

Intro to Machine Learning: Assignment #1

May 2, 2023

Dor Bourshan

1 Linear Algebra

1.1 Problem

תזכורת: מטריצה סימטרית A מעל \mathbb{R} ניתנת לכתיבה הבאה $A = QDQ^T$ כאשר :

- D מטריצה אלכסונית, ובאלכסון נמצאים הערכים העצמיים של D
- המטריצה Q היא מטריצה אורתוגונלית כאשר עמודות המטריצה הם וקטורים עצמיים.

סימון: עבור מטריצה B עם ערכים אי שלילים נסמן \sqrt{B} את המטריצה כך שהערך בתא (i, j) הוא $(\sqrt{b})_{i,j}$.

נראה שעבור מטריצה סימטרית A מעל \mathbb{R} התנאים הבאים שקולים :

- המטריצה חיובית.
- ניתן לכתוב $A = XX^T$.
- הערכים העצמיים של A אי שלילים.

נחל בשרשרת גרירות מעגלית :

- נניח הערכים העצמיים של A אי שלילים ונראה כי ניתן לכתוב $A = XX^T$.
מכך שהמטריצה A ניתנת לכתיבה $A = QDQ^T$ כאשר המטריצה D אלכסונית וכן באלכסון של D מופיעים ערכים אי שלילים, נטען כי השוויון הבא תקף $D = \sqrt{D}\sqrt{D}$. לפיכך A ניתנת לכתיבה כ $A = Q\sqrt{D}\sqrt{D}Q^T$. מהיותה של D אלכסונית כך גם \sqrt{D} ולכן מתקיים $\sqrt{D} = (\sqrt{D})^T$.

כידוע עבור מטריצות A, B מתקיים $(AB)^T = B^T A^T$ אם כך מתקיים :

$$\sqrt{D}Q^T = (\sqrt{D})^T Q^T = (Q\sqrt{D})^T$$

נסמן $X = Q\sqrt{D}$. נוכל לרשום

$$A = XX^T$$

כנדרש.

- נניח כי ניתן לכתוב $A = XX^T$ ונראה המטריצה A חיובית.
יהי $\mathbf{v} \in \mathbb{R}^n$, אנו מעוניינים להראות כי $\mathbf{v}^T A \mathbf{v} \geq 0$ ואכן

$$\mathbf{v}^T A \mathbf{v} = \mathbf{v}^T X X^T \mathbf{v} \stackrel{(1)}{=} (X^T \mathbf{v})^T (X^T \mathbf{v}) = \langle X^T \mathbf{v}, X^T \mathbf{v} \rangle \stackrel{(2)}{\geq} 0$$

כאשר :

(1) : תכונות transpose

(2) : תכונת מרחב מכפלה פנימית.

- תהיי המטריצה A מטריצה חיובית ונראה כי הערכים העצמיים של A אי שלילים.
יהי $\vec{0} \neq \mathbf{v} \in \mathbb{R}^n$ וקטור עצמי המתאים לערך עצמי λ . מהגדרת מטריצה חיובית לכל וקטור \mathbf{x} מתקיים כי $\mathbf{x}^T A \mathbf{x} \geq 0$ ולכן מתקיים כי:

$$0 \leq \mathbf{v}^T A \mathbf{v} = \mathbf{v}^T \lambda \mathbf{v} = \lambda \mathbf{v}^T \mathbf{v} = \lambda \langle \mathbf{v}, \mathbf{v} \rangle$$

מכך ש $\vec{0} \neq \mathbf{v}$ מתקיים כי $\langle \mathbf{v}, \mathbf{v} \rangle > 0$ מהגדרת מכפלה פנימית ולכן

$$\lambda \langle \mathbf{v}, \mathbf{v} \rangle \geq 0 \implies \lambda \geq 0$$

כדי להשלים את התמונה הע"ע המתאים לו"ע $\vec{0}$ הוא $\lambda = 0 \geq 0$.

לסיכום הערכים העצמיים של A אי שלילים.

□

1.2 Problem

יהיו $\alpha, \beta \geq 0$ ו $A, B \in \mathbb{R}^{n \times n}$ מטריצות חיוביות.

נראה כי $\alpha A + \beta B$ מטריצה חיובית.

יהי $\mathbf{v} \in \mathbb{R}^n$.

$$\mathbf{v}^T (\alpha A + \beta B) \mathbf{v} \stackrel{(1)}{=} (\alpha \mathbf{v}^T A + \beta \mathbf{v}^T B) \mathbf{v} \stackrel{(2)}{=} \alpha \mathbf{v}^T A \mathbf{v} + \beta \mathbf{v}^T B \mathbf{v} \stackrel{(3)}{\geq} 0$$

כאשר :

(1 + 2) : אסוציאטיבות ודיסטרטיביות של כפל מטריצות

(3) : מהיותן של A, B מטריצות חיוביות וכן $\alpha, \beta \geq 0$ מתקיים :

$$\mathbf{v}^T A \mathbf{v} \geq 0 \implies \alpha \mathbf{v}^T A \mathbf{v} \geq 0$$

$$\mathbf{v}^T B \mathbf{v} \geq 0 \implies \beta \mathbf{v}^T B \mathbf{v} \geq 0$$

□

נשים לב כי עבור $A = I$, המטריצה חיובית, כיוון שערכיה העצמיים הם חיוביים. מנגד, המטריצה αI עבור $\alpha < 0$ אינה חיובית כיוון שערכיה העצמיים שלילים.

לפיכך אין שמירה על אקסיומת כפל בסקלר.

2 Calculus and Probability

יהיו משתנים מקרים בלתי תלויים רציפים מהתפלגות $U([0, 1])$. נגדיר

$$Y = \max(X_1, \dots, X_n)$$

2.1 Problem

נמצא את ה PDF של Y , $\mathbb{E}[Y]$ וכן $\text{Var}(Y)$.

• נמצא את PDF

כידוע עבור $X \sim U([0, 1])$ מתקיים

$$F_X(x) = \begin{cases} x & 0 \leq x \leq 1 \\ 1 & x > 1 \\ 0 & x < 0 \end{cases}$$

לפיכך

$$\Pr(Y \leq y) = \Pr(\max(X_1, \dots, X_n) \leq y) = \Pr(X_1 \leq y, \dots, X_n \leq y) \stackrel{\text{i.i.d.}}{=} \prod_{i=1}^n \Pr(X_i \leq y) = (F_X(x))^n$$

אם כן

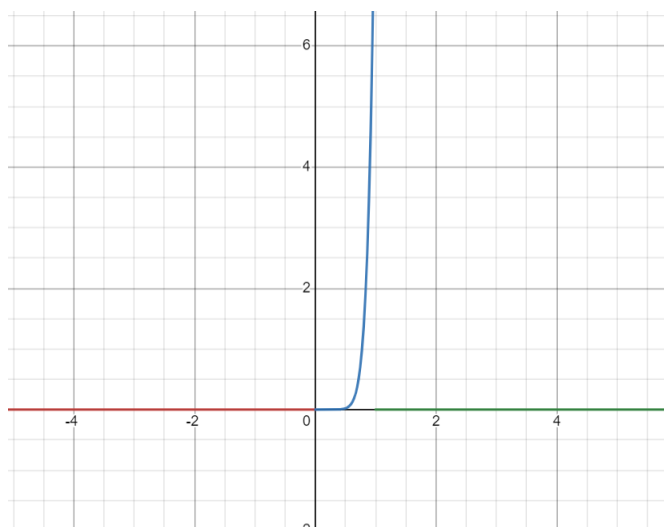
$$F_Y^n(x) = \begin{cases} x^n & 0 \leq x \leq 1 \\ 1 & x > 1 \\ 0 & x < 0 \end{cases}$$

אנו יודעים כי

$$f_X(x) = \frac{\partial}{\partial x} F_X(x)$$

ולכן

$$f_Y(y) = \begin{cases} ny^{n-1} & 0 \leq y \leq 1 \\ 0 & y > 1 \\ 0 & y < 0 \end{cases}$$



:1 Plot PDF

• נמצא $\mathbb{E}[Y]$

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} y \cdot f_Y(y) dy = \int_{-\infty}^0 y \cdot f_Y(y) dy + \int_0^1 y \cdot f_Y(y) dy + \int_1^{\infty} y \cdot f_Y(y) dy$$

$$\boxed{\mathbb{E}[Y] = \frac{n}{n+1}}$$

מתקיים כי

$$\int_{-\infty}^0 y \cdot 0 dy = \int_1^{\infty} y \cdot 0 dy = 0$$

ולכן

$$\mathbb{E}[Y] = \int_0^1 y \cdot f_Y(y) dy = \int_0^1 y \cdot ny^{n-1} dy = \int_0^1 ny^n dy = \frac{n \cdot y^{n+1}}{n+1} \Big|_0^1 = \frac{n}{n+1}$$

$$\boxed{\mathbb{E}[Y] = \frac{n}{n+1}}$$

• נמצא $\text{Var}(Y)$

כידוע מתקיים כי

$$\text{Var}(Y) = \mathbb{E}[Y^2] - \mathbb{E}^2[Y]$$

אם כך נמצא $\mathbb{E}[Y^2]$ כאשר ההסברים עבור המעברים דומים מאוד לעיל ולכן נקצר

$$\mathbb{E}[Y^2] = \int_0^1 y^2 \cdot f_Y(y) dy = \int_0^1 y^2 \cdot ny^{n-1} dy = \int_0^1 ny^{n+1} dy = \frac{n \cdot y^{n+2}}{n+2} \Big|_0^1 = \frac{n}{n+2}$$

מתקיים אם כך

$$\text{Var}(Y) = \mathbb{E}[Y^2] - \mathbb{E}^2[Y] = \frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2$$

$$\boxed{\text{Var}(Y) = \frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2}$$

• אסימפטוטיקה

$$\mathbb{E}[Y] = \mathbb{E}[\max(X_1, \dots, X_n)] = \frac{n}{n+1} \xrightarrow{n \rightarrow \infty} 1$$

$$\text{Var}(Y) = \frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2 \xrightarrow{n \rightarrow \infty} 1 - 1 = 0$$

□

3 Optimal Classifiers and Decision Rules

3.1 Problem

3.1.1 section

יהיו X, Y משתנים מקריים כך ש Y מקבל ערכים ב $\mathcal{Y} = \{1, \dots, L\}$. נגדיר 0-1 loss function ℓ_{0-1} שהוגדרה בכיתה.

נראה כי

$$h = \arg \min_{f: \mathcal{X} \rightarrow \mathcal{Y}} \mathbb{E} [\ell_{0-1}(Y, f(X))]$$

נתונה ע"י

$$h(x) = \arg \max_{i \in \mathcal{Y}} \Pr[Y = i \mid X = x]$$

לפי ההגדרה

$$\mathbb{E} [\ell_{0-1}(Y, f(X))] = \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \Pr[X = x, Y = y] \ell_{0-1}(y, f(x))$$

כאשר אנו משתמשים בהצגה זו רק לצורך הצגה נוחה וכן יכול להיות כי הקבוצה \mathcal{X} אינה בת מנייה.

נקבע x . לפיכך אנו נדרשים למצוא f המביאה למינימום את הביטוי הבא :

$$\sum_{y \in \mathcal{Y}} \Pr[X = x, Y = y] \ell_{0-1}(y, f(x)) = \sum_{y \in \mathcal{Y}} \Pr[Y = y \mid X = x] \Pr[X = x] \cdot \ell_{0-1}(y, f(x))$$

כעת נשים לב כי $\Pr[X = x]$ קבוע אי שלילי ולכן אנו נדרשים למעשה למנמם את הסכום הבא :

$$\sum_{y \in \mathcal{Y}} \Pr[Y = y \mid X = x] \ell_{0-1}(y, f(x))$$

מתקיים כי $\Pr[Y = y \mid X = x] \geq 0$ ולכן אנו נרצה להתעלם מהערך y כך ש $\Pr[Y = i \mid X = x]$ גדול ביותר עבורו (כאשר השתמשנו בהגדרת/ מסקנה מהגדרת ℓ_{0-1}). אם כן :

$$h(x) = \arg \max_{i \in \mathcal{Y}} \Pr[Y = i \mid X = x]$$

3.1.2 section

יהיו X, Y משתנים מקריים כך ש Y מקבל ערכים ב $\mathcal{Y} = \{0, 1\}$. נגדיר Δ פונקציית loss אסימטרית:

$$\Delta(y, \hat{y}) = \begin{cases} 0 & y = \hat{y} \\ a & y = 0, \hat{y} = 1 \\ b & y = 1, \hat{y} = 0 \end{cases}$$

כאשר $a, b \in (0, 1]$.

נמצא

$$h = \arg \min_{f: \mathcal{X} \rightarrow \mathcal{Y}} \mathbb{E}[\Delta(Y, f(X))]$$

לפי ההגדרה

$$\mathbb{E}[\Delta(Y, f(X))] = \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \Pr[X = x, Y = y] \Delta(y, f(x))$$

כאשר אנו משתמשים בהצגה זו רק לצורך הצגה נוחה וכן יכול להיות כי הקבוצה \mathcal{X} אינה בת מנייה.

נקבע x . לפיכך אנו נדרשים למצוא f המביאה למינימום את הביטוי הבא:

$$\sum_{y \in \mathcal{Y}} \Pr[X = x, Y = y] \Delta(y, f(x)) = \sum_{y \in \mathcal{Y}} \Pr[Y = y | X = x] \Pr[X = x] \cdot \Delta(y, f(x))$$

כעת נשים לב כי $\Pr[X = x]$ קבוע אי שלילי ולכן אנו נדרשים למעשה למנמם את הסכום הבא:

$$\Pr[Y = 0 | X = x] \Delta(0, f(x)) + \Pr[Y = 1 | X = x] \Delta(1, f(x))$$

האפשרויות שלנו נתונות ע"י

$$\Pr[Y = 0 | X = x] \Delta(0, 1) + \Pr[Y = 1 | X = x] \Delta(1, 1) \quad (1)$$

$$\Pr[Y = 0 | X = x] \Delta(0, 0) + \Pr[Y = 1 | X = x] \Delta(1, 0) \quad (2)$$

כלומר

$$\Pr[Y = 0 | X = x] \cdot a \quad (1)$$

$$\Pr[Y = 1 | X = x] \cdot b \quad (2)$$

אם כן כדי למנמם נגדיר:

$$h(x) = \begin{cases} 0 & \Pr[Y = 0 | X = x] \cdot a > \Pr[Y = 1 | X = x] \cdot b \\ 1 & \text{otherwise} \end{cases}$$

3.2 Problem 2

יהיו X, Y משתנים מקריים כך ש Y מקבל ערכים ב $\mathcal{Y} = \{0, 1\}$ ו X מקבל ערכים ב \mathcal{X} . נניח כי אנו מעוניינים למצוא פונקציית חיזוי $h: \mathcal{X} \rightarrow [0, 1]$ נגדיר Δ_{log}

$$\Delta_{log}(y, \hat{y}) = -y \log(\hat{y}) - (1 - y) \log(1 - \hat{y})$$

נמצא

$$h = \arg \min_{f: \mathcal{X} \rightarrow [0, 1]} \mathbb{E}[\Delta_{log}(Y, f(X))]$$

לפי ההגדרה

$$\mathbb{E}[\Delta(Y, f(X))] = \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \Pr[X = x, Y = y] \Delta_{log}(y, f(x))$$

כאשר אנו משתמשים בהצגה זו רק לצורך הצגה נוחה וכן יכול להיות כי הקבוצה \mathcal{X} אינה בת מנייה.

נקבע x . לפיכך אנו נדרשים למצוא f המביאה למינימום את הביטוי הבא :

$$\sum_{y \in \mathcal{Y}} \Pr[X = x, Y = y] \Delta_{log}(y, f(x)) = \sum_{y \in \mathcal{Y}} \Pr[Y = y | X = x] \Pr[X = x] \cdot \Delta_{log}(y, f(x))$$

כעת נשים לב כי $\Pr[X = x]$ קבוע אי שלילי ולכן אנו נדרשים למעשה למנמם את הסכום הבא :

$$\Pr[Y = 0 | X = x] \Delta_{log}(0, f(x)) + \Pr[Y = 1 | X = x] \Delta_{log}(1, f(x))$$

נסמן לשם הנוחות

$$\Pr[Y = 0 | X = x] = p_0 \quad (1)$$

$$\Pr[Y = 1 | X = x] = p_1 \quad (2)$$

כעת נכתוב מחדש את הסכום עם כתיבה מפורשת של Δ_{log}

$$-p_0 \cdot \log(1 - f(x)) + p_1 - p_1 \cdot \log(f(x))$$

ברור כי $0 < f(x) < 1$ אחרת הביטוי אינו מוגדר.

כעת נגזור לפי $f(x)$ ונשווה ל 0 כדי למצוא מינימום, ונקבל

$$\frac{p_0}{1 - f(x)} - \frac{p_1}{f(x)} = 0 \implies f(x) = p_1$$

כאשר השתמשנו בעובדה כי $p_0 + p_1 = 1$.

נוודא כי אכן מינימום ע"י גזירה נוספת :

$$\left(\frac{p_0}{1-f(x)} - \frac{p_1}{f(x)} \right)' = \frac{p_0}{(1-f(x))^2} + \frac{p_1}{(f(x))^2}$$

כעת נציב $f(x) = p_1$ ואנו מניחים כי $0, 1 \neq p_1$ ונקבל כי

$$\frac{p_0}{(1-p_1)^2} + \frac{p_1}{(p_1)^2} > 0$$

ולכן התקבל מינימום.

אם כך נגדיר

$$h(x) = \Pr[Y = 1 \mid X = x] = p_1$$

3.3 Problem

יהיו X, Y משתנים מקריים כך ש Y מקבל ערכים ב $\mathcal{Y} = \{0, 1\}$ וכן $\mathcal{X} = \mathbb{R}$ בהתאמה. נניח כי מתקיים $X_{|Y=0} \sim \mathcal{N}(\mu, \sigma_0^2)$ וכן $X_{|Y=1} \sim \mathcal{N}(\mu, \sigma_1^2)$ כאשר $\sigma_0 \neq \sigma_1$. נניח כי $\Pr[Y = 1] = p_1$.

נמצא את מסווג אופטימלי $h : \mathbb{R} \rightarrow \{0, 1\}$ שמביא למינימום את $\mathbb{E}[\ell_{0-1}(Y, h(X))]$

לפי ההגדרה אם כן :

$$\mathbb{E}[\ell_{0-1}(Y, h(X))] = \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \Pr[X = x, Y = y] \ell_{0-1}(y, h(x))$$

נקבע x

$$\begin{aligned} \sum_{y \in \mathcal{Y}} \Pr[X = x, Y = y] \ell_{0-1}(y, h(x)) &= \\ &= f_{X|Y=0}(x) \Pr[Y = 0] \ell_{0-1}(0, h(x)) + f_{X|Y=1}(x) \Pr[Y = 1] \ell_{0-1}(1, h(x)) \\ &= f_{X|Y=0}(x) \cdot (1 - p_1) \cdot \ell_{0-1}(0, h(x)) + f_{X|Y=1}(x) \cdot p_1 \cdot \ell_{0-1}(1, h(x)) \end{aligned}$$

כעת התנאי

$$h(x) = \begin{cases} 0 & \frac{f_{X|Y=0}(x)}{f_{X|Y=1}(x)} > \frac{p_1}{(1-p_1)} \\ 1 & \text{otherwise} \end{cases}$$

נפרש את הביטוי : נשים לב מהיות $f_{X|Y=y}$ פונקציית צפיפות של משתנה גאוסייני היא חיובית (ולכן נוכל לחלק בה) וההסתברות $p_1 \in (0, 1)$ אחרת לא יכול להיות כי $X|Y=y$ משתנה גאוסייני.

$$\begin{aligned} f_{X|Y=0}(x) \cdot (1-p_1) > f_{X|Y=1}(x) \cdot p_1 &\iff \frac{f_{X|Y=0}(x)}{f_{X|Y=1}(x)} > \frac{p_1}{(1-p_1)} \\ &\iff \frac{\frac{1}{\sigma_0\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma_0^2}\right)}{\frac{1}{\sigma_1\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma_1^2}\right)} > \frac{p_1}{(1-p_1)} \\ &\iff \frac{\sigma_1}{\sigma_0} \exp\left(\frac{(x-\mu)^2}{2\sigma_1^2} - \frac{(x-\mu)^2}{2\sigma_0^2}\right) > \frac{p_1}{(1-p_1)} \\ &\iff \left(\frac{(x-\mu)^2}{2\sigma_1^2} - \frac{(x-\mu)^2}{2\sigma_0^2}\right) > \ln\left(\frac{p_1}{(1-p_1)} \cdot \frac{\sigma_0}{\sigma_1}\right) \\ &\iff \left(\frac{x^2 - 2x\mu + \mu^2}{2\sigma_1^2} - \frac{x^2 - 2x\mu + \mu^2}{2\sigma_0^2}\right) > \ln\left(\frac{p_1}{(1-p_1)} \cdot \frac{\sigma_0}{\sigma_1}\right) \\ &\iff \left(\frac{x^2 - 2x\mu + \mu^2}{2\sigma_1^2} - \frac{x^2 - 2x\mu + \mu^2}{2\sigma_0^2}\right) > \ln\left(\frac{p_1}{(1-p_1)} \cdot \frac{\sigma_0}{\sigma_1}\right) \\ &\iff \left(\frac{\sigma_0^2 x^2 - \sigma_0^2 2x\mu + \sigma_0^2 \mu^2 - \sigma_1^2 x^2 + \sigma_1^2 2x - \sigma_1^2 \mu^2}{\sigma_1^2 \sigma_0^2}\right) > 2 \ln\left(\frac{p_1}{(1-p_1)} \cdot \frac{\sigma_0}{\sigma_1}\right) \\ &\iff \left(\frac{x^2 (\sigma_0^2 - \sigma_1^2) + 2x\mu (\sigma_1^2 - \sigma_0^2) + \mu^2 (\sigma_0^2 - \sigma_1^2)}{\sigma_1^2 \sigma_0^2}\right) > 2 \ln\left(\frac{p_1}{(1-p_1)} \cdot \frac{\sigma_0}{\sigma_1}\right) \\ &\iff x^2 (\sigma_0^2 - \sigma_1^2) + 2x\mu (\sigma_1^2 - \sigma_0^2) + \mu^2 (\sigma_0^2 - \sigma_1^2) > \sigma_1^2 \sigma_0^2 2 \ln\left(\frac{p_1}{(1-p_1)} \cdot \frac{\sigma_0}{\sigma_1}\right) \\ &\quad : (\sigma_0^2 - \sigma_1^2) = A \text{ וכן } \sigma_1^2 \sigma_0^2 2 \ln\left(\frac{p_1}{(1-p_1)} \cdot \frac{\sigma_0}{\sigma_1}\right) = C \text{ נסמן} \end{aligned}$$

נפריד כעת לשני מקרים (כיוון ש $A \neq 0$ לפי הנתון) :

$$: A > 0 \bullet$$

$$\begin{aligned} &\iff x^2 A - A 2x\mu + \mu^2 A > C \\ &\iff x^2 - 2\mu x + \mu^2 > \frac{C}{A} \\ &\iff (x - \mu)^2 > \frac{C}{A} \end{aligned}$$

אם $C < 0$: מתקיים לכל x

אם $C \geq 0$:

$$\iff (x - \mu)^2 > \frac{C}{A} \iff |x - \mu| > \sqrt{\frac{C}{A}} \iff x > \mu + \sqrt{\frac{C}{A}} \text{ or } x < -\mu + \sqrt{\frac{C}{A}}$$

$$A < 0 \bullet$$

$$\iff x^2 A - 2x\mu A + \mu^2 A > C$$

$$\iff x^2 - 2\mu x + \mu^2 < \frac{C}{A}$$

$$\iff (x - \mu)^2 < \frac{C}{A}$$

$$\text{אם } C \leq 0 :$$

$$\iff (x - \mu)^2 < \frac{C}{A} \iff |x - \mu| < \sqrt{\frac{C}{A}} \iff -\mu + \sqrt{\frac{C}{A}} < x < \mu + \sqrt{\frac{C}{A}}$$

$$\text{אם } C > 0 : \text{אין } x \text{ המקיים את הנ"ל.}$$

נסכם

$$h(x) = \begin{cases} 0 & A > 0, C < 0 \\ 0 & A > 0, C \geq 0, \left(x > \mu + \sqrt{\frac{C}{A}}\right) \vee \left(x < -\mu + \sqrt{\frac{C}{A}}\right) \\ 0 & A < 0, C \leq 0, -\mu + \sqrt{\frac{C}{A}} < x < \mu + \sqrt{\frac{C}{A}} \\ 1 & \text{otherwise} \end{cases}$$

$$C = \sigma_1^2 \sigma_0^2 2 \ln \left(\frac{p_1}{(1-p_1)} \cdot \frac{\sigma_0}{\sigma_1} \right)$$

$$A = (\sigma_0^2 - \sigma_1^2)$$

Programing Assignment

1. Visualizing the Hoeffding bound

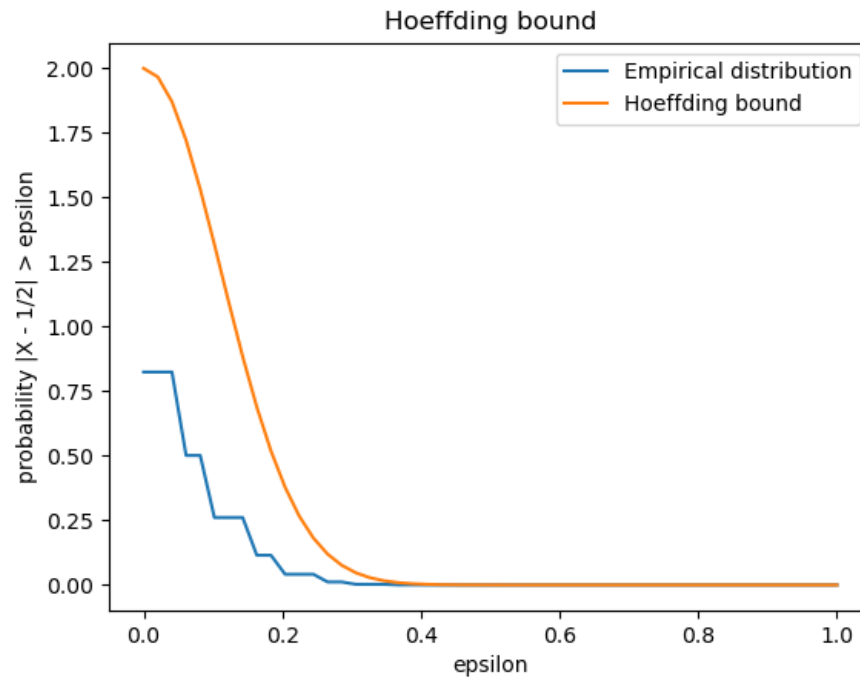


Figure 2: Visualizing the Hoeffding bound

2. Nearest Neighbor

2.1

Submitted to the programing section of the assignment.

2.2

אחוז הדיוק שהתקבל : 84.6%

ממסווג רנדומי לחלוטין היה מצופה שאחוז הדיוק יהיה (בתוחלת) 10% שכן יש עשרה מספרים שיכולים להתקבל בהסתברות שווה.

2.3

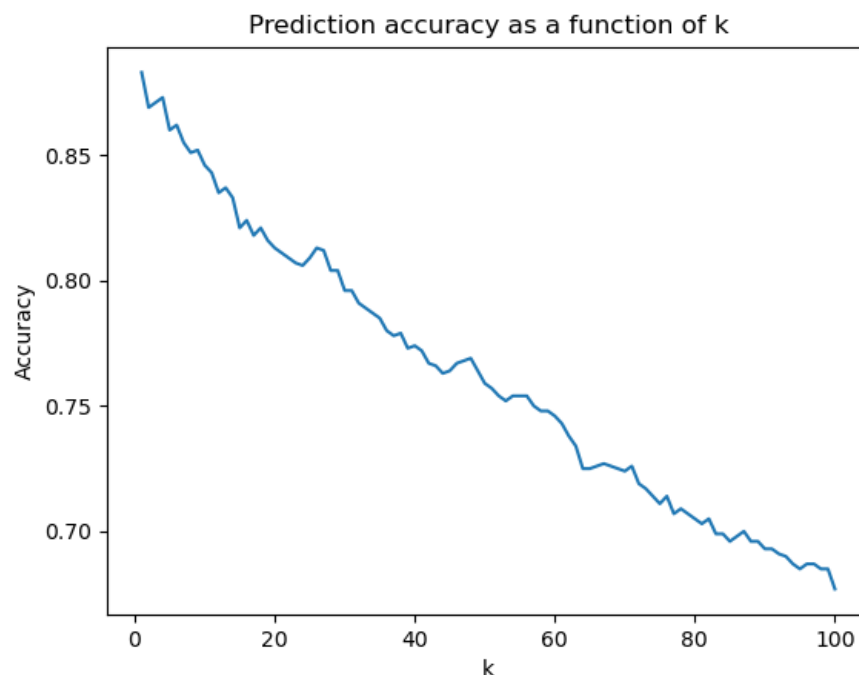
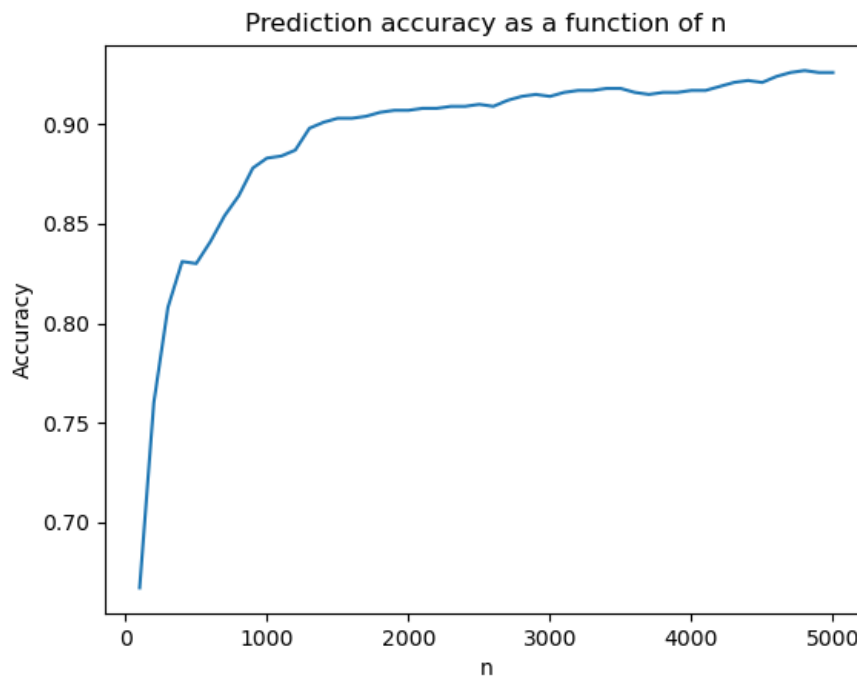


Figure 3: Prediction accuracy as a function of k

ניתן לראות כי ככל ש k גדל כך אחוז הדיוק קטן. ניתן להסביר זאת מפירוש ריצת האלגוריתם באופן הבא : ככל ש k גדל, "pool" התמונות שעומד לרשות המסווג גדול יותר לכן נצפה לראות יותר שגיאות. למשל ניתן הסבר שיכול לשכנע : תמונות של הספרה 1 ושל הספרה 7 יכולות להיות דומות, בהינתן תמונה של הספרה 7 שהאלגוריתם מקבל, הוא יחשב מרחקים מהתמונות הקרובות ביותר לפי המרחק האוקלידי. אם כן יש סיכוי לא מבוטל כי תמונות שמתייגות כ 1, קרובות לתמונה שהאלגוריתם קיבל ולכן נמצאות ב "pool" התמונות שעומד לרשות המסווג. אם התמונות התוייגות כ 1 ב "pool" הן השכיחות ביותר, האלגוריתם ישגה וייתייג את התמונה כתמונה של הספרה 1.

כבאופן מובהק ניתן לראות כי עבור $k = 1$ מתקבל הדיוק הטוב ביותר.

2.4

Figure 4: Prediction accuracy as a function of n

ניתן לראות כי ככל ש n גדל כך אחוז הדיוק עולה. ניתן להסביר שעבור n קטן, תמונה בעלת $Label = L_1$ הניתנת לאלגוריתם, תהיה קרובה לתמונה מתוייגת בעלת $Label = L_2$ כאשר $L_1 \neq L_2$ בסיכוי גבוה יותר (כיוון שמספר התמונות להשוות את המרחק קטן יחסית, והאלגוריתם קובע עבור התמונה היחידה, עד כדי מרחקים שווים, הקרובה ביותר אליו). ככל שיש יותר מידע כך סביר יותר שתמונה בעלת $Label = L_1$ הניתנת לאלגוריתם, תהיה קרובה ביותר לתמונה מתוייגת בעלת אותו L_1 .

ניתן לראות שהחל מ $n = 2000$ לערך, תוספת x תמונות מתוייגות אינה באופן מובהק כמו תוספת ל $100 < n < 1000$ ובאופן כללי ניתן לראות שהגרף מתנהג דומה לגרף \log . אפשר להסביר זאת שעבור מספיק תמונות מתוייגות (וכיוון שמספר התיוגים $\{0, 1, \dots, 9\}$ קטן יחסית והתמונות אינן מורכבות, יחסית) האלגוריתם יצליח לקבוע ברמת מובהקות טובה את התיוג המתאים.