מבוא ללמידה חישובית – תרגיל בית 1

שם : דורי רימון

ת"ז : 323996843

אלגברה לינארית:

שאלה 1:

שאלה 1

צ"ל להוכיח כי עבור מטריצה סימטרית A , $A = xx^t \iff A$ הנה PSD.

$\Leftarrow$ נניח כי $A = xx^t$. אזי לכל וקטור v ,

$$v^t A v = v^t x x^t v = (x^t v)^t (x^t v) = \| x^t v \|^2 \geq 0$$

ומכאן ש- A הנה PSD ע"פ הגדרה.

$\Rightarrow$ נניח כי $v^t A v \geq 0$, $\forall v$. נסמן את הוקטורים העצמיים של A ב- $\{v_i\}_{i=1}^n$ ואת הע"ע המתאימים $\{\lambda_i\}_{i=1}^n$.

$$\forall 1 \leq i \leq n , \quad 0 \underset{*}{\leq} v_i^t A v_i = v_i^t \lambda_i v_i = \lambda_i \cdot \| v_i \|^2$$

$$\Rightarrow \lambda_i \geq 0$$

* הנחה, הגדרת הוקטור v_i

* במקרה שהמטריצה היא 0, לא ניתן לפרק את הכתובים. אם כך, במקרה זה הע"ע הם וקטור העצמי (אין המטריצה היא סימטרית R ממשית אי-שלילית) ומכאן גם הע"ע הוא 0, ובפרט ≤ 0.

$$\Rightarrow D = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} = \begin{pmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{pmatrix} \begin{pmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{pmatrix} =$$

$$= \hat{D} \cdot \hat{D} = \hat{D}^t \cdot \hat{D}$$

$$\Rightarrow A = Q D Q^t = Q \hat{D}^t \hat{D} Q^t = (\hat{D} Q^t)^t (\hat{D} Q^t) \underset{*}{=} x^t x$$

$$x = \hat{D} Q^t$$

* לצורך הנוחות ניתן היה לסמן $y = x^t$ ולהציג בצורה המבוקשת את הנוסחה - $A = yy^t$.

<u>הסתברות ואינפי:</u>

שאלה 1:

שאלה 1

נתחיל מפישוט הביטוי ואחר אנחנו נגזור

$$x^t A x = x^t \begin{pmatrix} \sum_{j=1}^{n} A_{1j} x_j \\ \vdots \\ \sum_{j=1}^{n} A_{nj} x_j \end{pmatrix} = \sum_{i=1}^{n} x_i \cdot \left(\sum_{j=1}^{n} A_{ij} x_j \right) =$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j$$

ולסיום, נבצע את הגזירה הדרושה:

$$\frac{\partial x^t A x}{\partial x_i} = \underbrace{1 \cdot \sum_{j=1}^{n} A_{ij} \cdot x_j}_{} + \underbrace{1 \cdot \sum_{m=1}^{n} A_{mi} x_m}_{} = \left((A + A^t) x \right)_i$$

הווקטור היחיד
אשר מתאימה
הראשונה

המחובר היחיד
שמתאימה השורה
כאשר לוקחים את
מהיפוך של סימנים
(ושאר אין נגזרה)

וקיבלנו את הדרוש.

שאלה 2:

שאלה 2

א) $R(\lambda) = \ln(E[e^{\lambda x}])$. נגזור פעם ראשונה:

$$R(\lambda)' = \frac{\frac{\partial E[e^{\lambda x}]}{\partial \lambda}}{E[e^{\lambda x}]}$$

מכלל השרשרת: $E[e^{\lambda x}] = \int_{-\infty}^{\infty} e^{\lambda x} f_x(x)\,dx$. הפונק' שבתוך האינטגרל, וגם

נגזרתה רציפות בכל ת-ה, נוכל לגזור את האינטגרל ע"י הכנסת הגזירה לתוך האינטגרל כמו שהוכח בשאלה הקודמת:

$$\frac{\partial E[e^{\lambda x}]}{\partial \lambda} = \frac{\partial}{\partial \lambda}\int_{-\infty}^{\infty} e^{\lambda x} f_x(x)\,dx =$$

$$= \int_{-\infty}^{\infty} \frac{\partial}{\partial \lambda} e^{\lambda x} f_x(x)\,dx =$$

$$= \int_{-\infty}^{\infty} x e^{\lambda x} f_x(x)\,dx = E[xe^{\lambda x}]$$

אם כן מהנגזרה הראשונה קיבלנו

$$R(\lambda)' = \frac{E[xe^{\lambda x}]}{E[e^{\lambda x}]}$$

נגזור את הנגזרה שנית:

$$R(\lambda)'' = \frac{E[xe^{\lambda x}]' \cdot E[e^{\lambda x}] - E[e^{\lambda x}]' \cdot E[xe^{\lambda x}]}{E[e^{\lambda x}]^2}$$

בנגזרה של $E[xe^{\lambda x}]$ נשתמש באותה הוכחה של הנגזרה של $E[e^{\lambda x}]$ שהוכחנו, רק שנקבל במקום $E[xe^{\lambda x}]' = E[x^2e^{\lambda x}]$.

$$\Rightarrow R(\lambda)'' = \frac{E[x^2e^{\lambda x}]\,E[e^{\lambda x}] - (E[xe^{\lambda x}])^2}{E[e^{\lambda x}]^2} =$$

$$= \frac{E[x^2e^{\lambda x}]}{E[e^{\lambda x}]} - \left(\frac{E[xe^{\lambda x}]}{E[e^{\lambda x}]}\right)^2$$

כפי שרצינו.

ב) נזכור תיאום השונות של המשתנה החדש, כדי לחשב תוחלת — ואת המומנט השני. נעשה זאת:

$$E_g[X] = \int_{-\infty}^{\infty} x \cdot \frac{e^{\lambda x}}{E_f[e^{\lambda x}]} f_X(x)\,dx =$$

$$= \frac{1}{E_f[e^{\lambda x}]} \cdot \int_{-\infty}^{\infty} x\, e^{\lambda x} f_X(x)\,dx = \frac{E[xe^{\lambda x}]}{E[e^{\lambda x}]}$$

$$E_g[X^2] = \int_{-\infty}^{\infty} x^2 \cdot \frac{e^{\lambda x}}{E_f[e^{\lambda x}]} f_X(x)\,dx =$$

$$= \frac{1}{E_f[e^{\lambda x}]} \cdot \int_{-\infty}^{\infty} x^2 \cdot e^{\lambda x} \cdot f_X(x)\,dx = \frac{E[x^2 e^{\lambda x}]}{E[e^{\lambda x}]}$$

ומנוסחת שונות / כפי שכתבנו כי $Var_g(X) = E_g[X^2] - E_g[X]^2 =$

$$= B(\lambda)''$$

כעת להראות את החסם הנדרש: $Var_g(X) \le B^2$

כיוון שהתומך של X הוא $[-B, B]$, ניתן להסיק ש $E_g[X^2] \le B^2$

$$\Rightarrow Var_g(X) = E_g[X^2] - E_g[X]^2 \le E_g[X^2] \le B^2$$

כנדרש.

ב)

$* \; R(\lambda)'' \leq B^2$

$* \; R(0) = \ln(\mathbb{E}[e^{0\cdot x}]) = \ln 1 = 0$

$$* \; R'(0) = \frac{\mathbb{E}[x e^{0x}]}{\mathbb{E}[e^{0\cdot x}]} = \mathbb{E}[x] = 0$$

עפ"י המשפט היסודי של החדו"א ; $\int_a^b f(x)\,dx = F(b) - F(a)$.

$$\Rightarrow \int_0^\lambda \int_0^t R''(s)\,ds\,dt = \int_0^\lambda R'(t) - R'(0)\,dt = R(\lambda) - R(0) = R(\lambda)$$

$$; \int_0^\lambda \int_0^t R''(s)\,ds\,dt \leq \int_0^\lambda \int_0^t B^2\,ds\,dt = \int_0^\lambda t \cdot B^2\,dt = B^2 \frac{\lambda^2}{2}$$

$$\Rightarrow \ln(\mathbb{E}[e^{\lambda x}]) \leq B^2 \frac{\lambda^2}{2} \Rightarrow \mathbb{E}[e^{\lambda x}] \leq e^{B^2 \cdot \frac{\lambda^2}{2}}, \quad \forall \lambda \in \mathbb{R}$$

כנני.

<u>כללי החלטה וגבולות ריכוז:</u>

שאלה 1:

שאלה 1

הגדרנו את $h = \arg\min_{f: X \to Y} \mathbb{E}[\ell_{0-1}(f(X), Y)]$.

נרצה להוכיח כי $h(x) = \arg\max_{i \in Y = \{1,\dots,L\}} \mathbb{P}(Y = i \mid X = x)$

קבועה וסגורה כרגע, נעבוד בהתניה ב- $X = \bar{x}$

$$\Rightarrow L(h) = \mathbb{E}[\ell_{0-1}(h(\bar{x}), Y)] = \sum_{i=1}^{L} \ell_{0-1}(h(\bar{x}), i) \cdot \mathbb{P}(X = \bar{x}, Y = i)$$

תלות h הפונק׳ ב y.

$$\bullet\ \mathbb{P}(X = \bar{x}, Y = i) = \mathbb{P}(X = \bar{x}) \cdot \mathbb{P}(Y = i \mid X = \bar{x})$$

$$\Rightarrow L(h) = \mathbb{P}(X = \bar{x}) \cdot \sum_{i=1}^{L} \ell_{0-1}(h(\bar{x}), i) \cdot \mathbb{P}(Y = i \mid X = \bar{x}) =$$

$$= \left\{ \mathbb{P}(X = \bar{x}) \cdot \sum_{\substack{i=1 \\ i \neq k}}^{L} \mathbb{P}(Y = i \mid X = \bar{x}) \quad h(\bar{x}) = k,\ k \in Y = \{1,\dots,L\} \right.$$

$$\ell_{0-1}(h(\bar{x}), i) = \begin{cases} 1 & h(\bar{x}) \neq i \\ 0 & h(\bar{x}) = i \end{cases}$$

$$= \left\{ \mathbb{P}(X = \bar{x}) \cdot (1 - \mathbb{P}(Y = k \mid X = \bar{x})) \quad h(\bar{x}) = k \right.$$

לכן הערך המינימלי של $L(h)$ יתקבל כאשר

$$h(\bar{x}) = i = \arg\max_{i \in Y} \mathbb{P}(Y = i \mid X = \bar{x})$$

כפי שרצינו.

שאלה 2:

שאלה 2

א) אין להגשה..

ב) נחשב את הפונק' :

$$f_0(x, \mu_0, \Sigma) = \frac{1}{2\pi \cdot \sqrt{2}} \cdot e^{-\frac{1}{2} x^t \Sigma^{-1} x} =$$

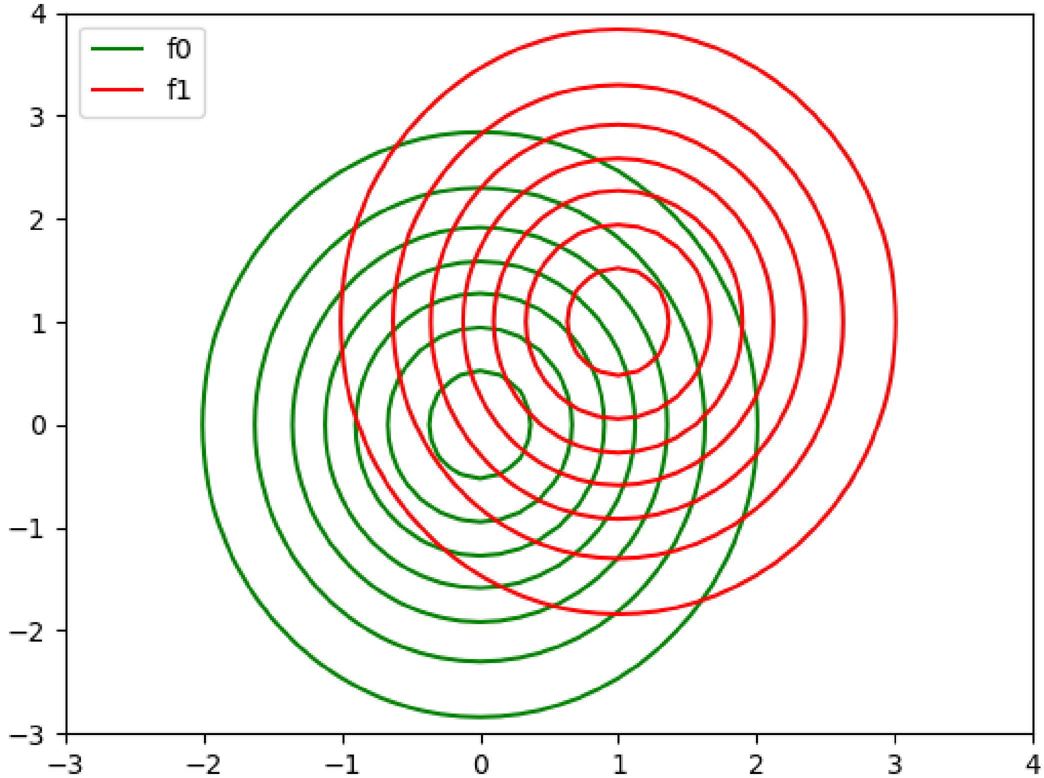
$$= \frac{1}{2\pi \cdot \sqrt{2}} \cdot e^{-\frac{1}{2} (x_1, x_2) \cdot \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}} =$$

$$= \frac{1}{2\pi \cdot \sqrt{2}} \cdot e^{-\frac{1}{2} (x_1^2 + \frac{1}{2} x_2^2)}$$

$$f_1(x, \mu_1, \Sigma) = \frac{1}{2\pi \cdot \sqrt{2}} \cdot e^{-\frac{1}{2} (x_1 - 1, x_2 - 1) \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} x_1 - 1 \\ x_2 - 1 \end{pmatrix}} =$$

$$= \frac{1}{2\pi \cdot \sqrt{2}} \cdot e^{-\frac{1}{2} ((x_1 - 1)^2 + \frac{1}{2} (x_2 - 1)^2)}$$

לסיום, הכנסתי את הפונק' הזו לקוד היופך את הגרף הרצוי.


f0
f1
4
3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3
4

5)

$$\mathbb{P}(y=1 \mid x) > \mathbb{P}(y=0 \mid x)$$

נוסחת בייס $\iff \dfrac{f_{X|y=1}(x)\,\mathbb{P}(y=1)}{f_X(x)} > \dfrac{f_{X|y=0}(x)\cdot\mathbb{P}(y=0)}{f_X(x)}$

$\mathbb{P}(y=1)=p$, $y\in\{0,1\}$ $\iff p\cdot f_1(x,\mu_1,\underset{\Sigma}{\overset{\shortparallel}{\Sigma_1}}) > (1-p)\cdot f_0(x,\mu_0,\underset{\Sigma}{\overset{\shortparallel}{\Sigma_0}})$

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

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

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

נוכל לפתח את הביטוי שבאגף ימין עוד יותר. לא נראה כמו שיך, אבל נקבל:

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

וקיבלנו ביטוי מסוג ולמד.

ד)

ראינו בכיתה כי עבור $d=1$ נקבל נקודה. ניתן אינטואיטיבית — להבין כי עבור $d=2$ נקבל ישר.

פורמלית, נקבל שוויון מתוך השוויון מהסעיף הקודם:

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

$$\Leftrightarrow \underbrace{2(\mu_1^t - \mu_0^t)\,\Sigma^{-1}x}_{\text{וקטור}} = \underbrace{2\ln\left(\frac{1-p}{p}\right) + (\mu_1^t - \mu_0^t)\,\Sigma^{-1}\cdot(\mu_0 + \mu_1)}_{\text{סקלר}}$$

כלומר הפתרון הוא מישור d - ממדי.

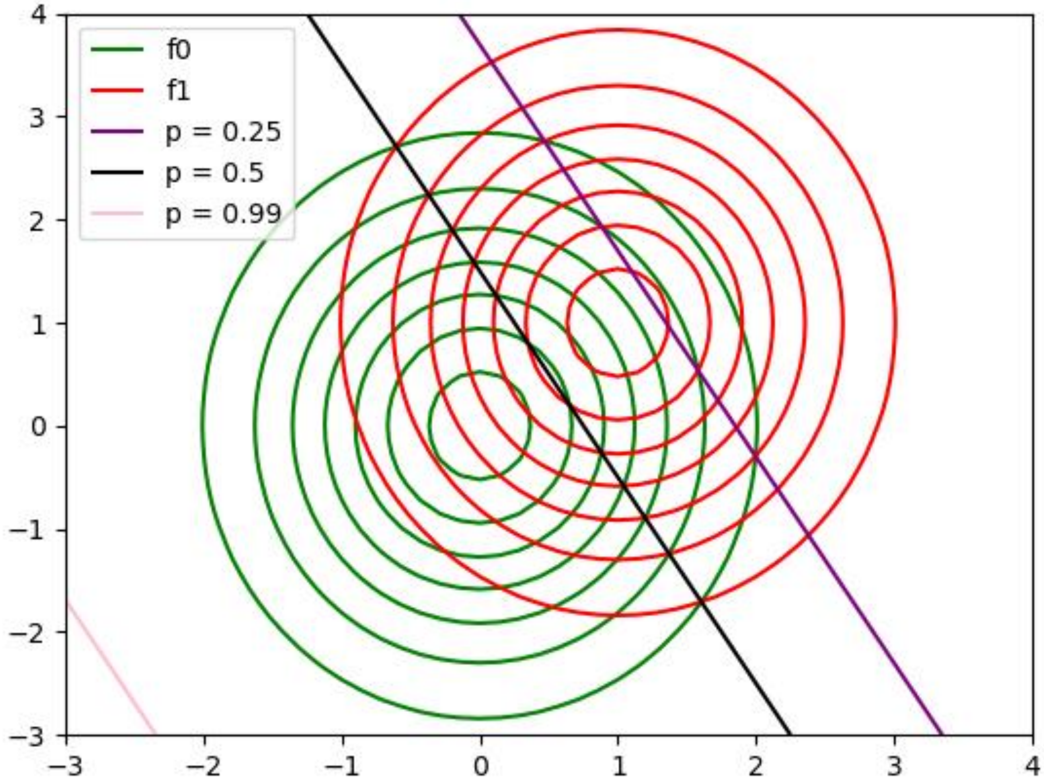
ה) מהסעיף הקודם קיבלנו משוואה. נציב את הנתונים הנתונים Σ, μ_1, μ_0:

$$(2,2)\cdot\begin{pmatrix}1 & 0\\ 0 & 1/2\end{pmatrix}x = 2\ln\left(\frac{1-p}{p}\right) + (1,1)\begin{pmatrix}1 & 0\\ 0 & 1/2\end{pmatrix}\begin{pmatrix}1\\ 1\end{pmatrix}$$

$$\Rightarrow (2,1)\,x = 2\ln\left(\frac{1-p}{p}\right) = \left(1,\frac{1}{2}\right)\begin{pmatrix}1\\ 1\end{pmatrix}$$

$$\Rightarrow 2x_1 + x_2 = 2\ln\left(\frac{1-p}{p}\right) + \frac{3}{2}$$

נוסיף את הישר לכל אחד מהערכים $p=0.25, 0.5, 0.99$.



f0
f1
p = 0.25
p = 0.5
p = 0.99
4
3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3
4

1)

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

$$\Leftrightarrow \frac{1-p}{p} \cdot \frac{\sigma_0}{\sigma_1} = e^{\frac{(x-\mu)^2}{2}\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)}$$

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

$$\vdots$$

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

נקודות של חיתוך.

שאלה 3:

שאלה 3

נתון $S = X_1 + \dots + X_n$ ואנו רוצים למצוא $N \in \mathbb{N}$ כך שלכל 6 $n \geq N$,

$$\mathbb{P}(S > n^2 + 0.2n) < 0.1$$

$$* \ \mathbb{E}[X_i] = \frac{-3+5}{2} = 1 \ , \ \forall\, 1 \leq i \leq n$$

כפי שהגדרנו בכיתה, נגדיר $\bar{X} = \frac{S}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i$.

$$* \ \mathbb{E}[\bar{X}] = \frac{1}{n} \cdot n = 1$$

$$\Rightarrow \mathbb{P}(|\bar{X} - 1| \geq \varepsilon) \leq 2e^{-\frac{\varepsilon^2}{2(5+3)^2} \cdot n}$$

הופדינג

כעת,

$$\mathbb{P}(S > n^2 + 0.2n) = \mathbb{P}\left(\frac{S}{n} > n + 0.2\right) =$$

$$= \mathbb{P}(\bar{X} - 1 > n - 0.8) \leq$$

מונוטוניות — פונק' ההסתברות — $\leq \mathbb{P}(|\bar{X} - 1| > n - 0.8) \leq$

הופדינג $\leq 2e^{-\frac{n(n-0.8)^2}{2 \cdot 8^2}} := \alpha$

נרצה לדרוש $\alpha < 0.1$.

$$\Rightarrow -\frac{n(n-0.8)^2}{2 \cdot 8^2} < \ln(0.05)$$

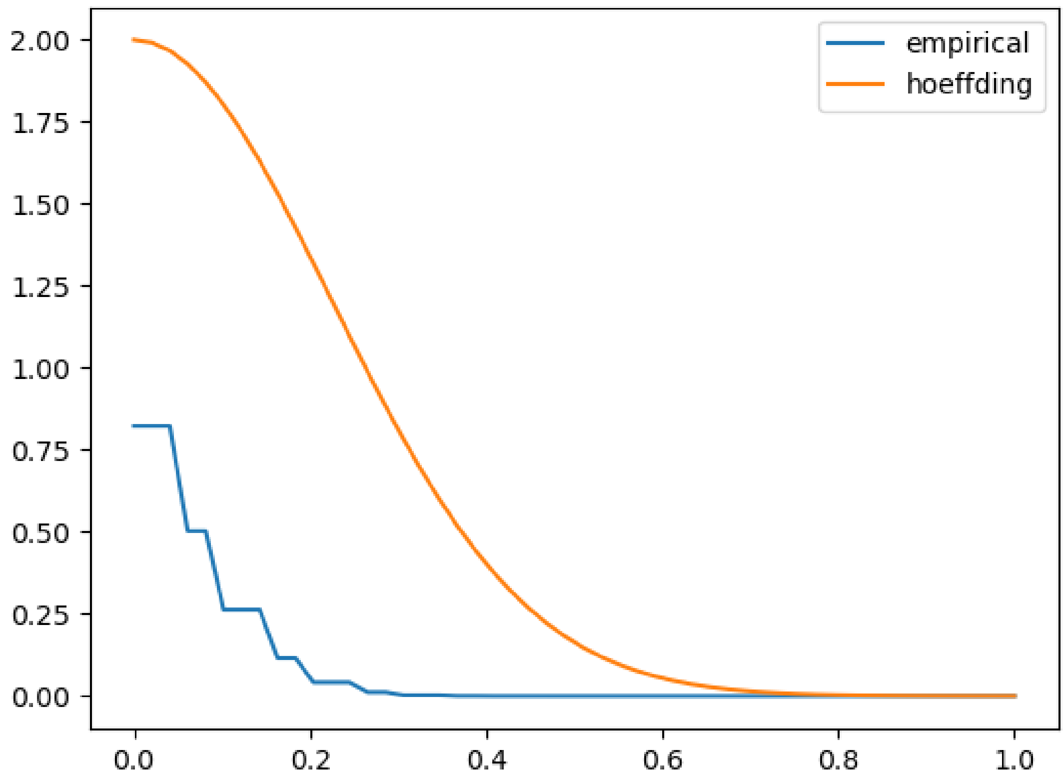
$$\Rightarrow n(n-0.8)^2 > -2 \cdot 8^2 \cdot \ln(0.05) \approx 383.45$$

אין שאלה שלמה. מבדיקה ידנית — פשוטה נמצאנו כי ה- N הרצוי הוא

$$N = 8$$

וקיבלנו את הרצוי.

שאלה 4:


empirical
hoeffding
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
0.0
0.2
0.4
0.6
0.8
1.0

<u>מטלה מעשית:</u>

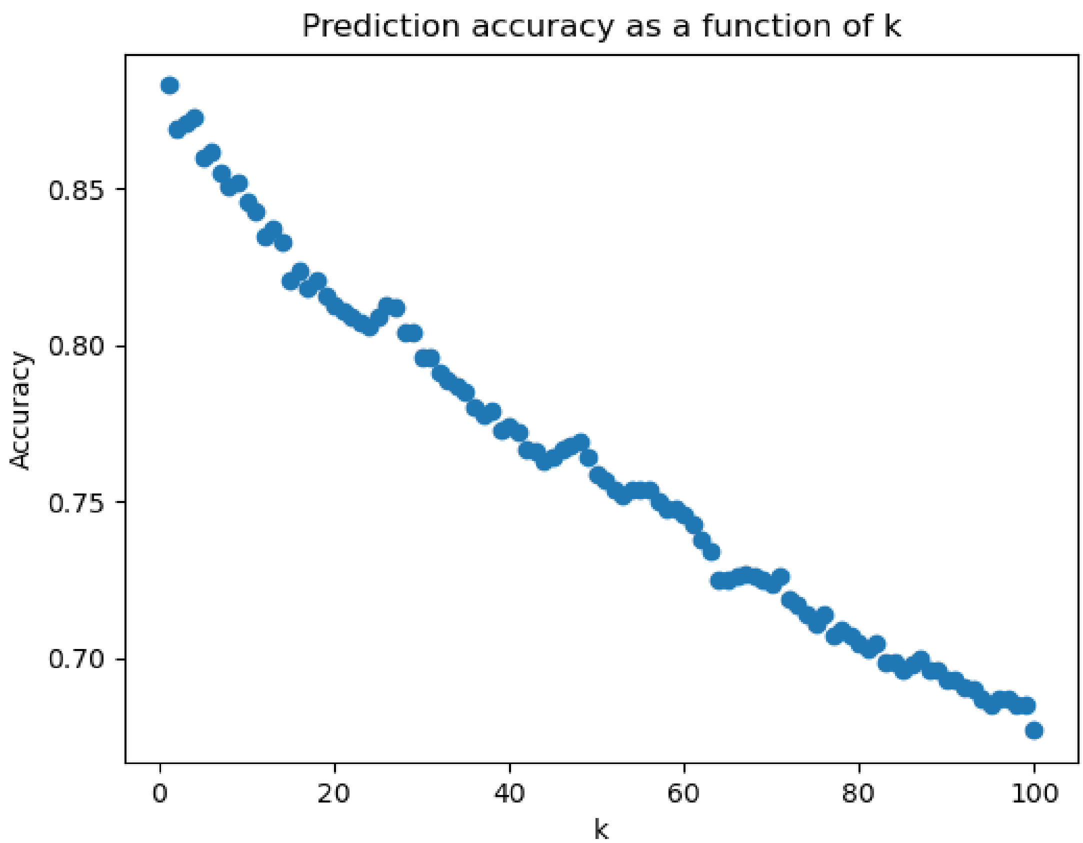
א. הקוד במודל, תחת השם knn.py. הפונק' הרלוונטית לסעיף זה היא knn.

ב. הדיוק תחת הנתונים הדרושים (n = 1000, k = 10) היה 84.6%.

הייתי מצפה מפרידיקט רנדומלי דיוק של כ – 10%. זאת כיוון שעבור כל תמונה, תחת הנחת התפלגות אחידה, הסיכוי של ניחוש רנדומלי להיות נכון הוא 1/10 (שכן ישנם עשרה קלאסטרים – 10 ספרות).

<u>הערה</u>: הקוד לסעיף זה נמצא גם הוא באותו הקובץ. הפונק' הרלוונטית היא predict.

ג.



בתוצאות ניתן לראות כי באופן כללי, ככל ש k גדול יותר, כך יורדת רמת הדיוק. בפרט, הדיוק הגבוה ביותר התקבל עבור k =1.

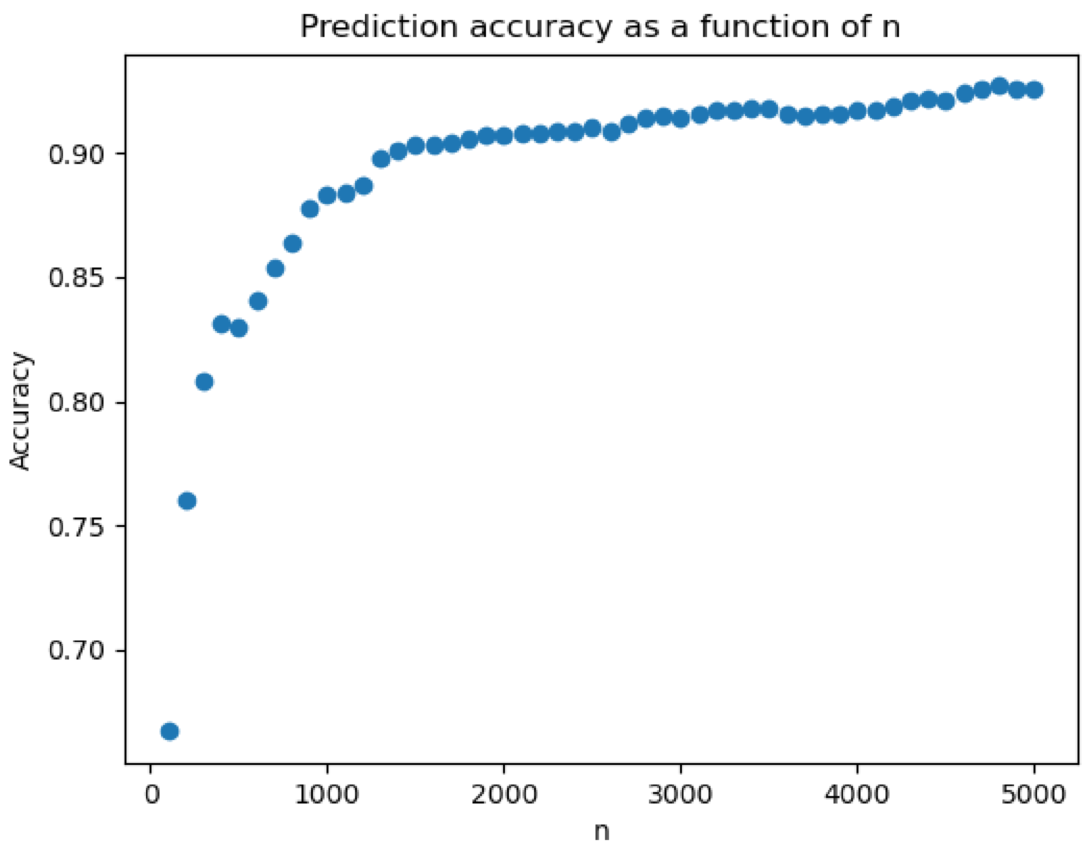
תחילה התוצאות הפתיעו אותי, אך אנסה להסבירן באופן הבא: באופן כללי ב data שניתן לאלגוריתם, יש דמיון גבוה בין ה – train data וה – test data.

על כן, בהתקבל קלט חדש, הוקטור הקרוב ביותר אליו הוא מדד איכותי ביותר. זאת ועוד, שהוספת וקטורים רחוקים יותר למדידות רק תפגע באיכות התוצאה.

מכאן שעבודה עם k =1, כלומר התייחסות רק לוקטור הקרוב ביותר, תיתן את התוצאות הטובות ביותר.

<u>הערה</u>: הקוד לסעיף זה נמצא גם הוא באותו הקובץ. הפונק' הרלוונטית היא plot_k_accuracy.

ד.



בתוצאות ניתן לראות כי באופן כללי, ככל ש n גדול יותר, כך עולה רמת הדיוק. בפרט, הדיוק הגבוה ביותר התקבל עבור $n = 5000$ (ייתכן כי עבור $n = 4800$, אך כוונתי פה הייתה לדון בתוצאות בצורה מעט יותר סכמטית).

ראשית ברור כי יותר תמונות לאימון ישפרו את דיוק האלגוריתם, שכן בעת קבלת תמונה חדשה, יש סיכוי גבוה יותר שתהיה תמונה עליה התאמנו הנמצאת באותו הקלאסטר וקרובה לתמונה הנמדדת.

דבר מעניין להבחין בו הוא כי מגמת השיפור הופכת להיות משמעותית קטנה יותר בערך כאשר $n > 1400$. הגיוני כי החל מנקודה מסויימת השיפוע יקטן שכן בשלב מסוים יהיו מספיק תמונות מקלאסטרים שונים שנמדדו כך שבסיכוי גבוה תמונה חדשה תמצא קרובה לתמונה שנמדדה כבר ונמצאת באותו הקלאסטר. במצב זה, הוספת תמונות נוספות תתרום פחות משתרמה בהתחלה.

<u>הערה</u>: הקוד לסעיף זה נמצא גם הוא באותו הקובץ. הפונק' הרלוונטית היא plot_n_accuracy.