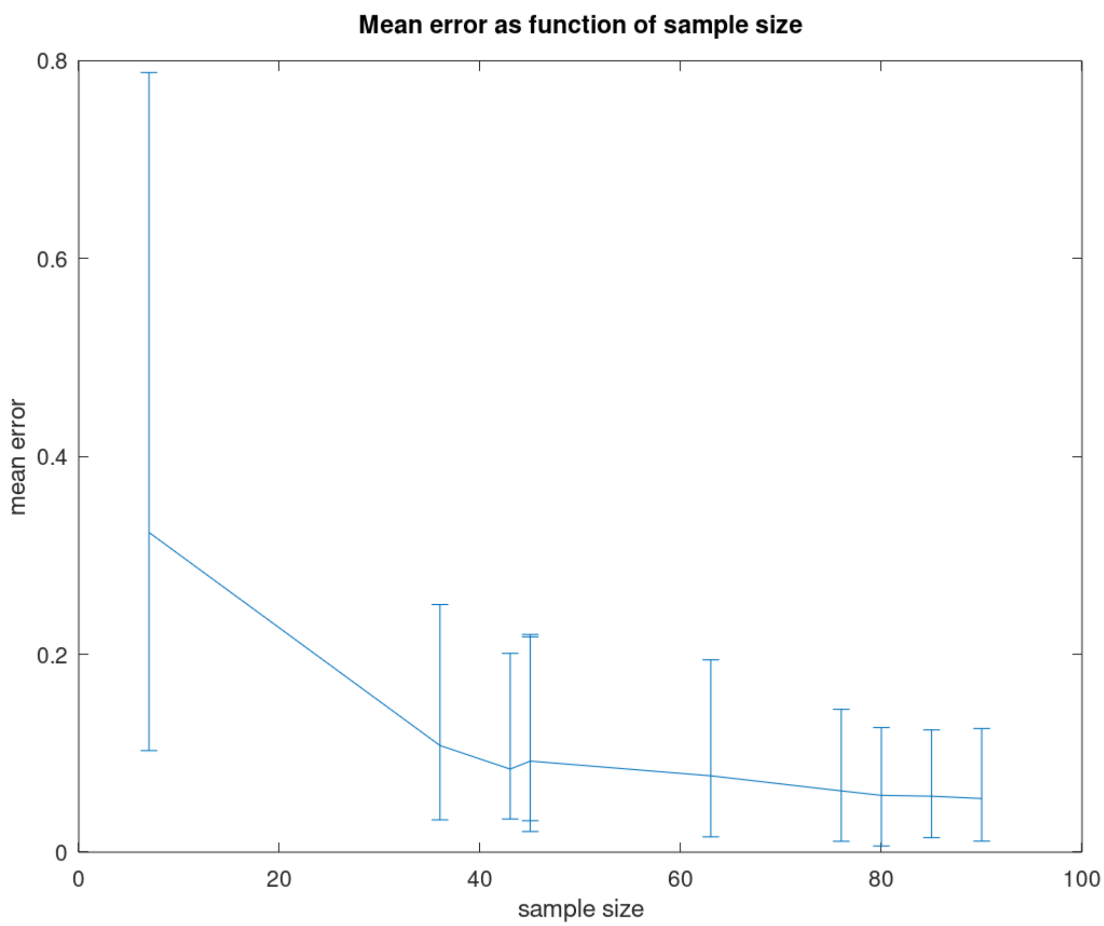
עבודה 1 – מבוא ללמידה וניתוח של מידע רב

מגישים: עומרי אטל 208625103, רעי וייס-ליפשיץ 208347039

שאלה 2

סעיף א



סעיף ב

השגיאה הממוצעת יורדת ככל שה sample size גדל.
הסבר: נשים לב תחילה שתמונות בעלות אותו לייבל יהיו דומות יחסית אחת לשנייה, כלומר המרחק האוקלידי שלהן יהיה קרוב ברוב המקרים. עבור k=1 אנו מריצים את למעשה את שיטת השכן הקרוב. ככל שהמדגם גדל, יש ל-predictor יותר דוגמאות מהלייבלים השונים, כך שלמעשה נוצרים גושים של נקודות בעלי אותו לייבל שגדלים יחד עם המדגם. מכאן, כאשר נפעיל את השכן הקרוב על כל דוגמה מה-test, יש סיכוי יותר גבוה שהדוגמה הזו "תיפול" בגוש הנכון (עם אותו הלייבל).

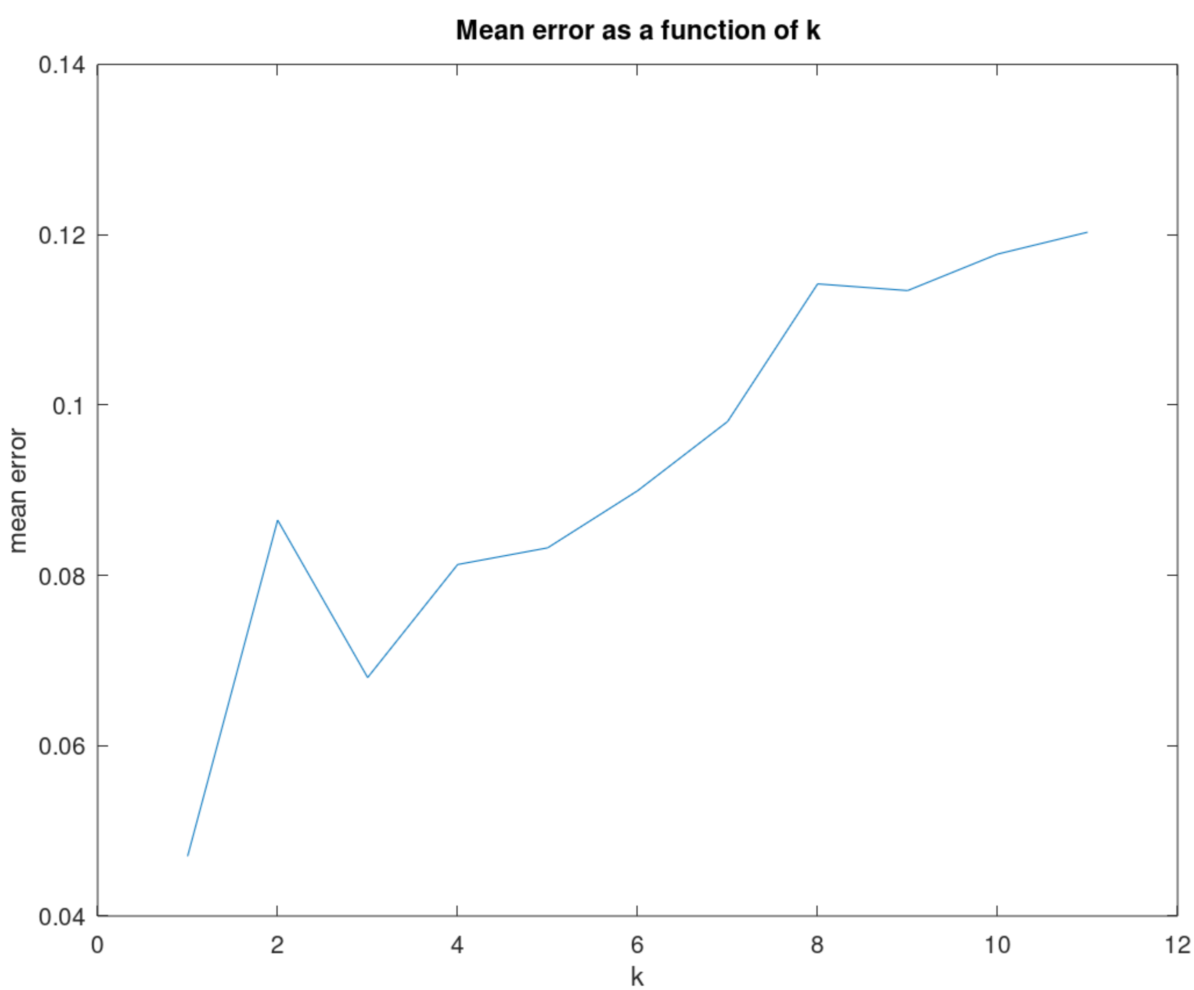
סעיף ג

כן, משום שב-gensmallm אנו יוצרים מדגם רנדומלי מה-train samples , כולכן בהרצות שונות יהיה לנו מדגמים שונים, מה שמשפיע על השגיאות.

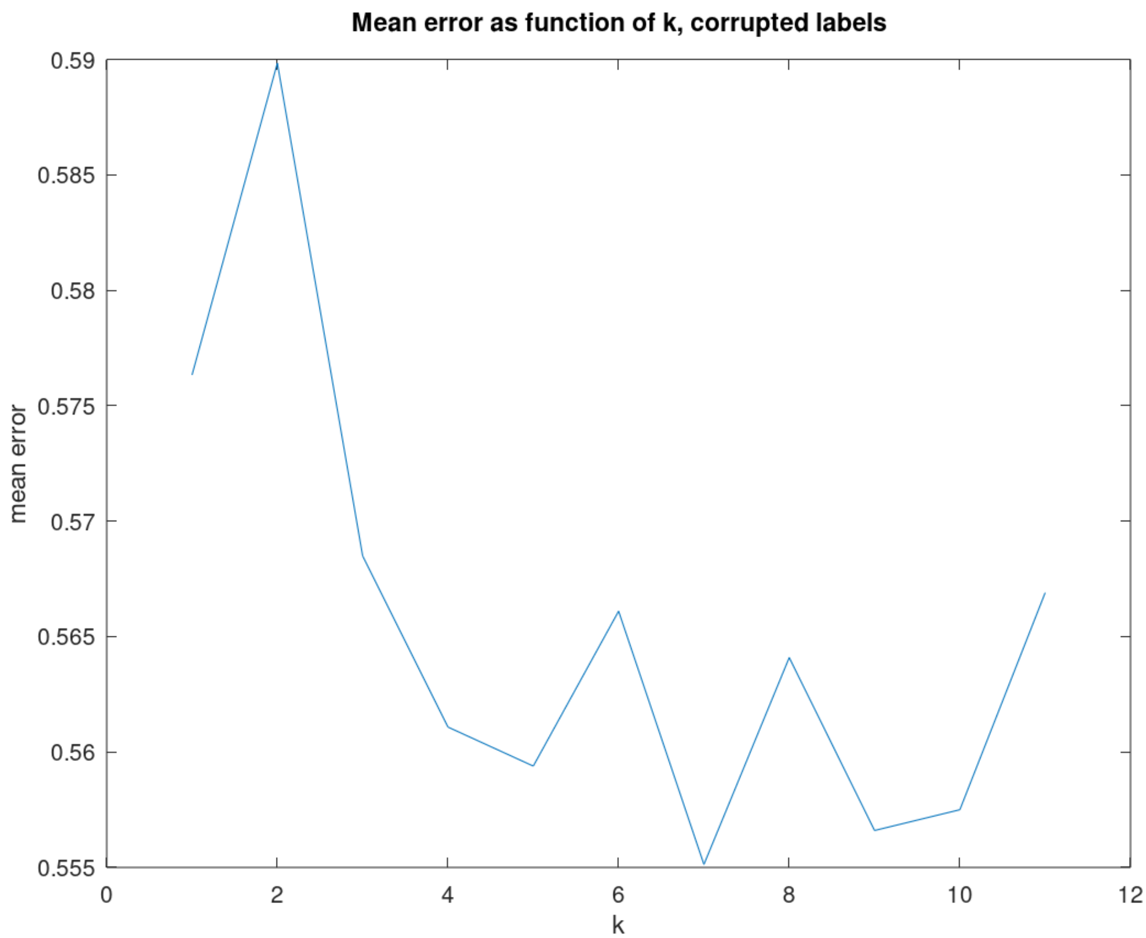
סעיף ד

טענו בסעיף ב שככל שהמדגם גדל כך לרוב השגיאות קטנות, ולכן בפרט זה נכון עבור השגיאה המקסימלית והמינימלית. בנוסף, ניתן לראות שהשגיאה המינימלית יחסית קבועה וחסומה על ידי 0, ומשום שהשגיאה הממוצעת יורדת, השגיאה המקסימלית חייבת לקטון.

סעיף ה



סעיף ו



סעיף ז

ערכים אופטימליים:

- ניסוי ראשון: $k = 1$
- ניסוי שני: $k = 7$

ההסבר לניסויים:

- בניסוי הראשון, אנו מגדילים את k בכל איטרציה, ככל שאנחנו מגדילים את k ככה יש יותר סיכוי שה- k-nn יתחשב בשכנים מאוד רחוקים ל-input שקיבלנו, ובפרט שכנים עם תווית שגויה.
- בניסוי השני, השגיאה הרבה יותר גבוהה, אבל נשים לב שככל ש-k עולה ככה השגיאה קטנה עד לנקודה מסוימת. מכיוון שה-training sample הושחת יש לו דוגמאות לא נכונות, אבל ככל שנגדיל את k נתחשב ביותר שכנים ולכן הסיכוי שניתקל בשכנים "אמיתיים" גדל ואז השגיאה תקטן עבור test samples לא מושחתים. מכיוון שקיימים test מושחתים אז טווח השגיאות גבוה בהרבה בהשוואה לניסוי הראשון, כי עבור test sample מושחת הסיכוי שהמדגם ייתן לו התווית הנכונה (שהיא למעשה שגויה) קלוש.

שאלה 3

סעיף א

$\mathcal{X} = \{(a,w)|a \in \mathbb{N}, w \in \mathbb{N}, a \leq 48,\ w \leq 4\}$

$\mathcal{Y} = \{lettuce, carrot\}$

סעיף ב

נגדיר בה"כ ש- lettuce = 1 ו- carrot = 0.

x	$h_{bayes}(x)$
(7,1)	1
(7,2)	1
(13,1)	0
(13,2)	1

$$error(h_{bayes}, \mathcal{D}) = \mathbb{P}_{(X,Y)\sim\mathcal{D}}[h_{bayes}(X) \neq Y] =$$
$$\sum_{x\in\mathcal{X}} \mathbb{P}[X = x](\mathbb{P}[Y \neq h_{bayes}(x)|X = x]) =$$
$$\sum_{x\in\mathcal{X}} \mathbb{P}[X = x](1 - \mathbb{P}[Y = h_{bayes}(x)|X = x]) =$$
$$\frac{1}{10}\cdot(1-1) + \frac{1}{2}\cdot(1-1) + \frac{3}{20}\cdot(1-1) + \frac{1}{4}\cdot(1-1) = 0$$

כלומר, $error(h_{bayes}, D) = 0$.

סעיף ג

age	preferred food	probability
7	carrot	0
7	lettuce	60%
13	carrot	15%
13	lettuce	25%

סעיף ד

גם פה נגדיר בה"כ ש- lettuce = 1 ו- carrot = 0.

x	$h'_{bayes}(x)$
7	1
13	1

$$error\left(h'_{bayes},\mathcal{D}'\right) = \mathbb{P}_{(X,Y)\sim\mathcal{D}'}\left[h'_{bayes}(X) \neq Y\right] =$$
$$\sum_{x\in X}\mathbb{P}[X=x]\left(\mathbb{P}\left[Y \neq h'_{bayes}(x)\middle|X=x\right]\right) =$$
$$\sum_{x\in X}\mathbb{P}[X=x]\left(1-\ \mathbb{P}\left[Y = h'_{bayes}(x)\middle|X=x\right]\right) =$$
$$\frac{3}{5}\cdot(1-1)+\frac{2}{5}\cdot\left(1-\frac{\mathcal{D}'(x,1)}{\mathcal{D}'(x)}\right) = \frac{2}{5}\cdot\left(1-\frac{0.25}{0.4}\right) = \frac{3}{20} = 0.15$$

כלומר, $error\left(h_{bayes},D\right) = \frac{3}{16}$

סעיף ה

למדנו בהרצאה ש-

$$\mathbb{E}_{S\sim D^m}\left[err\left(\widehat{h_S},\mathcal{D}\right)\right] = \frac{k-1}{k}\sum_{x\in\mathcal{X}}p_x(1-p_x)^m$$

כאשר $p_x = \mathbb{P}_{(X,Y)\sim\mathcal{D}}[X=x] = \sum_{y\in\mathcal{Y}}\mathcal{D}\left((x,y)\right)$

בשאלה זו, $k=2, m=2$. לכן:

$$\mathbb{E}_{S\sim D^2}\left[err\left(\widehat{h_S},\mathcal{D}\right)\right] = \frac{1}{2}\sum_{x\in\mathcal{X}}p_x(1-p_x)^2$$

עלינו לחשב את p_x לכל $x\in\mathcal{X}$.

$$x=(7,1)\Rightarrow p_x = \frac{1}{10}, x=(7,2)\Rightarrow p_x = \frac{1}{2}$$
$$x=(13,1)\Rightarrow p_x = \frac{3}{20}, x=(13,2)\Rightarrow p_x = \frac{1}{4}$$

ולכן סה"כ נקבל:

$$\frac{1}{2}\sum_{x\in\mathcal{X}} p_x(1-p_x)^2 = \frac{1}{2}\left[\frac{1}{10}\left(\frac{9}{10}\right)^2 + \frac{1}{2}\left(\frac{1}{2}\right)^2 + \frac{3}{20}\left(\frac{17}{20}\right)^2 + \frac{1}{4}\left(\frac{3}{4}\right)^2\right] = \frac{91}{400} = 0.2275$$

לסיכום, תוחלת שגיאת אלגוריתם ה $memorize$ כאשר $m = 2$ הוא:

$$\mathbb{E}_{S\sim D^2}\left[err\left(\widehat{h_s},\mathcal{D}\right)\right] = 0.2275$$

שאלה 4

סעיף א

נסמן $S = \big((a,y_1),..,(a,y_m)\big)$

$$\mathbb{P}\left[\widehat{h_s}(a) = 1\right] = \mathbb{P}\left[\widehat{h_s}(a) = 1 \middle|\ (a,1) \notin S\right]$$
$$+\mathbb{P}\left[\widehat{h_s}(a) = 1 | (a,1) \in S\right] = 0 + \beta_s = \beta_s \ \Rightarrow\ \mathbb{P}\left[\widehat{h_s}(a) = 1\right] = \beta_s$$

כך ש- $\beta_s = \frac{\sum_{i=1}^m \mathbb{I}[y_i=1]}{m}$, ולכן $\mathbb{P}\left[\widehat{h_s}(a) = 0\right] = 1 - \beta_s$

סעיף ב

נשים לב ש: $\mathcal{D}(a,1) = \psi, \mathcal{D}(a,0) = 1 - \psi$ לפי כלל בייס, ולכן לכל $(a,y) \in \mathcal{D}$ מתקיים ש-
$\mathbb{P}[y = 1] = \psi,\ P[y = 0] = 1 - \psi$

$$err\left(\hat{h}_s,\mathcal{D}\right) = \mathbb{P}_{(a,y)\in\mathcal{D}}\left[\hat{h}_s(a) \neq y\right] = \mathbb{P}\left[\ \hat{h}_s(a) \neq y \ \middle|\ y = 1\right]\ \cdot P[y = 1] +$$
$$\mathbb{P}\left[\ \hat{h}_s(a) \neq y \ \middle|\ y = 0\right] \cdot P[y = 0] = \ \mathbb{P}\left[\hat{h}_s(a) = 0\right] \cdot \psi +$$
$$\mathbb{P}\left[\hat{h}_s(a) = 1\right] \cdot (1-\psi) = (1-\beta_s)\psi + \beta_s(1-\psi) = \psi + \beta_s - 2\beta_s\psi$$

סעיף ג

נסמן $\gamma_s = m\beta_s$

$$\mathbb{E}_{S\sim D^m}[\beta_s] = \mathbb{E}_{S\sim D^m}\left[\frac{1}{m}\gamma_s\right] = \frac{1}{m}\mathbb{E}_{S\sim D^m}[\gamma_s]$$

נמצא את $\mathbb{E}_{S\sim D^m}[\gamma_s]$ כפונקציה של ψ. נשים לב ש- $\gamma_s = \sum_{i=1}^m \mathbb{I}[y_i = 1]$ כאשר $\mathbb{P}[y = 1] = \psi,$ $P[y = 0] = 1 - \psi$. כלומר γ_s הינו סכום של משתנים מקריים שמתפלגים ברנולית עם הסתברות ψ. לכן, $\gamma_s \sim B(m,\psi)$, וכידוע $\mathbb{E}[\gamma_s] = m\psi$. מכאן - $\mathbb{E}[\beta_s] = \frac{1}{m}\cdot m\psi = \psi$.

סעיף ד

$$\overline{err} = \mathbb{E}_{S\sim D^m}\left[err\left(\hat{h}_s,\mathcal{D}\right)\right] = \mathbb{E}_{S\sim D^m}[\psi + \beta_s - 2\beta_s\psi] = \psi + \psi - 2\psi^2 = 2\psi - 2\psi^2$$
$$= 2(\psi - \psi^2)$$

$$\overline{err} = \mathbb{E}_{S\sim D^m}\left[err\left(\hat{h}_s, \mathcal{D}\right)\right] = \mathbb{E}_{S\sim D^m}[\psi + \beta_s - 2\beta_s\psi] = \psi + \psi - 2\psi^2 = 2\psi - 2\psi^2$$
$$= 2(\psi - \psi^2) = 2\psi(1-\psi)$$

סעיף ה

תחילה נבחין ש-

$$h_{bayes}(a) = \begin{cases} 1\,, & \psi > \frac{1}{2} \\ 0, & \psi \leq \frac{1}{2} \end{cases}$$

$$err_{bayes} = err\left(h_{bayes}, \mathcal{D}\right) = \mathbb{P}_{(a,y)\in\mathcal{D}}\left[h_{bayes}(a) \neq y\right] = \mathbb{P}\left[h_{bayes}(a) = 0\right] \cdot \psi + \mathbb{P}\left[h_{bayes}(a) = 1\right] \cdot (1-\psi)$$

$$= \begin{cases} 0\cdot\psi + 1\cdot(1-\psi) = 1-\psi, & \psi > \frac{1}{2} \\ 1\cdot\psi + 0\cdot(1-\psi) = \psi, & \psi < \frac{1}{2} \end{cases} = \begin{cases} 1-\psi\,, & \psi > \frac{1}{2} \\ \psi, & \psi \leq \frac{1}{2} \end{cases}$$

סעיף ו

יהי $\varepsilon > 0$,

משום ש- $\frac{\overline{err}}{err_{bayes}} \geq 0$ אז מתקיים שלכל $2 \leq \varepsilon$ אכן $\frac{\overline{err}}{err_{bayes}} \geq \varepsilon - 2$.

אז נתבונן במקרה ש- $\varepsilon < 2$. נגדיר $\psi = \frac{\varepsilon}{4}$. נשים לב ש- $0 \leq \psi = \frac{\varepsilon}{4} < \frac{2}{4} = \frac{1}{2}$. לכן $err_{bayes} = \psi$.
אזי נקבל ש-

$$\frac{\overline{err}}{err_{bayes}} = \frac{2\psi(1-\psi)}{\psi} = 2(1-\psi) = 2 - 2\psi = 2 - 2\cdot\frac{\varepsilon}{4} = 2 - \frac{\varepsilon}{2} > 2 - \varepsilon$$

כנדרש.

שאלה 5

סעיף א

יהיו $(x_1, y_1), (x_2, y_2)$ כך ש- $(x_1, y_1), (x_2, y_2) \in S$ וגם $y_1 \neq y_2$. משום ש- $err\left(h_{bayes}, \mathcal{D}\right) = 0$ בהכרח מתקיים ש- $\mathcal{D}$ התפלגות דטרמיניסטית, כלומר שלכל $x \in \mathcal{X}$ קיים y יחיד בתמיכה של $\mathcal{D}$. לכן מתכונת דטרמיניסטיות של $\mathcal{D}$ מתקיים שלכל $x \in \mathcal{X}$, $\eta(x) = \mathbb{P}_{(X,Y)\sim\mathcal{D}}[Y = 1|X = x]$ מקיימת ש- $\eta(x)$ הוא 0 או 1. נשים לב שמשום ש- $y_1 \neq y_2$ מתקיים
$\eta(x_1) = \mathbb{P}_{(X,Y)\sim\mathcal{D}}[Y = 1|X = x_1] \neq \eta(x_2) = \mathbb{P}_{(X,Y)\sim\mathcal{D}}[Y = 1|X = x_2]$, ולכן

$|\eta(x_1) - \eta(x_2)| = 1$. נתון ש- η עבור $\mathcal{D}$ היא פונקציה $c - Lipschitz$ על פי המרחב האוקלידי ולכן בפרט עבור x_1, x_2 מתקיים $||x_1 - x_2|| \cdot c \geq |\eta(x_1) - \eta(x_2)|$, אז סה"כ לפי $|\eta(x_1) - \eta(x_2)| = 1$ נקבל $||x_1 - x_2|| \cdot c \geq 1 \Leftarrow \frac{1}{c} \geq ||x_1 - x_2||$.

<u>סעיף ב</u>

נסמן $f = f_S^{nn}$. צריך להוכיח ש- $err(f, \mathcal{D}) = \mathbb{P}_{(X,Y)\sim D}[h(X) \neq Y] = 0$, כלומר צריך להראות שלכל $(x, y) \in support(\mathcal{D})$, $f(x) = y$. יהיו $(x, y) \in \mathcal{X} \times \mathcal{Y}$ כך ש- $\mathcal{D}(x, y) > 0$.
נניח בשלילה ש- $f(x) = y' \neq y$, כלומר השכן הקרוב ביותר של x – במדגם הוא בעל תווית y', נסמנו x'. כלומר $f(x') = y'$.

<u>אבחנה</u>: המרחק $d = ||x - x'|| \leq \frac{2}{3c}$. לפי הנתון בו לכל $B \in \mathcal{B}$ קיים $(a, b) \in S$ עבורו $a \in B$, ובפרט עבור x', ולכן מכיוון ש- $\mathcal{B}$ מכסה את כל $\mathcal{X}$ נובע ש- $||x - x'|| \leq \frac{2}{3c}$, שכן קוטר הכדור הינו $\frac{2}{3c}$.

לפי האבחנה, $||x - x'|| \leq \frac{2}{3c}$, אבל $\frac{2}{3c} < \frac{1}{c}$ ולכן, $||x - x'|| \leq \frac{1}{c}$. לפי ההנחה, y' היא התווית של x ו- y' היא התווית של x', כך ש- $y' \neq y$. אך $||x - x'|| \leq \frac{1}{c}$ (*) , בסתירה להיות η פונקציה $c -$ $Lipschitz$ לפי סעיף א.

(*) הבהרה: הטענה שהוכחנו בסעיף א מתבססת על כל מדגם S, ולכן אם $(x, y) \notin S$ נוכל להוסיפו למדגם ובכך אילוץ המרחק על x, x' יתקיים.

<u>שאלה 6</u>

<u>סעיף א</u>

נטען ש $\inf_{h\in\mathcal{H}_2} err(\mathcal{H}_2, \mathcal{D}) \geq \inf_{h\in\mathcal{H}_4} err(\mathcal{H}_4, \mathcal{D}) \geq \inf_{h\in\mathcal{H}_6} err(\mathcal{H}_6, \mathcal{D})$, נוכיח זאת על ידי שנראה שלכל $f_{a_1a_2} \in \mathcal{H}_2$ קיימת $f_{a_1a_2} = f_{a_1a_2a_3a_4} \in \mathcal{H}_4$, ולכל $f_{a_1a_2a_3a_4} \in \mathcal{H}_4$ קיים $f_{a_1a_2a_3a_4} = f_{a_1a_2a_3a_4a_5a_6} \in \mathcal{H}_6$. ינבע ש- $\mathcal{H}_6$ תכיל לפחות את כל ההיפותזות של $\mathcal{H}_2, \mathcal{H}_4$ ולכן ינבע ששגיאת האפרוקסימציה לפי $\mathcal{H}_6$ היא הקטנה ביותר.

- תהא $f_{a_1a_2} \in \mathcal{H}_2$, אז נגדיר $f_{a_1a_2a_2a_2}$. נשים לב ש- $a_1 \leq a_2$ ולכן $f_{a_1a_2a_2a_2} \in \mathcal{H}_4$. לפי הגדרת $f_{a_1a_2}$, $f_{a_1a_2}(x) = \begin{cases} 1, & x \in [a_1, a_2] \\ 0, & x \notin [a_1, a_2] \end{cases}$, וגם משום ש-$a_1, a_2$ נמצא באינדקס אי-זוגי ב- $f_{a_1a_2a_2a_2}$ נובע ש-

$$f_{a_1a_2a_2a_2}(x) = \begin{cases} 1, & x \in [a_1, a_2] \; or \; x \in [a_2, a_2] \\ 0, & else \end{cases} = \begin{cases} 1, & x \in [a_1, a_2] \\ 0, & x \notin [a_1, a_2] \end{cases} = f_{a_1a_2}(x)$$

- תהא $f_{a_1a_2a_3a_4} \in \mathcal{H}_4$, אז נגדיר $f_{a_1a_2a_3a_4a_4a_4}$, נשים לב ש- $a_1 \leq a_2 \leq a_3 \leq a_4 \leq a_4 \leq a_4$, ולכן $f_{a_1a_2a_3a_4a_4a_4} \in \mathcal{H}_6$. משום ש- a_1, a_3, a_4 באינדקס אי-זוגי ב- $f_{a_1a_2a_3a_4a_4a_4}$ מתקיים:

$$f_{a_1a_2a_3a_4a_4a_4}(x) = \begin{cases} 1, & x \in [a_1, a_2]\ or\ x \in [a_3, a_4]\ or\ x \in [a_4, a_4] \\ 0, & else \end{cases}$$
$$= \begin{cases} 1, & x \in [a_1, a_2]\ or\ x \in [a_3, a_4] \\ 0, & else \end{cases} = f_{a_1a_2a_3a_4}(x)$$

<u>סעיף ב</u>

יהי $k \in \mathbb{N}$ זוגי כך ש- $k \geq 2$. יהי $m \leq k$ ומדגם S בגודל m.

נסמן ב -f^{nn} את ההיפותזה שחזרה מאלגוריתם *nearest neighbor* ביחס למדגם S. נראה תחילה ש- $f^{nn} \in \mathcal{H}_k$. נסמן $S = \left((x_1, y_1), \ldots, (x_m, y_m)\right)$. נראה שקיימת היפותזה $f_{a_1a_2\ldots a_k} \in \mathcal{H}_k$ כך ש- $f^{nn} = f_{a_1a_2\ldots a_k}$. נגדיר $I = \{i \in \mathbb{N} | i < m, y_i \neq y_{i+1}\}$.

המקרה הכללי:

נגדיר i_j להיות המספר ה-j בגודלו ב-I אז נגדיר $a'_j = \frac{x_{i_j}+x_{i_{j+1}}}{2}$.

<u>אבחנה</u>: $|I| < m \leq k$, $j = 1, \ldots, |I|$, וגם אם $|I| = m - 1$ אז בהכרח $y_1 \neq y_m$.

נגדיר את $a_1, \ldots, a_k$ לפי a'_j לכל $j = 1, \ldots, |I|$ באופן הבא:

אם $y_1 = 1$ אז נגדיר $a_1 = 0$ ולכל $i = 2, \ldots, |I| + 1$ נגדיר $a_i = a'_j$ כך ש- a'_j מסודרים בסדר ולה.

אם $y_1 = 0$ אז לכל $i = 1, \ldots, |I|$, נגדיר $a_i = a'_j$ כך ש- a'_j מסודרים בסדר עולה.

אם $y_m = 1$ אז נגדיר $a_k = 1$ ולכל $i = |I| + 1, \ldots, k - 1$ נגדיר $a_i = 1$ אם $y_1 = 0$, ואם $y_1 = 1$ אז לכל $i = |I| + 2, \ldots, k$.

אם $y_m = 0$ אז לכל $i = |I| + 1, \ldots, k$ נגדיר $a_i = a_{|I|}$ אם $y_1 = 0$, ואם $y_1 = 0$ אז לכל $i = |I| + 2, \ldots, k$.

נשים לב, שאם $|I| < m - 1$ אז a_k לא מוגדר a'_j, ולכן ניתן להגדיר אותו כרצוננו בלי לפגוע בערכים קיימים. אבל, אם $|I| = m - 1$ וגם $y_1 = 1$ אז בהכרח $y_m = 0$ לפי האבחנה, ולכן במקרה זה לא נוסיף את a_k ו- f תישאר ב- $\mathcal{H}_k$.

<u>אבחנה</u>: לכל $i = 1, \ldots, k - 1$ מתקיים ש- i אי-זוגי אם ורק אם לכל $x \in [a_i, a_{i+1}]$, $f_{a_1a_2\ldots a_k}(x) = 1$.

כעת נותר להראות ש- $f_{a_1a_2\ldots a_k} = f^{nn}$. יהי $x \in \mathcal{X}$. נראה ש- $f_{a_1a_2\ldots a_k}(x) = f^{nn}(x)$. נסמן ב- x_i את השכן הקרוב ביותר של x במדגם S. נניח בה"כ ש- $f^{nn}(x) = 1$, אז $y_i = 1$.

נניח ש- x_i מסודרים בסדר גודל עולה, ולהם סדרת לייבלים $(y_1, \ldots, y_i, \ldots, y_m)$. $y_i = 1$ והוא נמצא בתת-סדרה שכל ערכי y שלה הם 1. נתבונן ב- y_l וב- y_h, שהן התווית הראשונה והאחרונה בתת הסדרה הזו בהתאמה.

אם $l = 1$ אז בהכרח $a_1 = 0$. לכן $x \geq a_1$. נחלק לתתי מקרים:

- אם $h = m$ אז $|I| = 0$ ולכן $a_2 = a_3 = \cdots a_k = 1$ ולכן $x \leq a_2$ ו- $f_{a_1a_2\ldots a_k}(x) = 1$.

- אם $h < m$ אז $y_{h+1} = 0$ ולכן $h \in I$. לכן $a_2 = \frac{x_h + x_{h+1}}{2}$. אם $x > a_2$ אז x קרוב יותר ל- x_{h+1} מאשר x_h ובפרט x_i לא השכן הכי קרוב שלו, בסתירה להנחה. לכן $x \leq a_2$, ואז יתקיים $f_{a_1 a_2 \dots a_k}(x) = 1$.

אם $l > 1$ אז בהכרח $y_{l-1} = 0$, אזי $l - 1 \in I$, ולפי אופן הגדרת $f_{a_1 a_2 \dots a_k}$ עבור $a_j = \frac{(x_{l-1} + x_l)}{2}$ מתקיים ש-j אי-זוגי. בנוסף, נטען ש- $x \geq a_j$, באופן דומה למקרה שבו $l = 1$ ו- $h < m$ (שכן אחרת x יותר קרוב ל- x_{l-1} מאשר ל - x_l, ובפרט לא קרוב יותר ל- x_i). נחלק לתתי מקרים:

- $h = m$, אז לפי הגדרת $f_{a_1 a_2 \dots a_k}$ מתקיים ש- $a_{j+1} = \cdots = a_k = 1$ ומשום ש- a_j באינדקס אי זוגי וכי $x \geq a_j$ וגם $x \leq a_{j+1}$, אז $f_{a_1 a_2 \dots a_k}(x) = 1$.
- $h < m$, זהו מקרה זהה ל- $l = 1$ ו- $h < m$, ובפרט $x < a_{j+1}$, ולכן $f_{a_1 a_2 \dots a_k}(x) = 1$.

נותר להראות ש- $err(f^{nn}, S) = 0$, ואז f^{nn} יקיים את תנאי ה *ERM* עבור מחלקת ההיפותזות $\mathcal{H}_k$.
למדנו בהרצאה f^{nn} בעל שגיאה 0 על המדגם כאשר לכל תווית x_i במדגם S קיימת תווית יחידה y_i נטען שהדבר מתקיים עבור $S \sim \mathcal{D}^m$ כבשאלה בהסתברות 1. יהי $i \leq m$, נראה שלכל $j < i$ מתקיים שההסתברות שבמיקום ה-i נבחר x_j שכבר היה במדגם היא 0. נסמן ב- X_i מ"מ שמייצג את הבחירה ה-i במדגם.

$$\mathbb{P}\left[\forall j < i \,, X_i \neq x_j\right] = 1 - \mathbb{P}\left[\exists\, j < i, X_i = x_j\right] \geq 1 - \sum_{j=1}^{i-1} \mathbb{P}\left[X_i = x_j\right] = 1 - 0 = 1$$

לכן ההסתברות שנראה שתי דוגמאות זהות היא 0 ולכן לפי מה שנלמד בהרצאה, היפותזת ה *nearest neighbor* תיתן שגיאה 0 על המדגם ותקיים את תנאי ה *ERM* עבור $\mathcal{H}_k$.

<u>סעיף ג</u>

יהי $k \geq 2$ זוגי. אנחנו נראה שקיים מדגם בגודל $m = k + 1$, כך f^{nn} לא מקיימת את תנאי ה *ERM* עבור $\mathcal{H}_k$. נראה ש- $\inf_{h \in \mathcal{H}_k} err(h, S) > 0 = err(f^{nn}, S)$, ואז f^{nn} לא מקיים את הדרוש.
נגדיר: $S = \left((x_1, 1), (x_2, 0), (x_3, 1), \dots, (x_k, 0), (x_{k+1}, 1)\right)$, כך שלכל $i = 1, \dots, k$, $x_{i+1} > x_i$. נניח בשלילה שקיימת היפותזה $f_{a_1 a_2 \dots a_k} \in \mathcal{H}_k$ כך ש- $err\left(f_{a_1 a_2 \dots a_k}, S\right) = 0$. נראה שלכל x_i שתוויתו 1 במדגם S, קיים טווח אחד ויחיד $\left[a_j, a_{j+1}\right]$ שיכיל רק את $x_i \in S$ עבור j אי-זוגי.
נבחין שקיימים במדגם S $\frac{k}{2} + 1$ דוגמאות שתוויתם 1, וב-f קיימים $\frac{k}{2}$ טווחים שמכילים x יחיד. מעיקרון שובך היונים, קיים טווח שיכיל 2 איברים, בסתירה.

יהי x_i שתוויתו 1 במדגם S, לכן $f_{a_1 a_2 \dots a_k}(x) = 1$ כי היפותזה זו לא טועה על המדגם. אז קיים טווח $\left[a_j, a_{j+1}\right]$ כך ש- $x_i \in \left[a_j, a_{j+1}\right]$. נניח בשלילה שקיים $x_l > x_i$ בה"כ כך ש- $x_l \in \left[a_j, a_{j+1}\right]$. אם $y_l = 0$, הדבר לא אפשרי שכן אז $f_{a_1 a_2 \dots a_k}(x_l) = 1$ בסתירה לכך שהתווית שלו 0 במדגם. אם $y_l = 1$ אז לפי הגדרת S קיים x_r כך ש- $y_r = 0$ וגם $x_i < x_r < x_l$ ולכן גם $x_r \in \left[a_j, a_{j+1}\right]$, ואז $f_{a_1 a_2 \dots a_k}(x_r) = 1$ בסתירה לכך שהתווית שלו היא 0.

הוכחנו שלא קיימת $h \in \mathcal{H}_k$ כך ששגיאתה על המדגם היא 0. בנוסף, לפי סוף סעיף ב, גם עבור S בגודל זה, משום שכל $x_i \neq x_j$ $err(f^{nn}, S) = 0$, ולכן בהכרח $f^{nn} \notin \mathcal{H}_k$.