C = \begin{pmatrix} 5 & -1 \\ 5 & 7 \end{pmatrix}\begin{pmatrix} 5 & 5 \\ -1 & 7 \end{pmatrix} = \begin{pmatrix} 26 & 18 \\ 18 & 74 \end{pmatrix}$$

אבל בנוסף נקבל כי

$$C^T C = VD^T U^T UDV^T = VD^T DV^T$$

כאשר המעבר השני התאפשר מכיוון ש-U מטריצה אורתוגונלית.

כעת, יש לנו מטריצה ריבועית בין מטריצה אורתוגונלית והמשוחלפת שלה, נמצא ו״ע וע״ע.

$$\det(C^T C - \lambda I_n) = det\begin{pmatrix} 26-\lambda & 18 \\ 18 & 74-\lambda \end{pmatrix} = \lambda^2 - 100\lambda + 1{,}600 = (\lambda - 20)(\lambda - 80)$$

$$V_{\lambda=20} = \begin{pmatrix} 6 & 18 \\ 18 & 54 \end{pmatrix} = \begin{pmatrix} 6 & 18 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ 0 & 0 \end{pmatrix} \Rightarrow v_1 = \frac{1}{\sqrt{10}} \cdot \begin{pmatrix} 1 \\ 3 \end{pmatrix}$$

$$V_{\lambda=80} = \begin{pmatrix} -54 & 18 \\ 18 & -6 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 18 & -6 \end{pmatrix} = \begin{pmatrix} 3 & -1 \\ 0 & 0 \end{pmatrix} \Rightarrow v_2 = \frac{1}{\sqrt{10}} \cdot \begin{pmatrix} 3 \\ -1 \end{pmatrix}$$

כאשר המקדם הוא לצרכי נרמול

לכן:

$$V^T = \frac{1}{\sqrt{10}}\begin{pmatrix} 1 & 3 \\ 3 & -1 \end{pmatrix}$$

$$D = \begin{pmatrix} \sqrt{80} & 0 \\ 0 & \sqrt{20} \end{pmatrix}$$

כעת בכדי למצוא את U ניעזר בעובדה כי $CV = UD$:

$$CV = \frac{1}{\sqrt{10}}\begin{pmatrix} 5 & 5 \\ -1 & 7 \end{pmatrix}\begin{pmatrix} 1 & 3 \\ 3 & -1 \end{pmatrix} = \sqrt{10}\begin{pmatrix} 2 & 1 \\ 2 & -1 \end{pmatrix} = UD = U \cdot \sqrt{10}\begin{pmatrix} \sqrt{8} & 0 \\ 0 & \sqrt{2} \end{pmatrix} \Rightarrow$$

$$\Rightarrow U = \begin{pmatrix} \sqrt{1/2} & \sqrt{1/2} \\ \sqrt{1/2} & -\sqrt{1/2} \end{pmatrix}$$

יובל דלוס 305211880

7. (Power Iteration) In this section we will implement an algorithm for SVD decomposition, we will use the relation between SVD of A to EVD of A^TA that we saw in recitation. For some $A \in M_{m\times n}(\mathbb{R})$, define $C_0 = A^TA$.

Let $\lambda_1, \lambda_2, ...\lambda_n$ be the eigenvalues of C_0, with the corresponding eigenvectors $v_1, v_2, ...v_n$, ordered such that $\lambda_1 \geqslant \lambda_2 \geqslant ... \geqslant \lambda_n$.

Assume $\lambda_1 > \lambda_2$, where λ_1 is the largest eigenvalue and λ_2 is the second-largest one.

Define $b_{k+1} = \frac{C_0 b_k}{\|C_0 b_k\|}$, and initialize b_0 randomly.

Show that: $\lim_{k\to\infty} b_k = v_1$

Hint: use EVD decomposition of C_0 and represent b_0 accordingly. You can assume that $b_0 = \sum_{i=0}^n a_i v_i$, where $a_1 \neq 0$. As b_0 is initialized randomly, the probability of $a_1 = 0$ is zero.

7. ראשית נגדיר את המטריצה שלנו

$$A = UDV^T$$
$$C_0 = A^TA = VD^TDV^T$$

לכן נוכל להתייחס לו״ע של D^TD.

נביט על הווקטור ההתחלתי שלנו, נוכל לרשום אותו כצירוף לינארי של כל השאר:

$$b_0 = a_1v_1 + a_2v_2 + \cdots + a_nv_n$$

ונסתכל על התהליך הנדרש מאיתנו:

$$b_1 = \frac{C_0b_0}{\|c_0b_0\|} = \frac{1}{\|c_0b_0\|}\cdot C_0b_0 = \frac{1}{\|c_0b_0\|}\cdot C_0(a_1v_1 + a_2v_2 + \cdots + a_nv_n) \overset{(*)}{=} \frac{1}{\|c_0b_0\|}\cdot (a_1\lambda_1v_1 + a_2\lambda_2v_2 + \cdots + a_n\lambda_nv_n)$$

כאשר (*) נובע מכך שכל v_i הוא ו״ע, ולכן הוא אינו משתנה בפרט בהכפלה בע״ע המתאים לו.

באותו אופן:

$$b_2 = \frac{C_0b_1}{\|c_0b_1\|} = \frac{1}{\|c_0b_1\|}\cdot C_0b_1 = \frac{1}{\|c_0b_0\|}\cdot\frac{1}{\|c_0b_1\|}\cdot C_0(a_1\lambda_1v_1 + a_2\lambda_2v_2 + \cdots + a_n\lambda_nv_n) =$$
$$= \frac{1}{\|c_0b_0\|\cdot\|c_0b_1\|}\cdot(a_1\lambda_1^2v_1 + a_2\lambda_2^2v_2 + \cdots + a_n\lambda_n^2v_n)$$

ובאופן כללי:

$$b_k = \frac{1}{\prod_{i=0}^{k-1}\|c_0b_i\|}\cdot\left(a_1\lambda_1^kv_1 + a_2\lambda_2^kv_2 + \cdots + a_n\lambda_n^kv_n\right) =$$
$$= \frac{1}{\prod_{i=0}^{k-1}\|c_0b_i\|}\cdot\lambda_1^k\cdot\left(a_1v_1 + a_2\left(\frac{\lambda_2}{\lambda_1}\right)^k v_2 + \cdots + a_n\left(\frac{\lambda_n}{\lambda_1}\right)^k v_n\right)$$

ומכיוון שהנחנו כי λ_1 הוא הגדול ביותר, נותר לנו להסיק כי $\frac{\lambda_i}{\lambda_1} < 0$ לכל i, ולכן

$$li\,m_{k\to\infty}\left(\frac{\lambda_i}{\lambda_1}\right)^k = 0$$

לכן

$$\lim_{k\to\infty} b_k = lim_{k\to\infty} \frac{1}{\prod_{i=0}^{k-1}\|c_0 b_i\|} \cdot \lambda_1^k \cdot \left(a_1 v_1 + a_2 \left(\frac{\lambda_2}{\lambda_1}\right)^k v_2 + \cdots + a_n \left(\frac{\lambda_n}{\lambda_1}\right)^k v_n\right) = \frac{1}{\prod_{i=0}^{k-1}\|c_0 b_i\|} \cdot \lambda_1^k \cdot a_1 \cdot v_1$$

וקיבלנו את הווקטור המבוקש עד כדי ניפוח בסקלר.