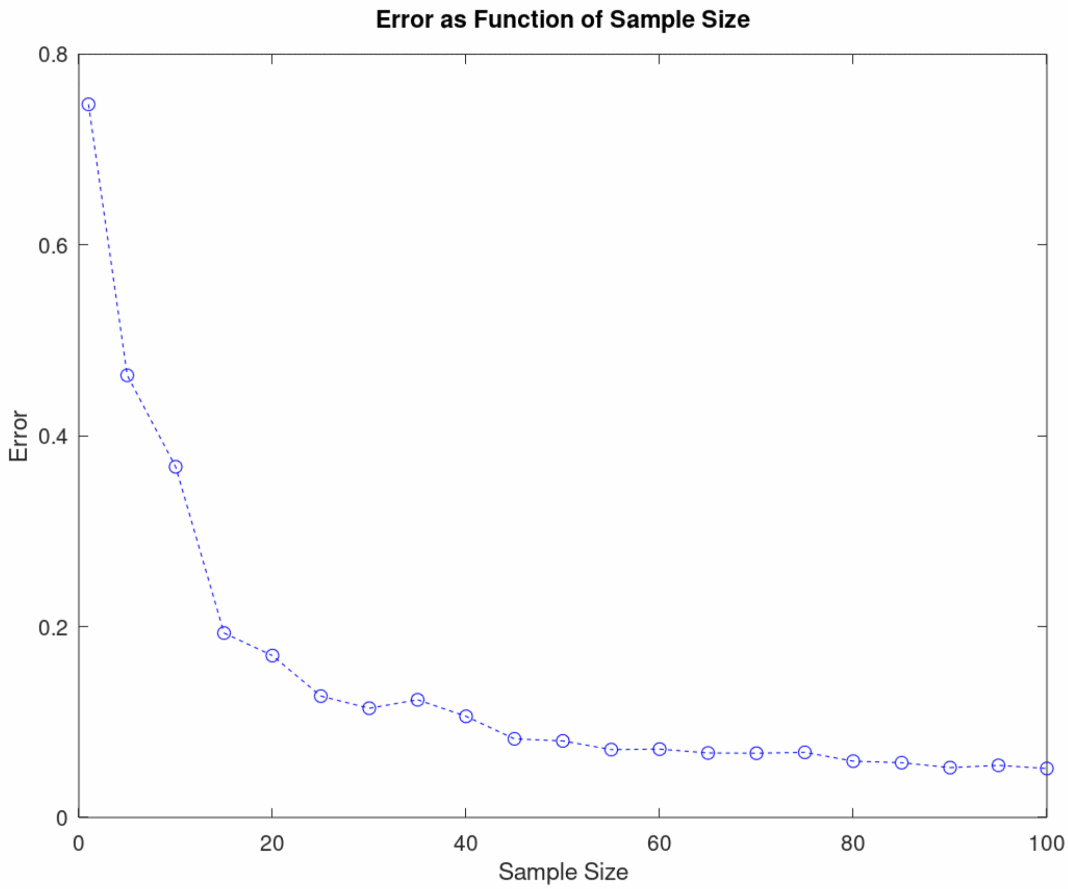
.2

.a



.b

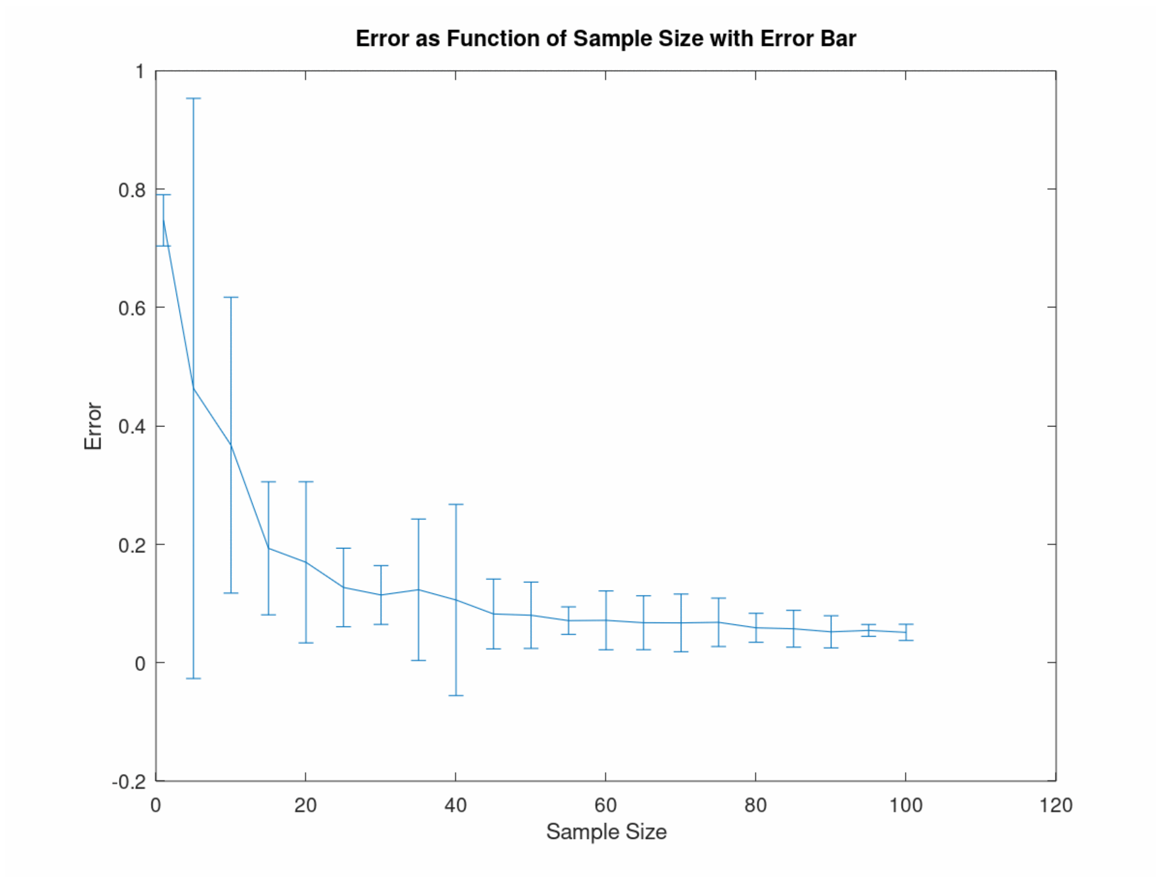
ניתן להבחין כי קיימת מגמה בערכי ממוצע השגיאה בגרף - ככל שגודל המדגם גדל כך ממוצע השגיאה קטן. ניתן להסביר זאת על ידי כך שככל שגודל המדגם גדול יותר, כך הוא מייצג באופן מהימן יותר את ההתפלגות, ולכן האלגוריתם יכול לחזות באופן מדויק יותר את התווית המתאימה של הדוגמאות.

.c

כפי שניתן לראות בטבלאת התוצאות המצורפת, אכן התקבלו תוצאות שונות עבור ריצות שונות עם אותו גודל מדגם. ניתן להסביר זאת על ידי כך שבכל מדגם עם אותו הגודל יכלו להתקבל דוגמאות שונות, הן מבחינת ערכי הדוגמאות והן מבחינת התוויות שלהן, ולכן האלגוריתם לומד בדרך שונה.

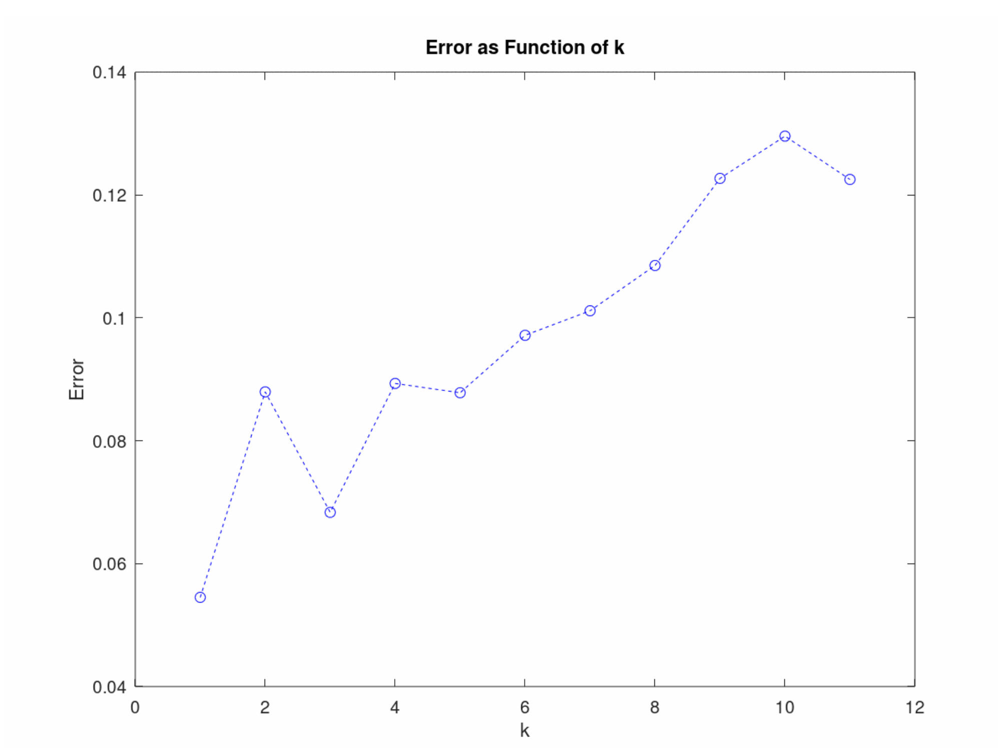
Sample Size\Iteration	1	2	3	4	5	6	7	8	9	10
1	0.75275398	0.75275398	0.75275398	0.75960832	0.72215422	0.75960832	0.72215422	0.72215422	0.76548348	0.76548348
5	0.23892289	0.36695226	0.57943696	0.54320685	0.38212974	0.45336597	0.72900857	0.62080783	0.47099143	0.2494492
10	0.19804162	0.3620563	0.44798042	0.37282742	0.4247246	0.37086903	0.374541	0.38506732	0.40685435	0.33292534
15	0.1130967	0.21566707	0.20636475	0.19706242	0.21664627	0.2244798	0.16866585	0.21542228	0.225459	0.1505508
20	0.25214198	0.17772338	0.15593635	0.14516524	0.15177479	0.19045288	0.20342717	0.17698898	0.12729498	0.11603427
25	0.11799266	0.09816401	0.16107711	0.14296206	0.1623011	0.10232558	0.13243574	0.09596083	0.15079559	0.10795594
30	0.12239902	0.12974296	0.13855569	0.11064871	0.12215422	0.11897185	0.10330477	0.1128519	0.08886169	0.09742962
35	0.08274174	0.10110159	0.13684211	0.100612	0.09008568	0.11970624	0.19436965	0.20220318	0.09791922	0.10795594
40	0.11113831	0.1130967	0.07980416	0.09914321	0.08812729	0.07980416	0.22741738	0.11456548	0.08151775	0.06585067
45	0.06829865	0.08078335	0.09130967	0.08641371	0.08151775	0.0626683	0.07955936	0.07564259	0.12166463	0.07637699
50	0.09449204	0.06413709	0.08053856	0.07735618	0.11627907	0.07319461	0.07931457	0.09155447	0.06536108	0.06022032
55	0.07099143	0.06536108	0.07833537	0.08225214	0.0624235	0.06976744	0.05899633	0.06315789	0.08078335	0.07955936
60	0.06756426	0.09522644	0.09840881	0.06389229	0.05581395	0.04871481	0.06829865	0.07392901	0.05801714	0.08665851
65	0.08445532	0.06168911	0.05189718	0.05801714	0.06022032	0.07686659	0.06805386	0.05091799	0.06756426	0.09645043
70	0.05997552	0.06609547	0.08984088	0.04455324	0.06658507	0.0624235	0.09253366	0.0746634	0.04381885	0.07294982
75	0.06340269	0.06487148	0.0621787	0.06585067	0.0626683	0.07809058	0.07588739	0.0619339	0.05361077	0.09449204
80	0.04700122	0.0621787	0.06829865	0.0499388	0.05826193	0.06854345	0.07148103	0.05067319	0.06609547	0.04798042
85	0.05116279	0.05973072	0.07270502	0.0499388	0.0626683	0.06756426	0.04161567	0.06487148	0.0494492	0.05458996
90	0.05826193	0.04895961	0.05312118	0.04504284	0.04308446	0.0501836	0.05507956	0.05214198	0.07025704	0.04651163
95	0.05899633	0.05091799	0.04895961	0.05875153	0.05679315	0.05189718	0.05116279	0.05850673	0.05410037	0.05630355
100	0.05434517	0.05507956	0.05116279	0.05801714	0.04773562	0.05556916	0.05091799	0.04847001	0.04773562	0.04430845

.d

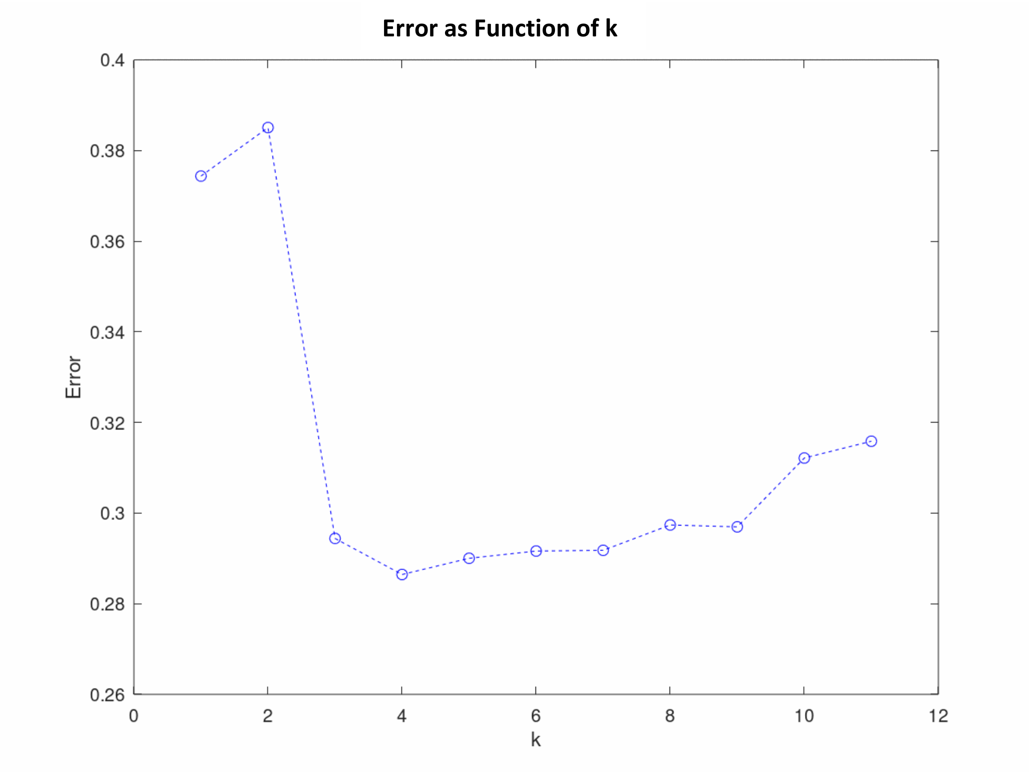


ניתן להבחין כי עבור מדגם בגודל אחד ערך הסוגר-טעות הינו נמוך מכיוון שזהו מקרה מנוון בו קיימת דוגמא אחת שהינה השכנה הקרובה ביותר עבור כל בדיקה, ולכן הערך הנ"ל מייצג את היחס בין התווית השכיחה ביותר שנבחרה, לבין התווית הנדירה ביותר שנבחרה מקבוצת הבדיקות. לאחר מכן, ניתן להבחין כי קיימת מגמה בערכי הסוגר-טעות – ככל שגודל המדגם גדול יותר, כך הערך הנ"ל קטן יותר. ניתן להסביר זאת על ידי כך שככל שגדלי המדגמים גדולים יותר, כך הם מייצגים באופן מהימן יותר את ההתפלגות, ומכאן השונות ביניהם קטנה יותר.

.e



.f



.g

הערך האופטימלי עבור הניסוי בסעיף e הינו k=1, בעוד שהערך האופטימלי עבור הניסוי בסעיף f הינו k=4. ניתן להבחין כי קיים הבדל בין שני הניסויים, אותו אפשר להסביר ע"י כך שהשחתה של חלק קטן מהדוגמאות והבדיקות מעלה צורך לכפר עליה בעזרת השוואה של כל בדיקה למספר רב יותר של דוגמאות.

3. a. $\chi = \{(x_1, x_2) \mid x_1 \in \{0,1,2,\dots,48\}, x_2 \in \{0,1,2,3,4\}\}$

$Y = \{0,1\}, (0 = carrot, 1 = lettuce)$

b. $h_{bayes}(x) = \begin{cases} 1\ ,\ x = (7,1) \\ 1\ ,\ x = (7,2) \\ 0\ ,\ x = (13,1) \\ 1\ ,\ x = (13,2) \\ randomaly,\ else \end{cases}$

$$err(h_{bayes}, D) = P_{(\chi,Y)\sim D}\left[h_{bayes}(\chi) \neq Y\right] = \sum_{(x,y)\in\chi\times Y: h_{bayes}(x)\neq y} P[\chi = x, Y = y] = \mathbf{0}$$

c.

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

d. $h'_{bayes}(x) = \begin{cases} 1\ ,\ x = 7 \\ 1\ ,\ x = 13 \\ randomaly,\ else \end{cases}$

$$err(h'_{bayes}, D') = P_{(\chi,Y)\sim D'}\left[h'_{bayes}(\chi) \neq Y\right] = \sum_{(x,y)\in\chi\times Y: h'_{bayes}(x)\neq y} P[\chi = x, Y = y] = \mathbf{0.15}$$

e. $E_{S\sim D^m}\left[err(\hat{h}_S, D)\right] = E_{S\sim D^m}\left[M_S * \frac{k-1}{k}\right] = E_{S\sim D^m}\left[M_S * \frac{1}{2}\right] = \frac{1}{2} E_{S\sim D^m}\left[\sum_{x\in\chi\setminus\chi_S} p_x\right] = \cdots = \frac{1}{2}\sum_{x\in\chi} p_x (1-p_x)^m$

כאשר המעבר הראשון נכון עבור התפלגות דטרמיניסטית, ושאר המעברים כפי שנראו בכיתה. עבור m=2 נקבל:

$$E_{S\sim D^2}\left[err(\hat{h}_S, D)\right] = \frac{1}{2} * [0.1 * 0.9^2 + 0.5 * 0.5^2 + 0.15 * 0.85^2 + 0.25 * 0.75^2] = \mathbf{0.2275}$$

4. a.

$$\left(P\left[\hat{h}_S(a)=1\right]\right)=\beta_S=\frac{\sum_{i=1}^{m} y_i}{m},m\neq 0$$

ניתן להסביר את החישוב הנ"ל ע"י כך שההסתברות שכלל החיזוי ייתן תווית '1' עבור 'a' תלויה ביחס בין מספר הדוגמאות שהתווית שלהן '1' $(\sum_{i=1}^{m} y_i)$ למספר הדוגמאות הכולל, מכיוון שכל הדוגמאות במדגם נמצאות באותה הנקודה, אזי ההסתברות שהאלגוריתם יחזה '1' הינה היחס הנ"ל ע"פ הגדרתו.

b.

ראשית נבחין כי $\psi=\eta(a)=P_{(\chi,Y)\sim D}[\,Y=1|\chi=a]=\frac{P_{(\chi,Y)\sim D}[Y=1\,,\chi=a]}{P_{(\chi,Y)\sim D}[\chi=a]}=P_{(\chi,Y)\sim D}[\chi=a,\ Y=1]=P[Y=1]$

$$\begin{aligned}\boldsymbol{err}\left(\hat{\boldsymbol{h}}_{\boldsymbol{S}},\boldsymbol{D}\right)&=P_{(\chi,Y)\sim D}\left[\hat{h}_S(\chi)\neq Y\right]=P\left[Y=1,\hat{h}_S(a)=0\right]+P\left[Y=0,\hat{h}_S(a)=1\right]\\&=P[Y=1]*P\left[\hat{h}_S(a)=0\right]+P[Y=0]*P\left[\hat{h}_S(a)=1\right]\\&=\psi*\left(1-P\left[\hat{h}_S(a)=1\right]\right)+(1-\psi)*P\left[\hat{h}_S(a)=1\right]=\psi*(1-\beta_S)+(1-\psi)*\beta_S\\&=\boldsymbol{\psi+\beta_S-2\psi\beta_S}\end{aligned}$$

c.

$$\begin{aligned}\boldsymbol{E_{S\sim D^m}[\beta_S]}&=\sum_{i=0}^{m}\frac{i}{m}*P\left[\beta_S=\frac{i}{m}\right]=\sum_{i=0}^{m}\frac{i}{m}*\binom{m}{i}\psi^i(1-\psi)^{m-i}=\sum_{i=1}^{m}\binom{m-1}{i-1}\psi^i(1-\psi)^{m-i}\\&=\sum_{k=0}^{n}\binom{n}{k}\psi^{k+1}(1-\psi)^{n-k}=\psi\sum_{k=0}^{n}\binom{n}{k}\psi^k(1-\psi)^{n-k}=\psi\big(\psi+(1-\psi)\big)^n=\boldsymbol{\psi}\end{aligned}$$

d.

$$\boldsymbol{E_{S\sim D^m}}\left[\boldsymbol{err}\left(\hat{\boldsymbol{h}}_{\boldsymbol{S}},\boldsymbol{D}\right)\right]=E[\psi+\beta_S-2\psi\beta_S]=\psi+E[\beta_S]-2\psi E[\beta_S]=\psi+\psi-2\psi^2=\boldsymbol{2\psi-2\psi^2}$$

e.

$$h_{bayes}(a)=\begin{cases}1,\ \psi\geq\frac{1}{2}\\0,\ \psi<\frac{1}{2}\end{cases}\ ,\quad \boldsymbol{err_{bayes}=\min\{\psi,(1-\ \psi)\}}$$

f. יהי $\varepsilon>0$.

נחלק למקרים:

i. $\psi\leq\frac{1}{2}$: במקרה זה $err_{bayes}=\min\{\psi,(1-\psi)\}=\psi$. לכן $\frac{\overline{err}}{err_{bayes}}=\frac{2\psi-2\psi^2}{\psi}=2-2\psi$, וצריך להראות כי קיים ψ עבורו $2-2\psi>2-\varepsilon$, משמע $\psi<\frac{\varepsilon}{2}$. ניתן לראות כי קיימים אינסוף ערכים $\boldsymbol{0<\psi<\frac{\varepsilon}{2}}$ המקיימים אי שוויון זה.

ii. $\psi>\frac{1}{2}$: במקרה זה $err_{bayes}=\min\{\psi,(1-\psi)\}=1-\psi$. לכן $\frac{\overline{err}}{err_{bayes}}=\frac{2\psi-2\psi^2}{1-\psi}=2\psi$, וצריך להראות כי קיים ψ עבורו $2\psi>2-\varepsilon$, משמע $\psi>1-\frac{\varepsilon}{2}$. ניתן לראות כי קיימים אינסוף ערכים $\boldsymbol{1>\psi>1-\frac{\varepsilon}{2}}$ המקיימים אי שוויון זה.

5. a. יהיו $(x_1,y_1),(x_2,y_2)\in S$ כך ש $y_1\neq y_2$.

η של D הוא $c-Lipschitz$ ביחס למרחק אוקלידי לכן $||x_1-x_2||\leq\frac{|\eta(x_1)-\eta(x_2)|}{c}$.

מכאן, על מנת להוכיח את הנדרש מספיק להראות כי $|\eta(x_1)-\eta(x_2)|=1$.

בה"כ נניח כי $y_1=0,y_2=1$.

D בעל $Bayes-error$ 0 ולכן דטרמיניסטי.

D דטרמיניסטי ו $(x_1,0)\in S$ ומכאן ש $\eta(x_1)=0$.

D דטרמיניסטי ו $(x_2,1)\in S$ ומכאן ש $\eta(x_2)=1$.

מכאן ניתן לראות כי $|\eta(x_1)-\eta(x_2)|=1$ כנדרש.

b. נניח בשלילה שקיימים $(x,y)\sim D$ כך ש $f_S^{nn}(x)\neq y$. נסמן ב x' את הדוגמא הקרובה ביותר ל x במדגם שעל פיה קבע האלגוריתם את $f_S^{nn}(x)=y'$. מכיוון ש $y\neq y'$ ע"פ סעיף א' נסיק כי $\left||x-x'|\right|\geq\frac{1}{c}$. יהי $B\in\mathcal{B}$ כך ש $x\in B$. ע"פ הנתון קיימת דוגמא (x'',y'') במדגם כך ש $x''\in B$. הרדיוס של B הינו $\frac{1}{3c}$ ומכאן ש $\left||x-x''|\right|\leq\frac{2}{3c}$.

מכאן ניתן להסיק כי $||x-x''||<\left||x-x'|\right|$, ומכך ש (x'',y'') שייכת למדגם נקבל סתירה לנכונות האלגוריתם $1-nn$ שבחר את הדוגמה הקרובה ביותר במדגם לנקודה x.

6. a. תהי $f_{a_1,a_2} \epsilon H_2$ פונקציה המקיימת $err_{H_2-app} = \inf_{h \in H_2} err(h,D) = err(f_{a_1,a_2},D)$.

נסתכל על $f_{a_1,a_2,a_2,a_2} \epsilon H_4$, ונראה כי לכל $x \in \chi$ מתקיים $f_{a_1,a_2}(x) = f_{a_1,a_2,a_2,a_2}(x)$.

יהי $x \in \chi$, נחלק למקרים:

i. $a_1 \le x \le a_2$: אזי לפי הגדרה $f_{a_1,a_2}(x) = f_{a_1,a_2,a_2,a_2}(x) = 1$.
ii. $x < a_1$ או $a_2 < x$: אזי לפי הגדרה $f_{a_1,a_2}(x) = f_{a_1,a_2,a_2,a_2}(x) = 0$.

הראינו כי לכל $x \in \chi$ מתקיים $f_{a_1,a_2}(x) = f_{a_1,a_2,a_2,a_2}(x)$, מכאן ש $err(f_{a_1,a_2},D) = err(f_{a_1,a_2,a_2,a_2},D)$.

לכן, $err_{H_2-app} = err(f_{a_1,a_2},D) = err(f_{a_1,a_2,a_2,a_2},D) \ge err_{H_4-app}$, אזי $err_{H_2-app} \ge err_{H_4-app}$.

באופן דומה ניתן להראות כי $err_{H_4-app} \ge err_{H_6-app}$, ובסה"כ הראינו כי $\boldsymbol{err_{H_2-app} \ge err_{H_4-app} \ge err_{H_6-app}}$.

b. נסמן ב h את ההיפותזה שהוחזרה ע"י אלגוריתם ה $nn-1$. היפותזה זו מחזירה עבור כל דוגמא מהמדגם את הערך של השכן הכי קרוב אליה, ובהנחה כי לא קיימות במדגם שתי דוגמאות באותה הנקודה, ההיפותזה תחזיר עבורה את הערך שלה מהמדגם. מכאן, $err(h,S) = 0$ שזוהי השגיאה המינימאלית האפשרית, וכל שנותר להראות כי קיימת היפותזה ב H_k השקולה להיפותזה h שהוחזרה ע"י אלגוריתם ה $nn-1$.

נסמן $S = ((x_1,y_1),\dots,(x_m,y_m))$ כך ש $x_1 < \dots < x_m$. בנוסף, נסמן $i_1 < \dots < i_l$ את האינדקסים בהם $y_{i_j} \ne y_{i_j+1}$. נבחין כי $l \le m-1$. נחלק למקרים:

i. $y_1 = 0$: נגדיר לכל $1 \le j \le l$ נגדיר $a_j = \frac{x_{i_j}+x_{i_j+1}}{2}$, ובנוסף $a_{l+1} = a_{l+2} = \dots = a_k = 1$. הגדרה זו מוגדרת היטב משום ש $l \le m-1 < m \le k$.
ניתן לראות כי $f_{a_1,\dots,a_k} \in H_k$, ונותר להראות כי לכל $x \in \chi$ מתקיים $h(x) = f_{a_1,\dots,a_k}(x)$.
יהי $x \in \chi$. נסתכל על (x_i,y_i) בסדרה S כך ש x_i הינו ה nn של x. ע"פ הגדרתו $h(x) = y_i$.
אם $y_i = 0$ מתקיים כי $x_i < a_1$ או $a_{i_j} \le x_i \le a_{i_j+1}$ כך ש i_j זוגי, ולפי הבנייה גם x נמצא בטווח זה ולכן $f_{a_1,\dots,a_k}(x) = 0 = y_i$.
אם $y_i = 1$ מתקיים כי $a_{i_j} \le x_i \le a_{i_j+1}$ כך ש i_j אי זוגי, ולפי הבנייה גם x נמצא בטווח זה ולכ ולכן $f_{a_1,\dots,a_k}(x) = 1 = y_i$.
בסה"כ הראינו כי $h(x) = f_{a_1,\dots,a_k}(x)$ לכל $x \in \chi$.

ii. $y_1 = 1$: נגדיר $a_1 = 0$, ולכל $1 \le j \le l-1$ נגדיר $a_{j+1} = \frac{x_{i_j}+x_{i_j+1}}{2}$, ובנוסף $a_l = a_{l+1} = \dots = a_k = 1$. הגדרה זו מוגדרת היטב משום ש $l \le m-1 < m \le k$.
ניתן לראות כי $f_{a_1,\dots,a_k} \in H_k$, ונותר להראות כי לכל $x \in \chi$ מתקיים $h(x) = f_{a_1,\dots,a_k}(x)$, באופן דומה לסעיף הקודם.

c. נגדיר $S = ((x_1,1),(x_2 0),(x_3,1),(x_4,0),\dots,(x_m,1))$. ע"פ הגדרת H_k, על מנת ש $f_{a_1,\dots,a_k} \in H_k$ תהיה שקולה להיפותזה שהוחזרה ע"י אלגוריתם ה $nn-1$, צריך "להציב" נקודה a_i לפני x_1, בין כל שתי נקודות במדגם, ואחרי x_m, משמע חייב "להציב" $k < m+1$ נקודות, ולכן לא קיימת $f_{a_1,\dots,a_k} \in H_k$ השקולה להיפותזה שהוחזרה ע"י אלגוריתם ה $nn-1$.