מבוא למתמטיקה שימושית

רועי נחבי - 208644854

<u>אלגברה לינארית</u>:

1) תהי A מט' אורתוגונלית.

אזי נוכיח כי לכל $x\in V$ $\quad \|Ax\|=\|x\|$

<u>הוכחה</u>: נזכר בהגדרות ונקבל:

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

אזי ראינו כי מט' המעבר אורתוגונלית נוכל לסמן

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

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

$$\|Ax\|^2=\langle Ax,Ax\rangle = (Ax)^TAx = x^TA^TAx =$$

מתכונות מכפלה סקלרית.

בקורס לינ' אחת 2 ראינו כי המכפלה Q מט' אורתו' היא המטריצה הזהה.

נוכל להוכיח זאת גם באופן ישיר, מכיוון שעמודות A הינן בסיס אורתו' אזי

$$(A^TA)_{ij}=v_i^Tv_j=\langle v_i|v_j\rangle$$

$\Leftarrow$ מכיוון ש $v_1,\dots,v_n$ בסיס אורתונורמלי מתקיים כי

$$\langle v_i|v_j\rangle=\delta_{ij}$$ הוא 1 כאשר $i=j$ ו-0 אחרת כנדרש.

$$A^TA=I_n \Leftarrow$$

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

$= \|x\|^2$

$\Leftarrow$ נורמות הן אי-שליליות, ולכן $\sqrt{\ }$ / $\|Ax\|^2 = \|x\|^2$

$\Downarrow$

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

כנדרש!

ב) נתון $A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 2 \end{pmatrix}$ ונחשב את ה-SVD

של A את המטריצות U, Σ, V^T כפי שראינו בהרצאה.

תחילה נחשב ע"ע של $A^T A$

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

נחשב ע"ע:

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

$$= (2-\lambda)\big((2-\lambda)(4-\lambda) - 4\big) + 2\big(2(\lambda-2)\big) =$$

$$= (2-\lambda)\big((\lambda-2)(\lambda-4) - 4\big) + 4(\lambda-2) =$$

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

$\lambda = 6, 2, 0 \;//$ ⇐

⇐ נקבל כי

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

<u>נחשב</u> ע״ע:

$$A^TA \cdot v = 0 \Rightarrow \left\{ \begin{matrix} v_1 + v_3 \\ v_2 - v_3 \\ v_1 - v_2 + 2v_3 \end{matrix} \right. = 0$$

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

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

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

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

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

⇐ <u>ננרמל</u>!

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

⇐ לבצע <u>נרמול</u>!

כעת נמצא את U לפי הנוסחה הבאה של וקטור
היות

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

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

זוהי
כמובן
בערך
מדויק.

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

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

וסיימנו!!

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

3) תהי A מט' כלשהי ו- $C_0 = A^TA$, יהיו $\lambda_1 \geq \dots \geq \lambda_n$
הע"ע של C_0 עם הו"ע $v_1, \dots, v_n$ בהתאמה. (v_i מנורמלים)
יהי $b_k \in \mathbb{R}^n$ מוגדר באופן הבא:

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

וקטור מנורמל.

צ"ל: עבור $a_1 \neq 0$ $\lim_{k \to \infty} b_k = \pm v_1$

הוכחה: נפרק את הנוסחה:

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

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

כי נורמה של מספר היא עצמו.

ניזכר באופן רקורסיבי

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

יהי $b_0 = \sum_{i=1}^{n} a_i v_i$ צ"ל כי עבור בסיס אורתונורמלי.

$$C_0^k b_0 = C_0^k \sum_{i=1}^{n} a_i v_i \underset{\text{הגדרת ו"ע}}{=} \sum_{i=1}^{n} a_i \cdot \lambda_i^k v_i =$$

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

⇐ כעת נשים לב כי $\left(\frac{\lambda_i}{\lambda_1}\right)^k \to 0$ לכל $2 \le i \le n$ מכיוון שידוע כי הע"ע מסודרים בסדר יורד.

⇐ כאשר נשאיף את $k \to \infty$ נקבל כי הנ"ל הופך ל-0.

⇐

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

ונציב אותו בגבול תוך פעולת הנרמול ונקבל

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

$$= \lim_{k\to\infty} \pm\frac{v_1}{\|\cancel{v_1}\|} = \pm v_1$$

$\|v_1\| = 1$ כי ו"ע מנורמלים

כנדרש!

∎

multivariate calculus:

4) יהי $x\in\mathbb{R}^d$ וקטור קבוע ו- $U\in\mathbb{R}^{d\times d}$ מט' אורתוגונלית.
נחשב את יעקוביאן של הפונק' $f:\mathbb{R}^d\to\mathbb{R}^d$

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

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

פתרון: ראשית לפי מימדי הוקטור נשים לב כי $J_x(f)\in\mathbb{R}^{d\times d}$

נרצה להיות מסוגלים לחשב את תוצאת הכפל ע"י
שנוכל כל פעם לגזור לפי משתנה אחר, הרי ראינו כי
כל שורה ב-J היא הגרדיאנט של הפונק' f_i

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

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

$$U_{i,j} = a_{i,j} = a_{j,i}$$

$$u = U^T\begin{pmatrix}x_1\\ \vdots\\ x_d\end{pmatrix} = \begin{pmatrix} \vdots \\ \sum_{j=1}^{d} a_{j,i}x_j \\ \vdots\end{pmatrix} =$$

סימון $\tilde{u}$

$$U\,\mathrm{diag}(\sigma) = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} =$$

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

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

מה קורה j.

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

נרצה כעת פורמלי:

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

קרי היעקוביאן יהיה גם אלכסונית (היות כי כל פונק׳ תלויה אחרת) ולכן

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

5) נעזר בכלל השרשרת על מנת לחשב את הגרדיאנט של

$$h(\sigma) = 1/2\,\|f(\sigma)-y\|^2$$

נגדיר:

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

$$h(\sigma) = 1/2\,\|g(\sigma)\|^2$$

נעזר בדוגמה שראינו בתרגול ונקבל

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

ראינו כי הגרדיאנט של הפונק׳ $1/2\,\|g(\sigma)\|^2$ הינו $(g(\sigma))^T$

ולכן:

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

6. נחשב את היעקוביאן של פונק' softmax

בתרגול ראינו כי במקרה ש $i=j$:

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

כעת נראה שבמקרה בו $i \neq j$ נקבל את היעקוביאן:

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

ונגזור מנה:

0 וזה קבוע.

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

ליניאריות הנגזרת

כל מה שלא x_j אפס

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

$$= -S_i \cdot S_j$$

$\Rightarrow$

$$J_{z(i,j)} f = S_i(\delta_{i,j} - S_j)$$

וכן לכל:

אילו $j=i$ אז מקבלים $S_i(1-S_i)$ כפי שרצינו

כנדרש, אחרת $S_i(0-S_j) = -S_iS_j$ כנדרש.

7) תהי $f(x,y)=x^3-5xy-y^5$ מצאו את ההסיאן של f.

במקרה זה נקבל שההסיאן יהיה מטריצה 2×2

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

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

$$\frac{\partial^2 f}{\partial x \partial y} = -5 = \frac{\partial^2}{\partial y \partial x}$$ נוע לינה תכונה זו מתבטא במקום 2.

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

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

Estimation Theory:

ש.) יהיו $X_1, X_2, \ldots \overset{iid}{\sim} P$ דגימות של התפלגות עם μ
תוחלת ושונות סופיות.
נראה כי האומד לתוחלת $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$ הוא
קונסיסטנטי.

פתרון: נזכיר בהגדרה של אומד קונסיסטנטי

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

כעת נזכיר במקרה שלנו ונוכיח ע״י א״ש צ׳בישב שקיים הנ״ל.

יהי $\varepsilon > 0$ ויהי $n \in \mathbb{N}$,

$$\theta_n = \hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} x_i \quad , \quad \hat{\mu}$$ הקבוע אליו אנו אומדים

נזכיר:

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

נרצה להשתמש בא״ש צ׳בישב ולכן נחשב תוחלת ושונות

$|X - E(X)|$

נשים לב כי כיוון $E(x_i) = \hat{\mu}$ עבור כל i, וכיוון ש- $x_i \overset{iid}{\sim}$ מתקיים:

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

$\Leftarrow$ נוכל להשתמש בא״ש צ׳בישב ולקבל

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

$$= \frac{var\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)}{\varepsilon^2} \underset{\text{סכום שונויות מ"מ ב"ת}}{=} \frac{\frac{1}{n^2}\cdot \cancel{n}\, var(x_i)}{\varepsilon^2} =$$

$$= \frac{1}{n}\cdot\frac{var(x_i)}{\varepsilon^2}$$

כעת נציב בהגדרת הגבול ונשתמש בנוסח המקורי ונקבל:

$$\lim_{n\to\infty} \mathbb{P}\left(|\hat{\mu}_n - \hat{\mu}| > \varepsilon\right) < \frac{1}{n}\underbrace{\frac{var(x_i)}{\varepsilon^2}}_{\text{קבוע}} \xrightarrow{n\to\infty} 0$$

$$\Leftarrow \quad \lim_{n\to\infty} \mathbb{P}\left(|\hat{\mu}_n - \hat{\mu}| > \varepsilon\right) = 0$$

קרי $\hat{\mu}$ קונסיסטנטי לפי הגדרה כנדרש!

מש״ל

4. יהיו $x_1,\dots,x_n \overset{iid}{\sim} N(\mu,\Sigma)$ עם n דגימות כאשר $\mu\in\mathbb{R}^d$ וקטור התוחלת ו-Σ מט' השונות.

נסיק מהי פונ' log-likelihood עבור $N(\mu,\Sigma)$.

ונזכר כי עבור מ"ט רב-ממדי נורמלי מתקיים:

$$\forall i\in[n] \quad f(x_i) = \frac{1}{\sqrt{(2\pi)^d|\Sigma|}}\cdot \exp\left(-\frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

כמו ראינו במקרה הפרטי נגדיר כי $\Theta=\{\mu,\Sigma\}$

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

נזכר:

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

<= כאן נפעיל $\log$ על הביטוי:

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

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

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

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

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

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