# שאלה 1 (מטריצות)

(a) (⇒) נניח ש-$A-C$ היא PSD. נשתמש בלמה 2 שראינו בתרגול: היות ו-PSD, אזי כל הערכים העצמיים שלה אי שליליים (כלומר, יש להם שורש ממשי). נסמן את $0 \le \lambda_1, \dots, \lambda_n$ הע"ע של $A$, אזי $0 \le \sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}$ מוגדרים היטב. נשים לב בנוסף לכך כי $A = QDQ^T$ כאשר $Q$ אורתוגונלית ו-$D$ אלכסונית עם הע"ע $\lambda_1, \dots, \lambda_n$ על האלכסון. נשים לב כי נוכל לרשום את המכפלה הנ"ל כך:

$$A = QDQ^T = Q \cdot \begin{pmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_n\end{pmatrix} \cdot Q^T = \underbrace{Q \cdot \begin{pmatrix}\sqrt{\lambda_1} & & 0\\ & \ddots & \\ 0 & & \sqrt{\lambda_n}\end{pmatrix}}_{\text{נסמן ב-}X\text{ את המכפלה הזו}} \cdot \begin{pmatrix}\sqrt{\lambda_1} & & 0\\ & \ddots & \\ 0 & & \sqrt{\lambda_n}\end{pmatrix} \cdot Q^T = X \cdot X^T$$

כי מתקיים:

$$X^T = \left(Q \cdot \begin{pmatrix}\sqrt{\lambda_1} & & 0\\ & \ddots & \\ 0 & & \sqrt{\lambda_n}\end{pmatrix}\right)^T = \begin{pmatrix}\sqrt{\lambda_1} & & 0\\ & \ddots & \\ 0 & & \sqrt{\lambda_n}\end{pmatrix}^T \cdot Q^T = \begin{pmatrix}\sqrt{\lambda_1} & & 0\\ & \ddots & \\ 0 & & \sqrt{\lambda_n}\end{pmatrix} \cdot Q^T$$

ולכן אכן $X$ שהגדרנו מקיים $A = XX^T$ כנדרש.

(⇐) נניח כי מתקיים $A = X \cdot X^T$ ונוכיח כי $A$ היא PSD. יהי $v$ וקטור, מתקיים:

$$v^TAv = v^TXX^Tv = (X^Tv)^T \cdot (X^Tv) \underset{\text{מלמה 2}}{=} \|X^Tv\|^2 \ge 0$$

ולכן מתקיים לכל וקטור $v$ כי $v^TAv \ge 0$, כלומר $A$ היא PSD.

□

(b) יהיו מטריצות $A,B\in R^{n\times n}$ PSD ויהיו $\alpha,\beta\ge0$ וגם וקטור $V\in R^n$. מתקיים:

$$V^T(\alpha A+\beta B)V=V^T\alpha AV+V^T\beta BV=\underbrace{\alpha}_{\ge0}\cdot\underbrace{V^TAV}_{\ge0}+\underbrace{\beta}_{\ge0}\cdot\underbrace{V^TBV}_{\ge0}\ge0$$

ולכן עבור כל הגדרה, קיבלנו כי $\alpha A+\beta B$ הינו PSD.

לעומת מטריצות לכל $A,B\in R^{n\times n}$ שהן PSD ולכל $\alpha,\beta\ge0$ לא דווקא, מתקיים אך לא שאמור

סימטריות שליליות, אם נכפול מטריצה C ב-A שהינה PSD. $\to$ $\gamma<0$

אז $\gamma\cdot C<0$ כלומר לא PSD, ואין סגירות לכפל בסקלר, ולכן לא מהווה

וקטורי מעל R.

□

שאלה 1 (הסתברות והתפלגות)

(a) יהיו $X_1,\dots,X_n$ משתנים i.i.d כך שלכל $1\le i\le n$ מתקיים $X_i\in U[0,1]$ ונגדיר $Y=\max(X_1,\dots,X_n)$

נרצה לחשב את פונקציית הצפיפות של $Y$, לכן נחשב תחילה את ההצטברות.

$$F_Y(x)=P(Y\le x)=P(\max(X_1,\dots,X_n)\le x)=P(X_1\le x,\dots,X_n\le x)=(P(X_1\le x))^n=(F_{X_1}(x))^n$$

$X_1,\dots,X_n$ הם i.i.d

כמו כן מתקיים להתפלגות, עבור $X_1\in U[0,1]$ מתקיים:

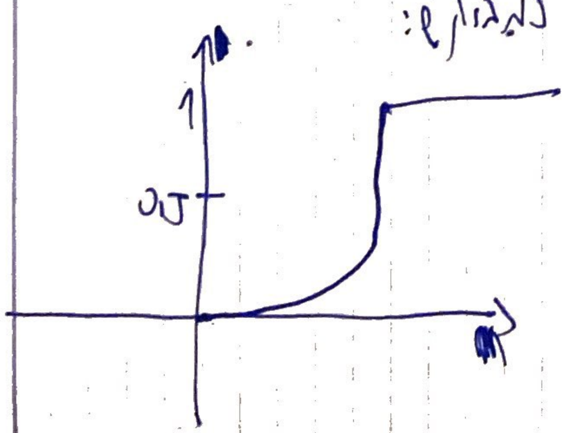
$$F_{X_1}(x)=\begin{cases}0 & x<0\\ x & x\in[0,1]\\ 1 & x>1\end{cases}\Rightarrow f_{X_1}(x)=\begin{cases}0 & x\notin[0,1]\\ 1 & x\in[0,1]\end{cases}$$

שאר ערכים נשמרים

כעת, נגזור את ההצטברות של $Y$:

$$f_Y(x)=(F_Y(x))'=\frac{d}{dx}(F_{X_1}(x))^n=n\cdot(F_{X_1}(x))^{n-1}\cdot f_{X_1}(x)=n x^{n-1}\cdot I_{[0,1]}$$

נצייר את ההצטברות של $Y$ כמבוקש:



כעת נחשב תוחלת ושונות עבור Y

$$E[Y] = \int_0^1 x\cdot f_Y(x)\,dx = \int_0^1 x n x^{n-1}\,dx = n\int_0^1 x^n\,dx = \frac{n}{n+1}$$

$y\neq 0$ ולתחום עבור

ולכן כאשר $n\to\infty$ נקבל: $\boxed{E[Y]\xrightarrow{n\to\infty}1}$ כי $\lim_{n\to\infty}\frac{n}{n+1}=1$

וכעת נחשב את השונות. נחשב את המומנט השני לפי הגדרה:

$$E[Y^2] = \int_0^1 x^2 f_Y(x)\,dx = n\int_0^1 x^{n+1}\,dx = \frac{n}{n+2}$$

$\boxed{\text{ולכן השונות היא}}$

ולכן:

$$Var(Y) = E[X^2]-(E[X])^2 = \frac{n}{n+2} - \frac{n^2}{(n+1)^2}$$

נחשב את הגבול כאשר $n\to\infty$:

$$\lim_{n\to\infty} Var(Y) = \frac{n}{n+2} - \frac{n^2}{n^2+2n+1} = 1-1=0$$

קיבלנו $\boxed{\text{השונות שואפת ל-0 כאשר } n\to\infty.}$

(b) לא פתרתי...

---

שאלה 2 (הבונוס)

$q=m$

~~נגדיר: נגדיר $h(q)=E[|X-q|]$ ונראה ש... $m$ ...~~

$$\cancel{h(q) = E[|X-q|] = \int_{-\infty}^{\infty}|x-q|f_X(x)\,dx = \int_{-\infty}^{q}(q-x)f_X(x)\,dx + \int_q^{\infty}(x-q)f_X(x)\,dx}$$

~~נגזור לפי $q$ ונשווה ל-0:~~

$$\cancel{\frac{d}{dq}(h(q)) = (q-x)\cdot\underbrace{\frac{d}{dq}\int_{-\infty}^{q}f_X(x)\,dx}_{=f_X(q)} + \int_{-\infty}^{q}f_X(x)\,dx + (x-q)\cdot\underbrace{\frac{d}{dq}\int_q^{\infty}f_X(x)}_{=f_X(q)} - \int_q^{\infty}f_X(x)\,dx =}$$

$$\cancel{= \underbrace{(q-x+x-q)}_{=0}\cdot f_X(q) + \int_{-\infty}^{q}f_X(x)\,dx - \int_q^{\infty}f_X(x)\,dx = 0 \Rightarrow \int_{-\infty}^{q}f_X(x)\,dx = \int_q^{\infty}f_X(x)\,dx}$$

~~ואז נקבל לפי הגדרת החציון: $P(X\le q)=P(X\ge q)$, כלומר $P(X\le q)=P(X\ge q)=\frac{1}{2}$~~

~~ולכן זה $m$ שמקיים את תנאי החציון!!~~

$\square$

## שאלה 2 (המשך)

כדי למצוא המינימום מספיק לגזור את h ונקבל:

$$h(y) = E[|X-y|] = \int_{-\infty}^{\infty} |x-y| f_X(x)\,dx = \int_{-\infty}^{y} (y-x) f_X(x)\,dx + \int_{y}^{\infty} (x-y) f_X(x)\,dx =$$

$$= y\int_{-\infty}^{y} f_X(x)\,dx - \int_{-\infty}^{y} x f_X(x)\,dx + \int_{y}^{\infty} x f_X(x)\,dx - y\int_{y}^{\infty} f_X(x) =$$

$$= yF_X(y) - y(1-F_X(y)) + \int_{y}^{\infty} x f_X(x)\,dx - \int_{-\infty}^{y} x f_X(x)\,dx =$$

$$= 2yF_X(y) - y + \int_{y}^{\infty} x f_X(x)\,dx - \int_{-\infty}^{y} x f_X(x)\,dx$$

נגזור לפי y:

$$h'(y) = (2yF_X(y))' - yf_X(y) - yf_X(y) - 1 = 2y f_X(y) - 2yf_X(y) - 1 + 2F_X(y) =$$

$$= 2F_X(y) - 1$$

ואם נשווה ל-0 נקבל $F_X(y)=\frac{1}{2}$, וזה בדיוק m לפי הגדרה

שזה גם מה שרצינו. ניתן לומר שזה מינימום כי אנו יודעים מהשאלה

נקבל שאנו יכולים (מניחים שאנו מוגדרים) לשאלות האלה.

$\square$

שאלה 1 [illegible]

(a) נקבע $X=x$ מסוים:

$$L(h)=E[loss(Y,h(X))]=\sum_{i=1}^{L} loss(i,h(x))\cdot P(X=x,Y=i)$$

ולכן לפי [illegible]:

$$P(X=x,Y=i)=P(X=x)\cdot P(Y=i|X=x)$$

ולכן L יהיה:

$$L(h)=P(X=x)\cdot\sum_{i=1}^{L} loss(i,h(x))\cdot P(Y=i|X=x)$$

והיות ש- $h(x)\in\{1,\dots,L\}$:

$$loss(i,h(x))=\begin{cases}0 & h(x)=i\\ 1 & else.\end{cases}$$

ולכן נקבל:

$$L(h)=\begin{cases}P(X=x)\cdot(1-P(Y=1|X=x)) & h(x)=1\\ \vdots & \\ P(X=x)\cdot(1-P(Y=L|X=x)) & h(x)=L\end{cases}$$

כלומר, המינימום של $L(h)$ הוא עבור $h(x)=i$ שמקיים:

$h(x)=i=\arg\max_{i\in y} P(Y=i|X=x)$.

וזה מה שרצינו להוכיח. ∎

(b) נקבע $X=x$ ונקבל:

$$L(h)=E[\Delta(Y,h(X))]=P(X=x,Y=0)\cdot\Delta(0,h(x))+P(X=x,Y=1)\cdot\Delta(1,h(x))=$$

$$=P(X=x)\cdot P(Y=0|X=x)\cdot\Delta(0,h(x))+P(X=x)\cdot P(Y=1|X=x)\cdot\Delta(1,h(x))=$$

$$=P(X=x)\cdot\left(P(Y=0|X=x)\Delta(0,h(x))+P(Y=1|X=x)\Delta(1,h(x))\right)$$

$P(X=x)$ קבוע, לכן עלינו להגיע למינימום את הביטוי בסוגריים, [illegible] $P(X=x)$

עבור $h(x)=1$ נקבל: $a\cdot P(Y=0|X=x)$

ועבור $h(x)=0$ אנו נקבל: $b\cdot P(Y=1|X=x)$

מתי נעדיף לנחש $h(x)=1$:

$$b\cdot P(Y=1|X=x) > 4\cdot P(Y=0|X=x)$$

כלומר

$$\frac{b}{4} > \frac{P(Y=0|X=x)}{P(Y=1|X=x)}$$

אם לעומת $\frac{b}{4} < \frac{P(Y=0|X=x)}{P(Y=1|X=x)}$ נעדיף לנחש $h(x)=0$

**הערה:** אם $P(Y=1|X=x)=0$ אז לא נוכל לחלק ולכן נחשב את המקרה הזה בנפרד

במקרה זה, $h(x)=0$ כי ברור. נסכם את $h(x)$:

$$h(x)=\begin{cases} 0 & 0=P(Y=1|X=x) \lor \frac{b}{4} < \frac{P(Y=0|X=x)}{P(Y=1|X=x)} \\ 1 & \text{else} \end{cases}$$

**הערה נוספת:** במקרה של שוויון ניתן לבחור שרירותית אם לנחש 1 או 0...

$\square$

שאלה 2 (מודל גאוסיאני ומודל ...)

(a) לפי הגדרה, אם כל הפיצ'רים בלתי תלויים גם מתקיים:

$$\prod_{i=1}^{d} f(x_i, \mu_i, \Sigma_{i,i}) = \prod_{i=1}^{d}\frac{1}{\sqrt{2\pi\Sigma_{ii}}}e^{-\frac{(x_i-\mu_i)^2}{2\Sigma_{i,i}}} = \frac{1}{(2\pi)^{\frac{d}{2}}\left(\prod_{i=1}^{d}\Sigma_{i,i}\right)^{\frac{1}{2}}}e^{-\frac{1}{2}\sum_{i=1}^{d}\frac{(x_i-\mu_i)^2}{\Sigma_{i,i}}}$$

ומכיוון ש-$\Sigma$ אלכסונית כל שאר האיברים אפסים, $\left(\prod_{i=1}^{d}\Sigma_{ii}\right)^{\frac{1}{2}} = |\Sigma|^{\frac{1}{2}}$ וגם מתקיים:

$$\sum_{i=1}^{d}\frac{(x_i-\mu_i)^2}{\Sigma_{ii}} = (x-\mu)^T\Sigma^{-1}(x-\mu)$$

ולכן קיבלנו את הדרוש.

$\square$

(b) נוכיח את כלל ההחלטה. כי: לפי כלל בייס אופטימלי:

$P(Y=1|X=x) > P(Y=0|X=x)$ } נשתמש בבייס

$$\frac{f_{X|Y=1}(x)P(Y=1)}{\cancel{f_X(x)}} > \frac{f_{X|Y=0}(x)P(Y=0)}{\cancel{f_X(x)}}$$

} נסמן $P(Y=1):=p$

ונניח $\Sigma_0=\Sigma_1=\Sigma$

$$p\cdot f_1(x;\mu_1,\Sigma) > (1-p) f_0(x;\mu_0,\Sigma)$$

$$\frac{1-p}{p} < \frac{\frac{1}{\cancel{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}}} e^{\left(-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\right)}}{\frac{1}{\cancel{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}}} e^{\left(-\frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0)\right)}}$$

$$\frac{1-p}{p} < e^{-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)+\frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0)}$$

} נוציא ln ונכפיל ב-2

$$2\ln\left(\frac{1-p}{p}\right) < (x-\mu_0)^T\Sigma^{-1}(x-\mu_0) - (x-\mu_1)^T\Sigma^{-1}(x-\mu_1)$$

ונשים לב שזו מטריצה אלכסונית, לכן:

$$RHS = \sum_{i=1}^{d}\frac{(x_i-\mu_{0,i})^2}{\Sigma_{i,i}} - \sum_{i=1}^{d}\frac{(x_i-\mu_{1,i})^2}{\Sigma_{i,i}} = \sum_{i=1}^{d}\frac{\cancel{x_i^2}-2x_i\mu_{0,i}+\mu_{0,i}^2-\cancel{x_i^2}+2x_i\mu_{1,i}-\mu_{1,i}^2}{\Sigma_{i,i}} =$$

$$= \sum_{i=1}^{d}\frac{2x_i(\mu_{1,i}-\mu_{0,i}) - (\mu_{1,i}-\mu_{0,i})(\mu_{1,i}+\mu_{0,i})}{\Sigma_{i,i}} = \sum_{i=1}^{d}\frac{(\mu_{1,i}-\mu_{0,i})(2x_i-\mu_{1,i}-\mu_{0,i})}{\Sigma_{i,i}} =$$

$$= \underline{(\mu_1^T-\mu_0^T)\Sigma^{-1}(2x-\mu_0-\mu_1)}$$

כלומר הכלל שלנו נכון עבור x, נשים לב שאם (אם) המכפלה הנ"ל גדולה מ- $2\ln\left(\frac{1-p}{p}\right)$

אחרת נבחר לסווג 1 ואחרת 0.

$\square$

הערה: אם $0=P(Y=1)$ אז לכל x נסווג $h(x)=0$ ולכן נניח $0\neq P(Y=1)$

כדי שהשאלה תהיה מעניינת.

(5) נסמן $V=(\mu_1^T-\mu_0^T)\Sigma^{-1}$ את המקדמים כי $\sum_{i=1}^{d} v_i x_i = V\cdot x$. אזי, לפי סעיף קודם

מהאופן שמסווג קודם נקבל:

$$2\ln\left(\frac{1-p}{p}\right)=(\mu_1^T-\mu_0^T)\Sigma^{-1}(2x-\mu_0-\mu_1)$$

ואם נסמן את המקדם שמופיע בו ב-$a\in\mathbb{R}$ נקבל: $\sum_{i=1}^{d} v_i x_i - a = 0$ וזה מסווג

ליניארי כלומר מתאר היפר-מישור ב-$\mathbb{R}^d$.

$\square$

(1)

נא. נב. אם נניח ש $\Sigma_0=\sigma_0^2$, $\Sigma_1=\sigma_1^2$ אזי ~~[illegible]~~

נסווג ל-1 אם:

$$P(Y=1|X=x) > P(Y=0|X=x)$$

$$\frac{f_X(x|Y=1)\,P(Y=1)}{\cancel{f_X(x)}} > \frac{f(x|Y=0)P(Y=0)}{\cancel{f_X(x)}}$$

$$\frac{f_X(x|Y=1)}{f_X(x|Y=0)} > \frac{P(Y=0)}{P(Y=1)}$$

$$\Downarrow \quad p := P(Y=1)$$

$$\frac{\dfrac{1}{\sqrt{2\pi\sigma_1^2}}e^{-\frac{1}{2}(x-\mu_1)^T\sigma_1^{-2}(x-\mu_1)}}{\dfrac{e^{-\frac{1}{2}(x-\mu_0)^T\sigma_0^{-2}(x-\mu_0)}}{\sqrt{2\pi\sigma_0^2}}} > \frac{1-p}{p}$$

$$\cancel{e^{\frac{1}{2}(x-\mu_0)^T\sigma_0^{-2}(x-\mu_0)-\frac{1}{2}(x-\mu_1)\ldots} > \frac{1-p}{p}\cdot\frac{\sigma_1}{\sigma_0}}$$

$$e^{\frac{1}{2}(x-\mu_0)^T\sigma_0^{-2}(x-\mu_0)-\frac{1}{2}(x-\mu_1)^T\sigma_1(x-\mu_1)} > \frac{1-p}{p}\cdot\frac{\sigma_1}{\sigma_0}$$

$$\Downarrow \quad \mu_0=\mu_1=\mu$$

$$e^{\frac{1}{2}(x-\mu)^2\sigma_0^{-2}-\frac{1}{2}(x-\mu)\sigma_1^{-2}} > \frac{1-p}{p}\cdot\frac{\sigma_1}{\sigma_0}$$

$$\frac{(x-\mu)^2}{2}\cdot\left(\frac{1}{\sigma_0^2}-\frac{1}{\sigma_1^2}\right) > \ln\left(\frac{1-p}{p}\cdot\frac{\sigma_1}{\sigma_0}\right)$$

$$(x-\mu)^2 > \underbrace{\frac{2\sigma_0^2\sigma_1^2\ln\left(\frac{1-p}{p}\cdot\frac{\sigma_1}{\sigma_0}\right)}{\sigma_1^2-\sigma_0^2}}_{=:c}$$

כלומר אם $x>\mu+\sqrt{c}$ או $x<\mu-\sqrt{c}$ נסווג ל-1 ואחרת נסווג ל-0.

$\square$

אחד נימן

(a) [illegible] גוף נימן הגה.

(b) אני ~~[illegible]~~ מבנה שמחה (המוציא יבוא) כל ~~[illegible]~~ מספר בסיכוי $\frac{1}{10}$ (כלומר אחידה על המספרים). יש לי ~~[illegible]~~ 84.8% בעזרת התפלגות בינומית.

שבנינו

ריוק

(c) גוף מבצע. ~~[illegible]~~ יש שר-א בכו לא הוא [illegible] בכל אות כי המשוגע שה-א לפי לא יהיה באותו מקום אינו [illegible] דף-טל כי אם א אחד מבין אותם השופטים לגמרי, אם הוא קטן מבי לא נכון "לפעמים את כל החומר בכל אחד", אז המשמעות היא קרובה ~~[illegible]~~ הוא מקוששת אז במאה הבא.

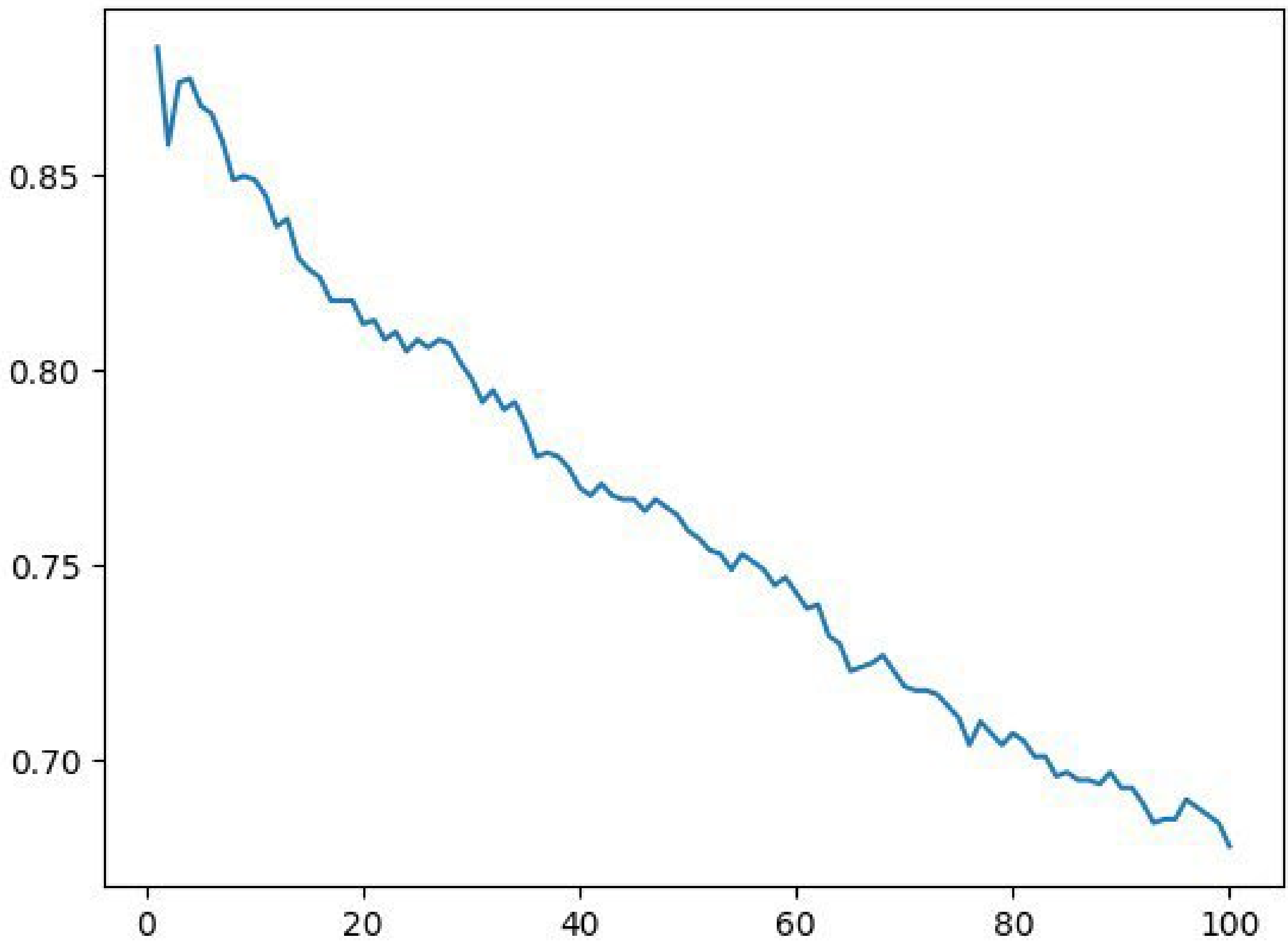
למרות ~~[illegible]~~, מעצמו שבוטף רק את אופטימה במבחן הזה...

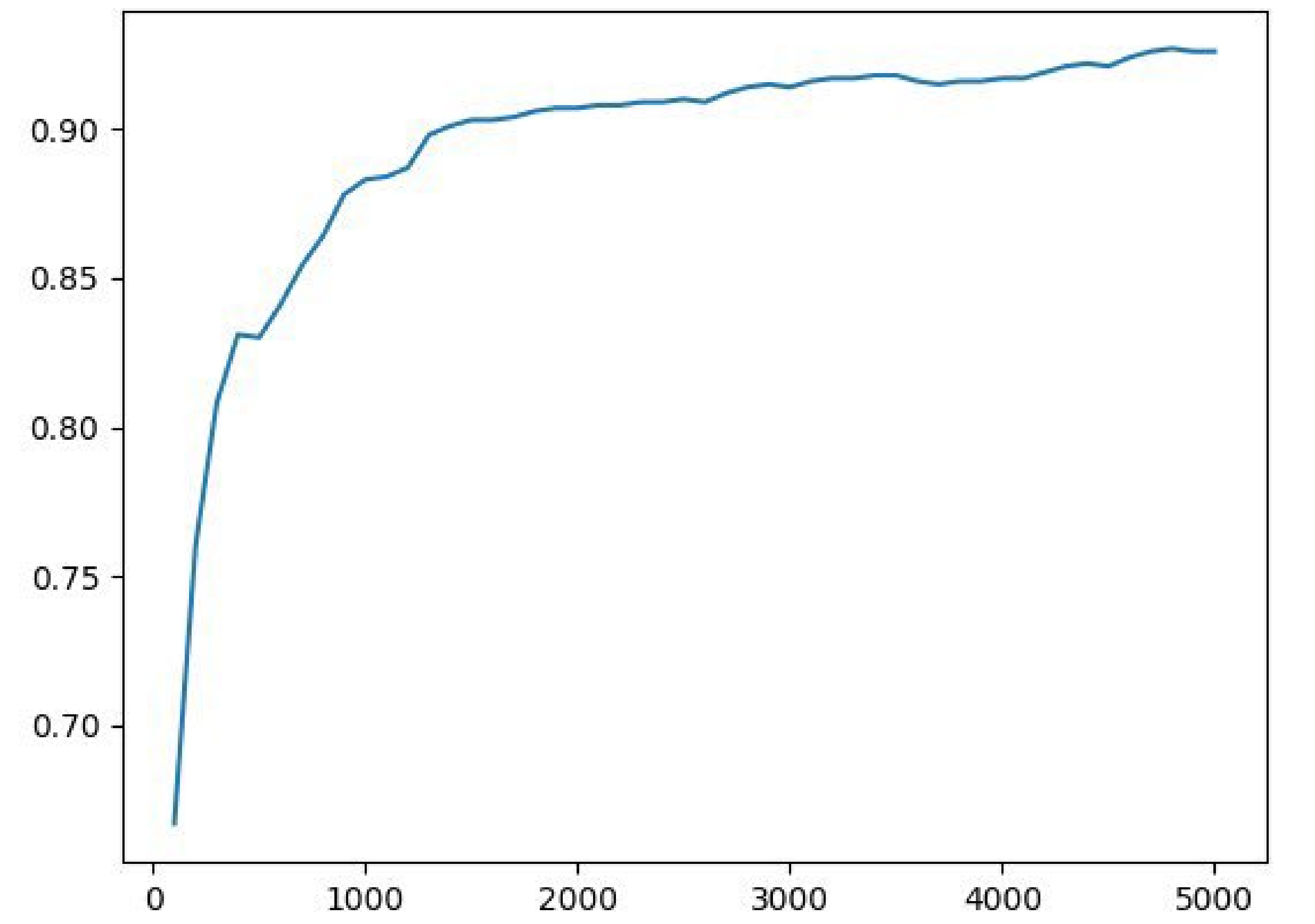
לגמ

(d) גוף מבצע. הייתי שלביון אל מבין שאמנו מוסמכים את המודל עם ~~[illegible]~~ יוצא מידע מיקר, וכל הבוגר ~~[illegible]~~ של האמות מכלל את כל הגבולות המקסימות ולכן גולמי שלא ישתבר כבר שמטיל את האמות למבחן. הוויק הכי ~~[illegible]~~ המגבלה עיון ~~[illegible]~~ h=4

טול

| |
|---|
| דרכים לאמולים הבאים, החשון לטורף c ולגשי לטורף b! |


0.85
0.80
0.75
0.70
0
20
40
60
80
100


0.90
0.85
0.80
0.75
0.70
0
1000
2000
3000
4000
5000