מבוא למערכות לומדות (236756) | תרגיל בית 2 גדול

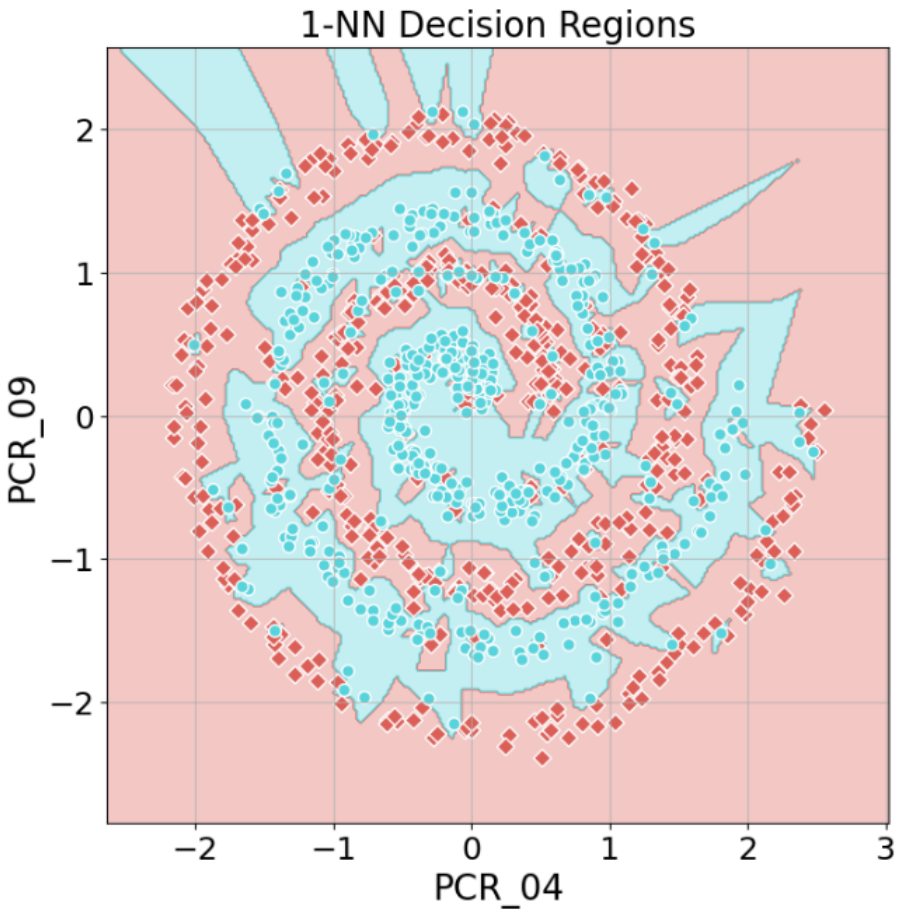
ליאל פרבר | 214413437

ראובן טימסיט | 330083858

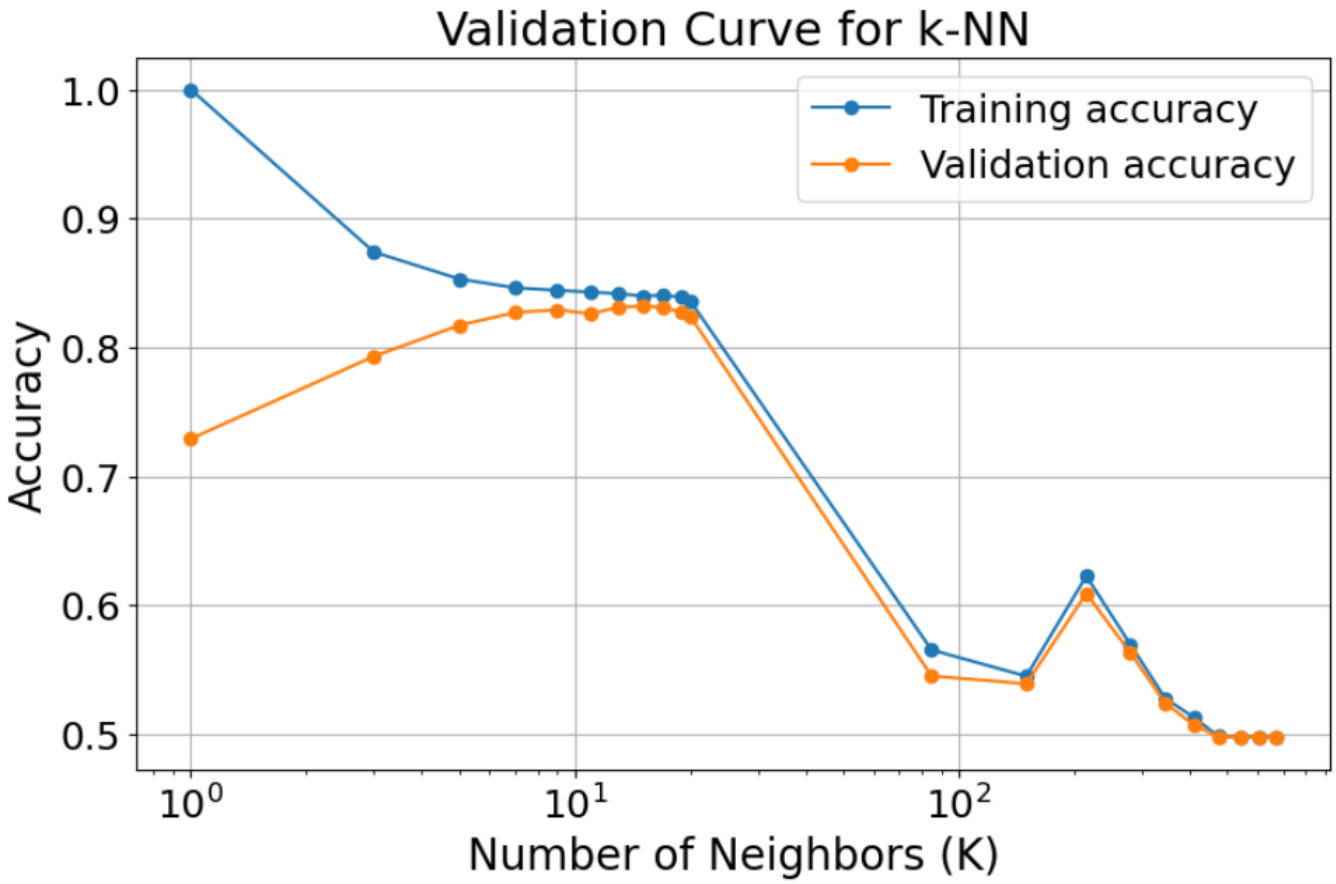
30 ביולי 2024

שאלה 1

להלן המסווג המתקבל:



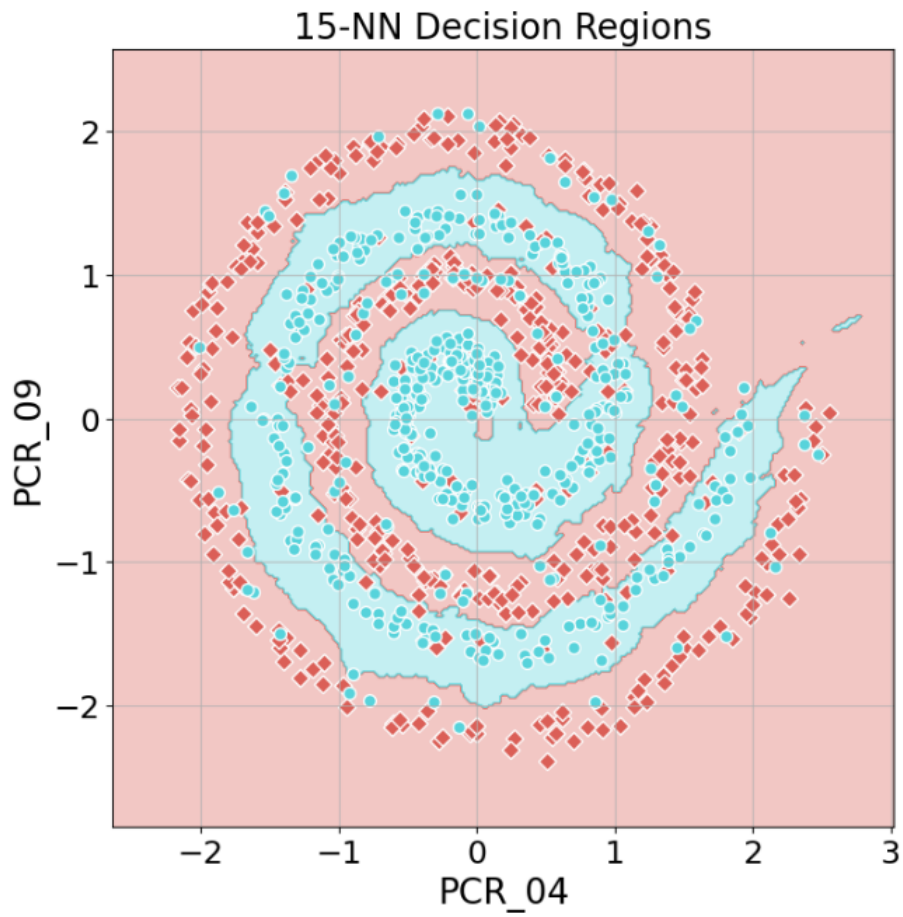
שאלה 2



נחשב (במחברת הפייתון) את k הטוב ביותר (שמניב אחוז דיוק מקסימלי ב- $validation\ set$):

- ה- k הטוב ביותר הוא $k = 15$. עבור k זה:
  - ממוצע אחוז הדיוק של המודל על ה- $training\ set$ הוא 0.8399.
  - ממוצע אחוז הדיוק של המודל על ה- $validation\ set$ הוא 0.8320.
- ה- k שגורמים ל- $overfitting$ הם אלו שמניבים $average\ training\ accuracy > 0.8399$. זאת משום שהם גורמים למודל להשקיע יותר מידי מאמצים בהתאמתו לחיזוי ה- $training\ set$ וכתוצאה ה- $average\ validation\ accuracy$ אינו מקסימלי. ה- k הללו הם $[1, 3, 5, 7, 9, 11, 13, 17]$. למשל עבור $k = 1$ נקבל דיוק מקסימלי על $training\ set$ ודיוק של כ- 0.725 על $validation\ set$.
- ה- k שגורמים ל- $underfitting$ הם אלו שמניבים $average\ training\ accuracy < 0.8399$. זאת משום שהם גורמים למודל להשקיע פחות מידי מאמצים בהתאמתו לחיזוי ה- $training\ set$ וכתוצאה ה- $average\ validation\ accuracy$ אינו מקסימלי. ה- k הללו הם $[19, 20, 85, 150, 215, 280, 345, 410, 475, 540, 605, 670]$. למשל עבור $k = 670$ נקבל דיוק נמוך מאוד (כמעט מינימלי) על $training\ set$ וה- $validation\ set$ עקב סיווג $majority$ שהמסווג מבצע לכל נקודה.

שאלה 3



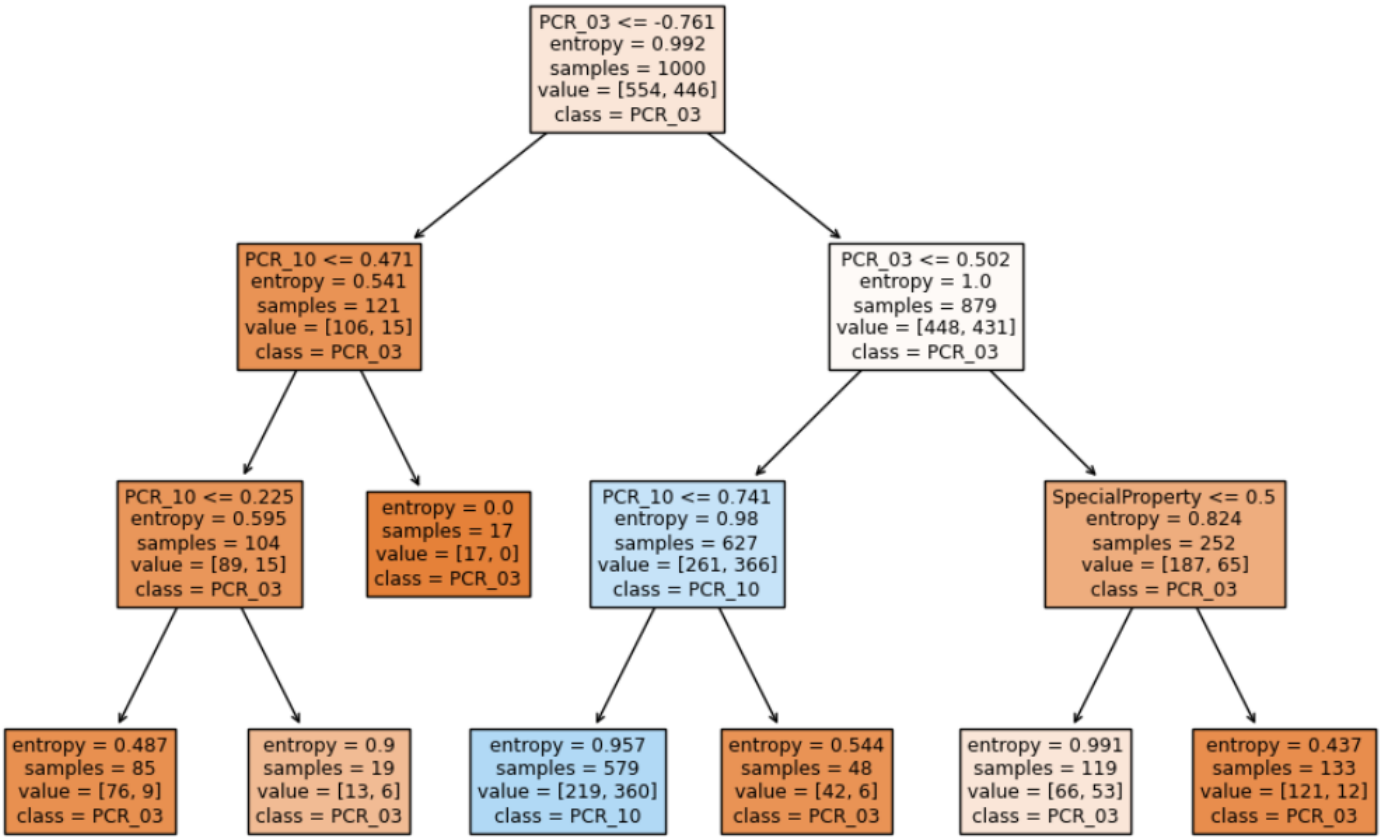

```
Test Accuracy: 0.82
```

כפי שניתן לראות אחוז דיוק המודל בחיזוי קבוצת המבחן הוא 0.82.

שאלה 4

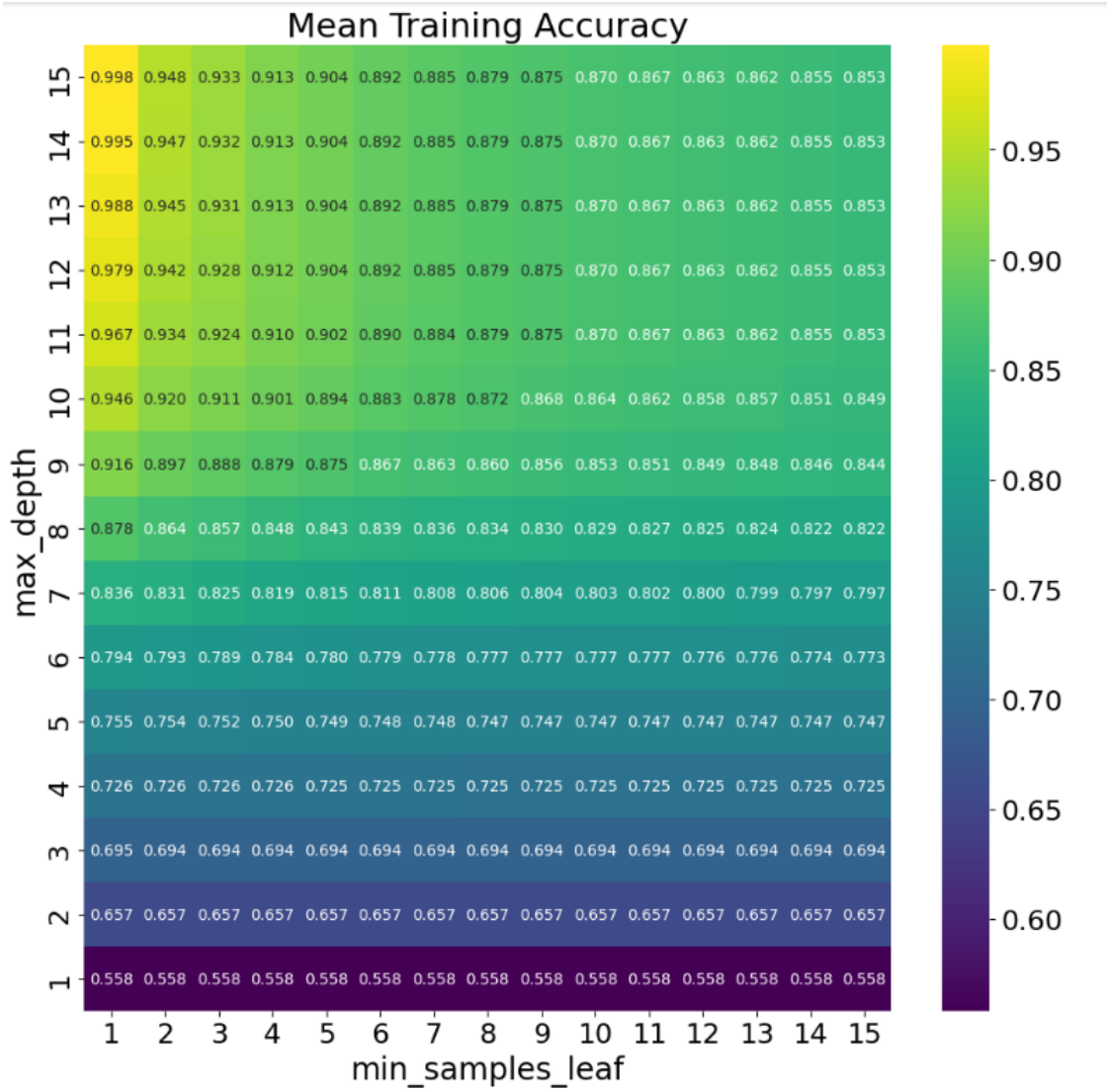
גבולות ההחלטה של המודל עם $k = 1$ מורכבים ומפוזרים יותר מאשר במודל עם $k = 15$, מה שמצביע על $overfitting$ גדול יותר אצל $k = 1$. כמו כן, נשים לב שב- $k = 1$ יש ניסיון לסווג נכון גם דוגמאות רועשות (דוגמאות כחולות שמוקפות בדוגמאות אדומות) בעוד שב- $k = 15$ זה לא מתרחש ודוגמאות אלו מסווגות באופן שגוי. לפיכך, $k = 1$ פחות מתאים מאשר $k = 15$ לחיזוי דוגמאות מבחן ולכן אחוז הדיוק של $k = 15$ על קבוצת מבחן (או ולידציה) נמוך יותר משל $k = 15$.

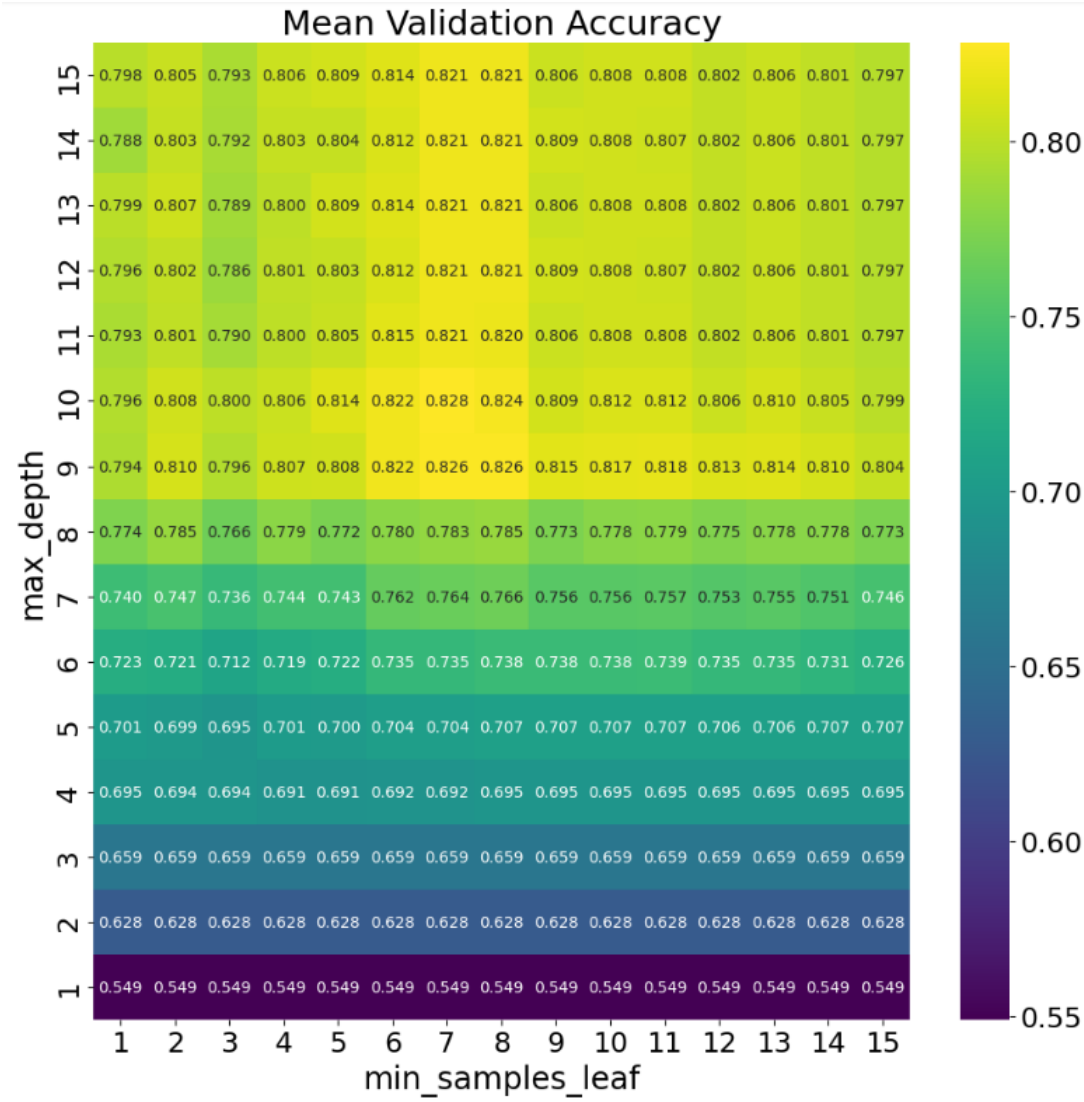
שאלה 5



אחוז הדיוק של העץ על קבוצת האימון הוא 0.695.

שאלה 6


Mean Training Accuracy
max_depth
min_samples_leaf
15 0.998 0.948 0.933 0.913 0.904 0.892 0.885 0.879 0.875 0.870 0.867 0.863 0.862 0.855 0.853
14 0.995 0.947 0.932 0.913 0.904 0.892 0.885 0.879 0.875 0.870 0.867 0.863 0.862 0.855 0.853
13 0.988 0.945 0.931 0.913 0.904 0.892 0.885 0.879 0.875 0.870 0.867 0.863 0.862 0.855 0.853
12 0.979 0.942 0.928 0.912 0.904 0.892 0.885 0.879 0.875 0.870 0.867 0.863 0.862 0.855 0.853
11 0.967 0.934 0.924 0.910 0.902 0.890 0.884 0.879 0.875 0.870 0.867 0.863 0.862 0.855 0.853
10 0.946 0.920 0.911 0.901 0.894 0.883 0.878 0.872 0.868 0.864 0.862 0.858 0.857 0.851 0.849
9 0.916 0.897 0.888 0.879 0.875 0.867 0.863 0.860 0.856 0.853 0.851 0.849 0.848 0.846 0.844
8 0.878 0.864 0.857 0.848 0.843 0.839 0.836 0.834 0.830 0.829 0.827 0.825 0.824 0.822 0.822
7 0.836 0.831 0.825 0.819 0.815 0.811 0.808 0.806 0.804 0.803 0.802 0.800 0.799 0.797 0.797
6 0.794 0.793 0.789 0.784 0.780 0.779 0.778 0.777 0.777 0.777 0.777 0.776 0.776 0.774 0.773
5 0.755 0.754 0.752 0.750 0.749 0.748 0.748 0.747 0.747 0.747 0.747 0.747 0.747 0.747 0.747
4 0.726 0.726 0.726 0.726 0.725 0.725 0.725 0.725 0.725 0.725 0.725 0.725 0.725 0.725 0.725
3 0.695 0.694 0.694 0.694 0.694 0.694 0.694 0.694 0.694 0.694 0.694 0.694 0.694 0.694 0.694
2 0.657 0.657 0.657 0.657 0.657 0.657 0.657 0.657 0.657 0.657 0.657 0.657 0.657 0.657 0.657
1 0.558 0.558 0.558 0.558 0.558 0.558 0.558 0.558 0.558 0.558 0.558 0.558 0.558 0.558 0.558
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60



- הקומבינציה הטובה ביותר של $(min\_samples\_leaf, max\_depth)$ היא $(7, 10)$.
- נשים לב שעבור קומבינציה זו אנו מקבלים אחוז דיוק ממוצע של 0.878 על קבוצת האימון. לכן, באופן דומה לשאלה 2:
  - קומבינציה שגורמת ל- $overfitting$ היא כזו שמניבה $mean\ training\ accuracy > 0.878$. למשל, $(1, 15)$ שדורשת שנסווג דוגמאות עד שיש בעלה דוגמה יחידה וכתוצאה העץ המתקבל עמוק ואנו מבצעים מאמצים רבים מידי שפוגעים ב- $mean\ training\ accuracy$.
  - קומבינציה שגורמת ל- $underfitting$ היא כזו שמניבה $mean\ training\ accuracy < 0.878$. למשל, $(15, 1)$ שדורשת שנסווג דוגמאות בעץ ישר בשורש לפי סיווג $majority$ ולכן אנו מבצעים מאמצים מעטים מידי שפוגעים ב- $mean\ training\ accuracy$.

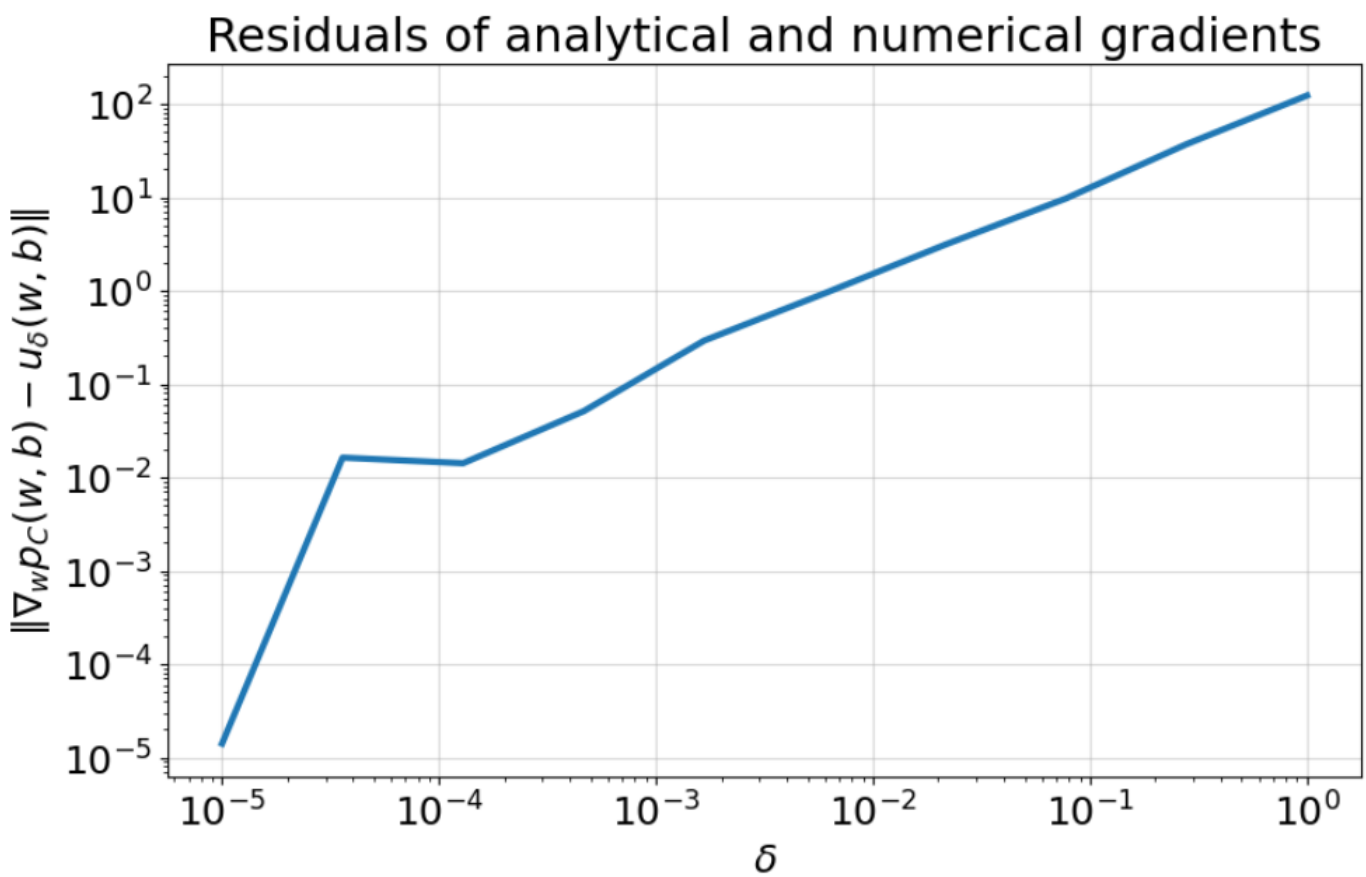
שאלה 7

ב- $Grid\ search$ שלנו בדקנו 15 אפשרויות שונות לכל פרמטר. כלומר, בסה"כ בדקנו $15^2 = 225$ קומבינציות שונות. אילו היינו מוסיפים פרמטר שלישי לחיפוש עם x ערכים אפשריים אז היינו צריכים לבדוק $225x$ קומבינציות שונות של 3 הפרמטרים.

שאלה 8

לאחר אימון עץ החלטה בעומק מקסימלי 10 עם מספר דוגמאות מינמילי בעלים 7, קיבלנו $test\ accuracy = 0.828$.

שאלה 9



כפי שניתן לראות עבור ערכי δ קטנים ההפרש בין הגרדיאנט האנליטי (ביחס ל- w) לגרדיאנט הנומרי (ביחס ל- w) קטן. עבור ערכי δ גדולים ההפרש גדל, כלומר השגיאה של הגרדיאנט הנומרי (ביחס לאנליטי) גדלה. כאשר $\delta \longrightarrow 0$ מתקיים שההפרש הנ"ל שואף ל- 0. ואכן, זו התוצאה לה אנו מצפים שכן $u_\delta(w,b) \underset{\delta\to 0}{\longrightarrow} \nabla_w p_C(w,b)$ לפי הגדרת הגרדאינט ונגזרות חלקיות.

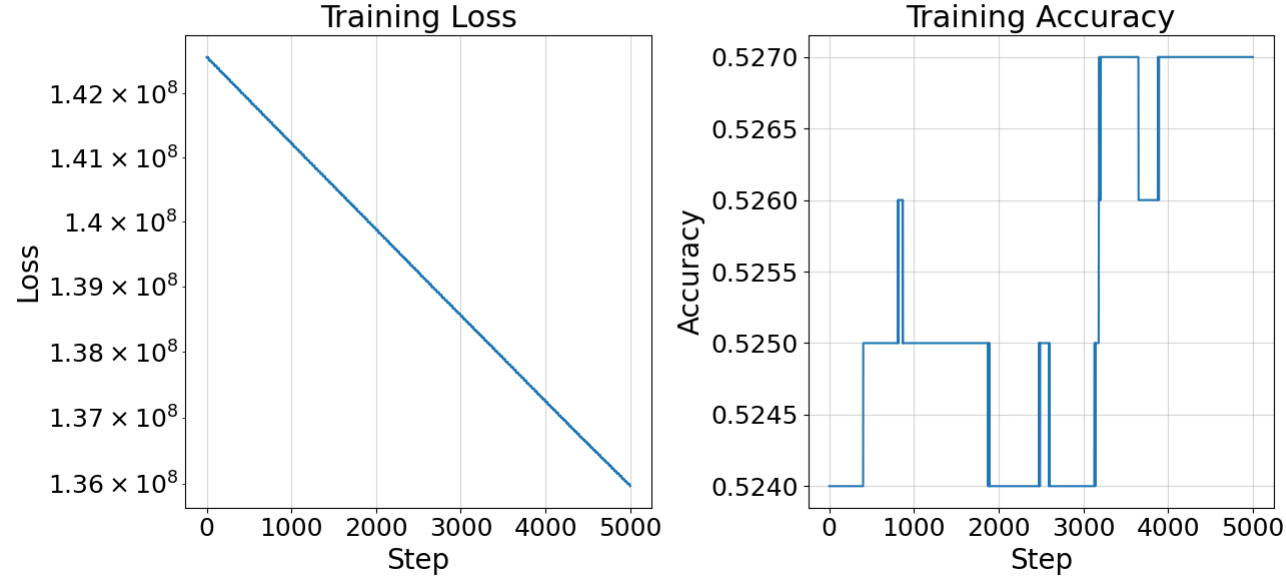
שאלה 10

אנו רואים שבמשך הזמן ה- $Train\ Loss$ קטן, כאשר ב- 500 הצעדים הראשונים קצב הדעיכה שלו גדול מאוד ומ- 500 עד 4800 קצב הדעיכה קטן. לקראת 4800 צעדים (ומעלה) ה- $Train\ Loss$ מייצב על ערך מינימלי. במקביל, ה- $Train\ Accuracy$ חווה עליות וירידות במשך 5000 הצעדים בניגוד למצופה - אנו מצפים שככל שה- $Train\ Loss$ קטן כך יגדל ה- $Train\ Accuracy$. נשים לב שב- 500 הצעדים הראשונים ה- $Train\ Accuracy$ עולה כמצופה ביחס לירידה הגדולה של ה- $Train\ Loss$. לקראת 4800 צעדים (ומעלה) ה- $Train\ Accuracy$ מתייצב כמצופה כפי שה- $Train\ Loss$ התייצב.

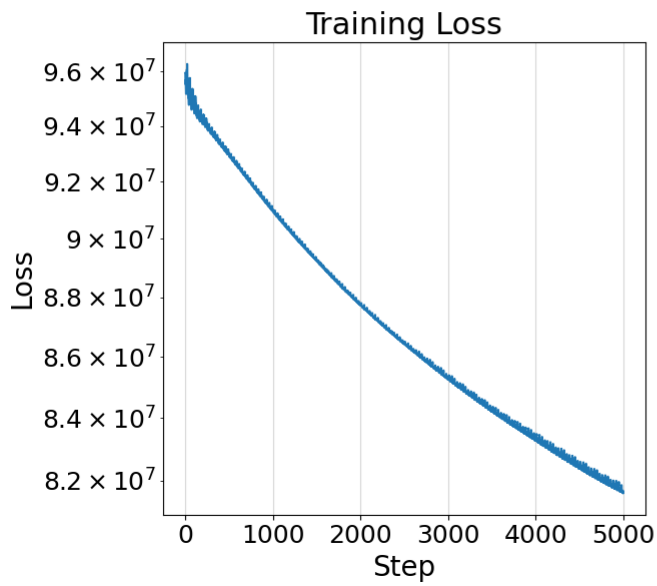
נדרשים אלפי צעדים כדי להתייצב על $Train\ Loss$ מינימלי (מינימום מקומי - לא בהכרח גלובלי) משום שהפרמטר $learning\ rate$ של SGD קטן מאוד. בנוסף, משום שהפרמטר C של SGD גדול מאוד אז נוצר $overfitting$ של קבוצת האימון שככל הנראה פוגע ב- $Train\ Accuracy$. נשים לב שיש הרבה דוגמאות רועשות ובכלל קבוצת האימון מאוד מבולגנת ונראה כי לא ניתן להשיג אחוז דיוק של יותר מכ- 0.6. לכן, כמצופה המודל לא מייצר מסווג טוב. למשל, מצעד 500 ועד צעד 2200, ומצעד 3300 ועד צעד 4500 יש דעיכה של $Train\ Accuracy$ עקב הניסיון הנואש של המודל להתאים עצמו לדוגמאות רועשות מבולגנות מאוד.

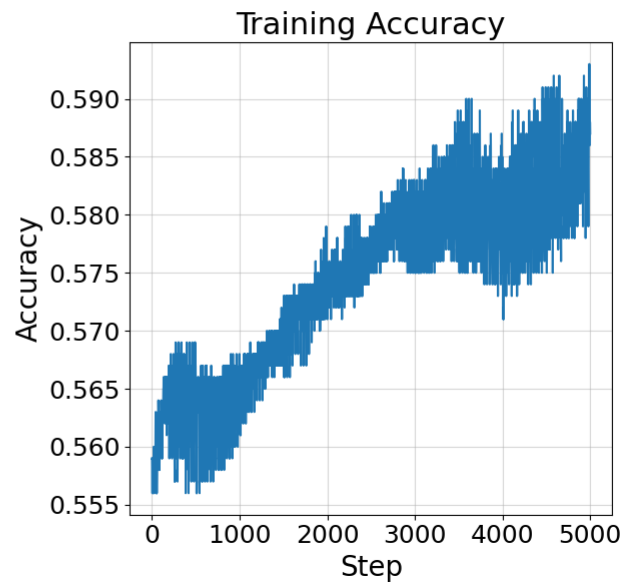
שאלה 11

```
max accuracy and min loss for lr = 1e-11

max accuracy: 0.527
max accuracy iteration: 3188

min loss: 135962642.458
min loss iteration: 5000
```

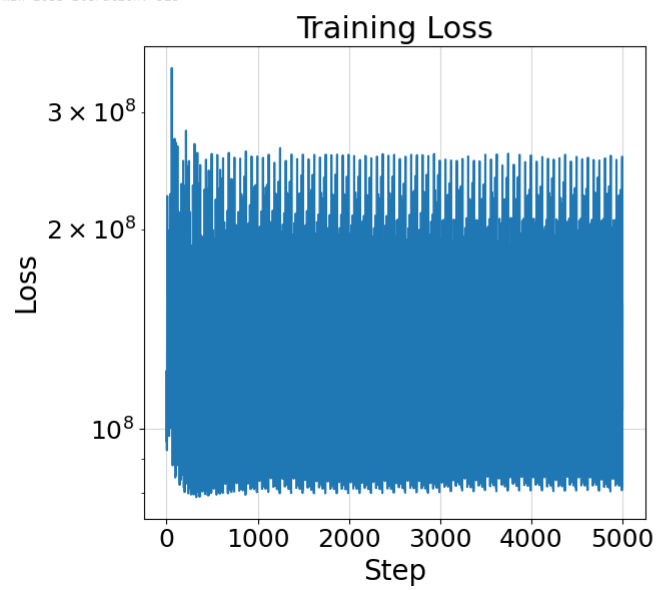


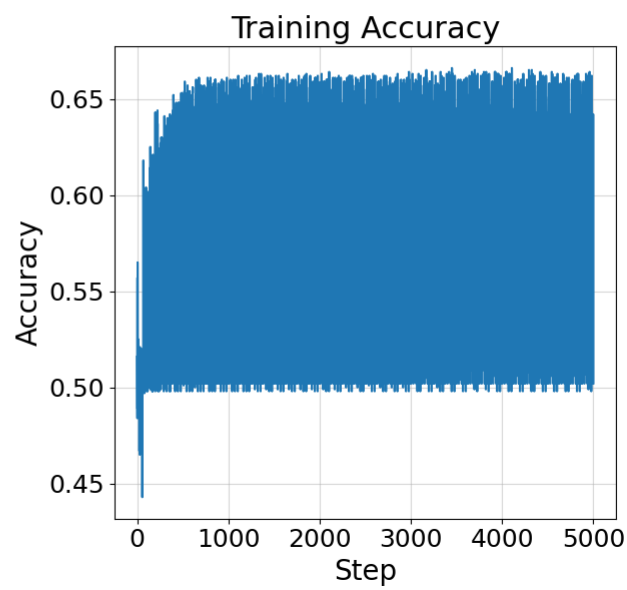
```
max accuracy and min loss for lr = 1e-09

max accuracy: 0.593
max accuracy iteration: 4991

min loss: 81601976.827
min loss iteration: 4999
```



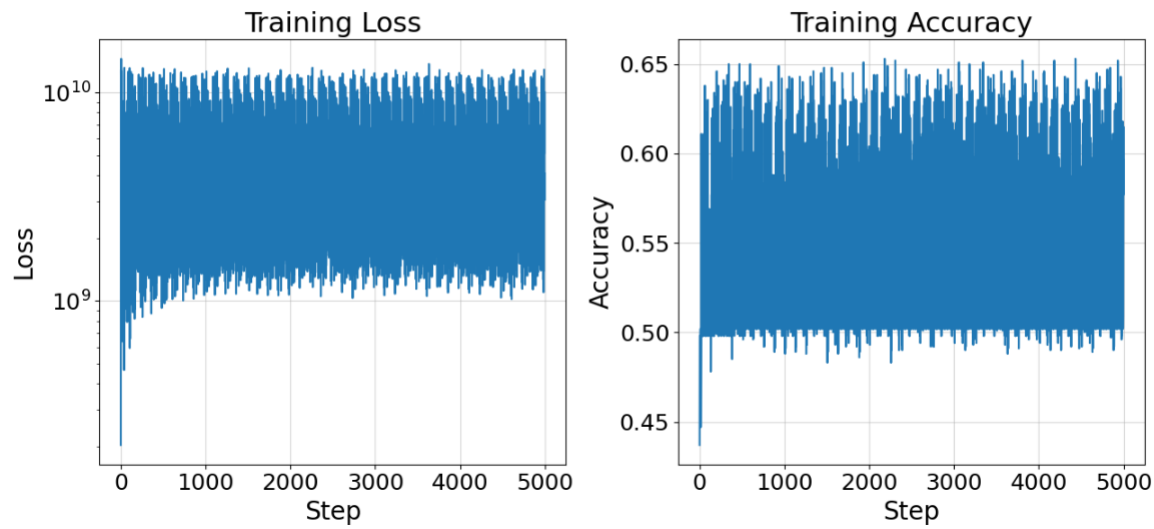



```
max accuracy and min loss for lr = 1e-07

max accuracy: 0.666
max accuracy iteration: 3451

min loss: 78912272.209
min loss iteration: 328
```

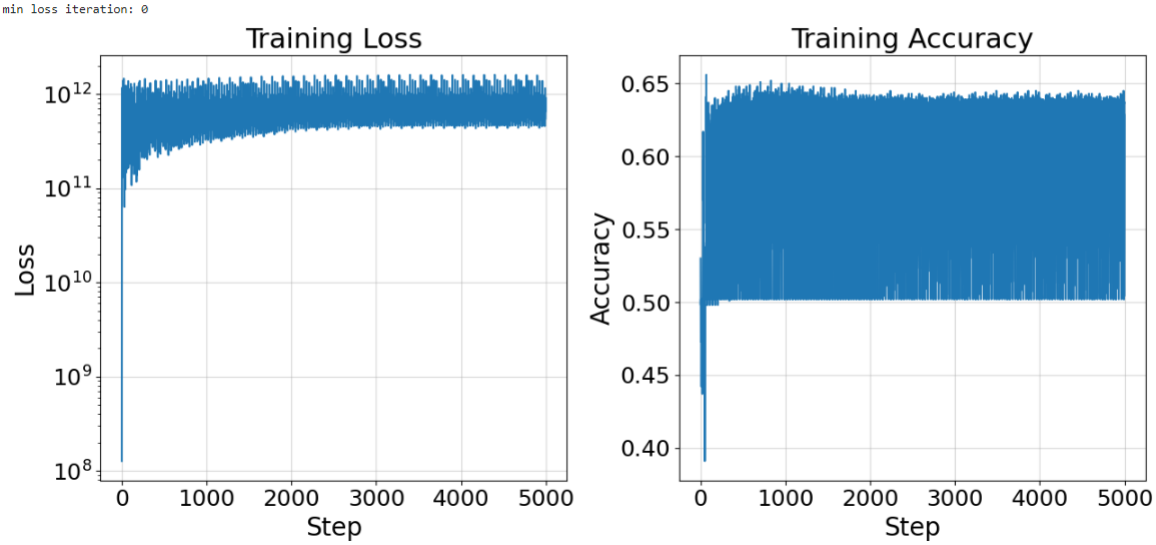




```
max accuracy and min loss for lr = 1e-05

max accuracy: 0.653
max accuracy iteration: 2182

min loss: 203454696.261
min loss iteration: 0
```




```
max accuracy and min loss for lr = 0.001

max accuracy: 0.656
max accuracy iteration: 66

min loss: 126705278.306
min loss iteration: 0
```



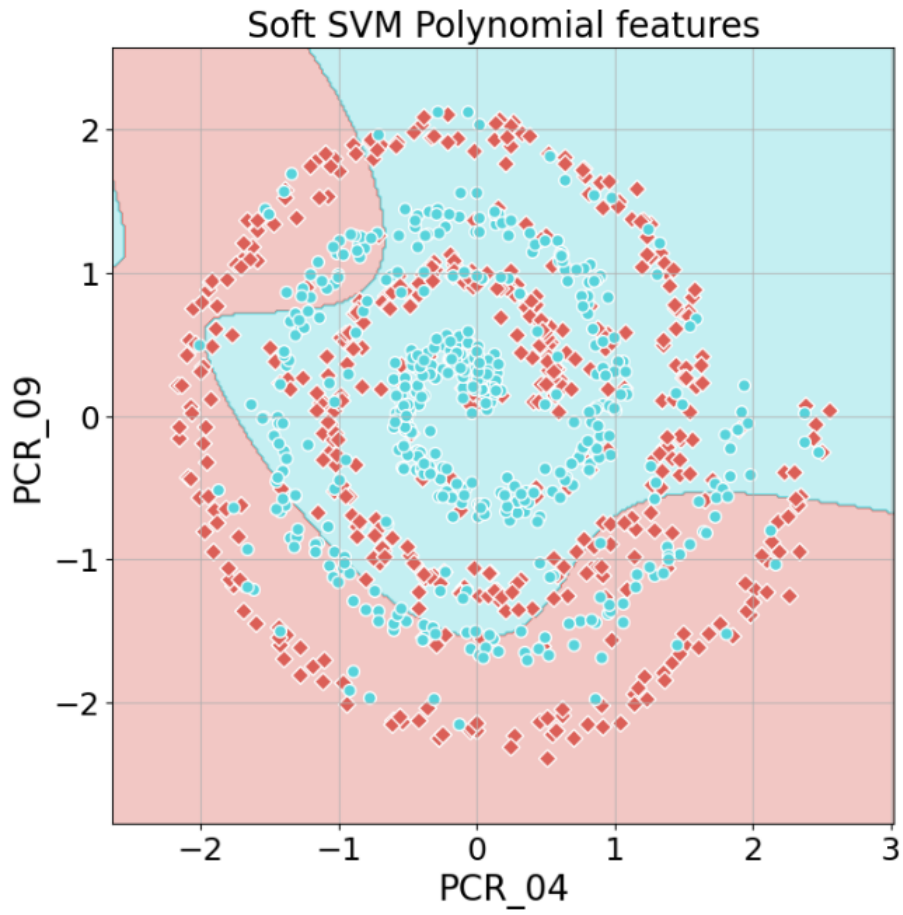
לאחר שניסינו קצבי למידה שונים, גילינו כי השימוש ב- $lr = 1e^{-7}$ מספק את התוצאות הטובות ביותר עבור תהליך האימון שלנו :

- קצבי למידה נמוכים יותר כמו $lr = 1e^{-9}$ או $lr = 1e^{-11}$ הובילו להתכנסות איטית מאוד של ה- $training\ loss$ וה- $training\ accuracy$.

- עבור $lr = 1e^{-5}$ נקבל $training\ loss$ גדול יותר (פי כמה עשרות).

- עבור $lr = 0.001$ נקבל אומנם התכנסות קצת יותר מהירה ל- $training\ accuracy$. אבל עבור $lr = 1e^{-7}$ נקבל $training\ accuracies$ גבוהים יותר בחלק מהמקרים. כמו כן, עבור $lr = 0.001$ ה- $training\ loss$ גדול יותר (פי כ- 10^4).

שאלה 12




```
Train Accuracy: 0.538
Test Accuracy: 0.576
```

כפי שניתן לראות אחוז הדיוק על קבוצת האימון הוא 0.538 ועל קבוצת המבחן הוא 0.576.

שאלה 13

a.

כלל החיזוי הוא:

$$h(x) = sign\left(\sum_{i\in[m],\alpha_i>0} \alpha_i \cdot y_i \cdot K(x, x_i)\right) = sign\left(\sum_{i\in[m],\alpha_i>0} \alpha_i \cdot y_i \cdot e^{-\gamma\cdot\|x-x_i\|_2}\right) \underset{\alpha=\overline{1}}{=}$$

$$= sign\left(\sum_{i\in[m]} y_i \cdot e^{-\gamma\cdot\|x-x_i\|_2}\right)$$

לכל $x \in D$.

b.

מתקיים:

$$h(x) = sign\left(\sum_{i\in[m]} y_i \cdot e^{-\gamma\cdot\|x-x_i\|_2}\right) = sign\left(\sum_{i\in[m]} y_i \cdot K(x, x_i)\right) =$$

$$= sign\left(\sum_{j\in\{i\in[m]|y_i=1\}} K(x, x_j) - \sum_{j\in\{i\in[m]|y_i=-1\}} K(x, x_j)\right)$$

לכל $x \in D$.

.c

נתון שעבור הדוגמה הנתונה $x \in D$, עם $y = 1$, קיים $(x_p, y_p) \in S$ כך ש- $y = y_p$ וגם $\|x - x_p\|_2 < \delta - 1$. מתקיים $p \in \{i \in [m] \,|y_i = 1\}$ ולכן:

$$\sum_{j\in\{i\in[m]|y_i=1\}} K(x, x_j) = \sum_{j\in\{i\in[m]|y_i=1\}\setminus\{p\}} e^{-\gamma\cdot\|x-x_j\|_2} + e^{-\gamma\cdot\|x-x_p\|_2} >$$

$$\underset{e^t>0 \; for \; each \; t\in\mathbb{R}}{>} e^{-\gamma\cdot\|x-x_p\|_2} \underset{\|x-x_p\|_2<\delta-1 \; and \; \gamma>0}{>} e^{-\gamma\cdot(\delta-1)}$$

.d

- נתון ש- $\frac{m}{2} = |\{x_i|y_i = 1\}| = m - |\{x_i|y_i = -1\}|$ ולכן $|\{x_i|y_i = -1\}| = \frac{m}{2}$.$\otimes$
- נתון שעבור הדוגמה הנתונה $x \in D$, עם $y = 1$, קיים $(x_p, y_p) \in S$ כך ש- $1 = y = y_p$ וגם $\|x - x_p\|_2 < \delta - 1$.$\circledast$
- נתון שלכל $i \neq j \in [m]$ מתקיים $\|x_i - x_j\|_2 > 3\delta$.$\circledast\circledast$

לכן:

$$\sum_{j\in\{i\in[m]|y_i=-1\}} K(x, x_j) = \sum_{j\in\{i\in[m]|y_i=-1\}} e^{-\gamma\cdot\|x-x_j\|_2} =$$

$$= \sum_{j\in\{i\in[m]|y_i=-1\}} e^{-\gamma\cdot\|x_j-x\|_2} =$$

$$= \sum_{j\in\{i\in[m]|y_i=-1\}} e^{-\gamma\cdot\|x_j-x_p+x_p-x\|_2} \underset{triangle \; inequality \; and \; \gamma>0}{\leq}$$

$$\leq \sum_{j\in\{i\in[m]|y_i=-1\}} e^{-\gamma\cdot\left|\|x_j-x_p\|_2-\|x_p-x\|_2\right|} \leq$$

$$\leq \sum_{j \in \{i \in [m] | y_i = -1\}} e^{-\gamma \cdot \left(\|x_j - x_p\|_2 - \|x_p - x\|_2 \right)}$$

מתקיים $p \notin \{i \in [m] \,|\, y_i = -1\}$ כי $y_p = 1$. מ- $\circledast\circledast$ לכל $j \in \{i \in [m] \,|\, y_i = -1\}$ מתקיים $\|x_j - x_p\| > 3\delta$. ולכן $\gamma > 0$ $\gamma \|x_j - x_p\| > 3\gamma\delta$. בנוסף, מ- $\circledast$ נקבל $\gamma \|x_p - x\|_2 < \gamma(\delta - 1)$. לכן:

$$\sum_{j \in \{i \in [m] | y_i = -1\}} e^{-\gamma \cdot \left(\|x_j - x_p\|_2 - \|x_p - x\|_2\right)} =$$

$$= \sum_{j \in \{i \in [m] | y_i = -1\}} e^{-\gamma \|x_j - x_p\|_2 + \gamma \|x_p - x\|_2} <$$

$$< \sum_{j \in \{i \in [m] | y_i = -1\}} e^{-3\gamma\delta + \gamma(\delta - 1)} \underset{\otimes}{=} \frac{m}{2} \cdot e^{-3\gamma\delta + \gamma(\delta - 1)} =$$

$$= \frac{m}{2} \cdot e^{-\gamma \cdot (2\delta + 1)} \underset{2\delta + 1 > 2\delta - 1 \; and \; \gamma > 0}{<} \frac{m}{2} \cdot e^{-\gamma \cdot (2\delta - 1)}$$

e.

משום ש- $\gamma = ln\left(\frac{m}{2}\right)$ מתקיים:

$$\sum_{j \in \{i \in [m] | y_i = 1\}} K(x, x_j) - \sum_{j \in \{i \in [m] | y_i = -1\}} K(x, x_j) >$$

$$\underset{sections \; c+d}{>} e^{-\gamma \cdot (\delta - 1)} - \frac{m}{2} \cdot e^{-\gamma \cdot (2\delta - 1)} \underset{\gamma = ln\left(\frac{m}{2}\right)}{=}$$

$$= e^{ln\left(\left(\frac{m}{2}\right)^{-(\delta - 1)}\right)} - \frac{m}{2} \cdot e^{ln\left(\left(\frac{m}{2}\right)^{-(2\delta - 1)}\right)} = \left(\frac{m}{2}\right)^{-(\delta - 1)} - \frac{m}{2} \cdot \left(\frac{m}{2}\right)^{-(2\delta - 1)} =$$

$$= \left(\frac{m}{2}\right)^{1 - \delta} - \left(\frac{m}{2}\right)^{2 - 2\delta} = \left(\frac{m}{2}\right)^{1 - \delta} \left(1 - \left(\frac{m}{2}\right)^{1 - \delta}\right)$$

משום ש- $\delta > 1$ מתקיים $1 - \delta < 0$. משום ש- $1 \leq \frac{m}{2}$ (כי m זוגי - אחרת $\left|\{x_i|y_i = 1\}\right| = \frac{m}{2} \notin \mathbb{N}$ וזו סתירה) נקבל $1 = \left(\frac{m}{2}\right)^0 > \left(\frac{m}{2}\right)^{1-\delta}$. כלומר, $1 - \left(\frac{m}{2}\right)^{1-\delta} > 0$. כמו כן, $\left(\frac{m}{2}\right)^{1-\delta} > 0$ ולכן:

$$\sum_{j\in\{i\in[m]|y_i=1\}} K\left(x, x_j\right) - \sum_{j\in\{i\in[m]|y_i=-1\}} K\left(x, x_j\right) > 0$$

כלומר:

$$h\left(x\right) = sign\left(\sum_{j\in\{i\in[m]|y_i=1\}} K\left(x, x_j\right) - \sum_{j\in\{i\in[m]|y_i=-1\}} K\left(x, x_j\right)\right) = 1$$

f.

תחת ההנחות הנתונות, כולל $\gamma = ln\left(\frac{m}{2}\right)$:

עבור דוגמאות x עם תיוג $y = 1$ הראנו ש- $h\left(x\right) = 1 = y$. עבור דוגמאות x עם תיוג $y = -1$ מתקיים:

קיים $\left(x_p, y_p\right) \in S$ כך ש- $y = y_p$ וגם $\left\|x - x_p\right\|_2 < \delta - 1$. מתקיים $p \in \left\{i \in [m] \left|y_i = -1\right.\right\}$ ולכן:

$$\sum_{j\in\{i\in[m]|y_i=-1\}} K\left(x, x_j\right) = \sum_{j\in\{i\in[m]|y_i=-1\}\setminus\{p\}} e^{-\gamma\cdot\left\|x-x_j\right\|_2} + e^{-\gamma\cdot\left\|x-x_p\right\|_2} >$$

$$\underset{e^t>0 \ for \ each \ t\in\mathbb{R}}{>} e^{-\gamma\cdot\left\|x-x_p\right\|_2} \underset{\left\|x-x_p\right\|_2<\delta-1 \ and \ \gamma>0}{>} e^{-\gamma\cdot(\delta-1)}$$

מתקיים $p \notin \{i \in [m] \,|\, y_i = 1\}$ כי $y_p = -1$. לכל $j \in \{i \in [m] \,|\, y_i = 1\}$ מתקיים $\|x_j - x_p\| > 3\delta$. ולכן $\gamma > 0$ $\gamma \|x_j - x_p\| > 3\gamma\delta$. בנוסף, $\gamma \|x_p - x\|_2 < \gamma(\delta - 1)$. לכן:

$$\sum_{j \in \{i \in [m] | y_i = 1\}} K(x, x_j) = \sum_{j \in \{i \in [m] | y_i = 1\}} e^{-\gamma \cdot \|x_j - x_p + x_p - x\|_2} =$$

$$\underset{\textit{triangle inequality and } \gamma > 0}{\leq}$$

$$\sum_{j \in \{i \in [m] | y_i = 1\}} e^{-\gamma \cdot \left(\|x_j - x_p\|_2 - \|x_p - x\|_2\right)} =$$

$$= \sum_{j \in \{i \in [m] | y_i = 1\}} e^{-\gamma \|x_j - x_p\|_2 + \gamma \|x_p - x\|_2} <$$

$$< \sum_{j \in \{i \in [m] | y_i = 1\}} e^{-3\gamma\delta + \gamma(\delta - 1)} =$$

$$= |\{i \in [m] \,|\, y_i = 1\}| \cdot e^{-3\gamma\delta + \gamma(\delta - 1)} =$$

$$= \frac{m}{2} \cdot e^{-\gamma \cdot (2\delta + 1)} \underset{2\delta + 1 > 2\delta - 1 \textit{ and } \gamma > 0}{<} \frac{m}{2} \cdot e^{-\gamma \cdot (2\delta - 1)}$$

לפיכך:

$$\sum_{j\in\{i\in[m]|y_i=1\}} K\left(x,x_j\right) - \sum_{j\in\{i\in[m]|y_i=-1\}} K\left(x,x_j\right) <$$

$$< \frac{m}{2}\cdot e^{-\gamma\cdot(2\delta-1)} - e^{-\gamma\cdot(\delta-1)} \underset{\gamma=ln\left(\frac{m}{2}\right)}{=}$$

$$= \left(\frac{m}{2}\right)^{2-2\delta} - \left(\frac{m}{2}\right)^{1-\delta} = \left(\frac{m}{2}\right)^{1-\delta}\left(\left(\frac{m}{2}\right)^{1-\delta} - 1\right)$$

משום ש- $1-\left(\frac{m}{2}\right)^{1-\delta} > 0$ ו- $\left(\frac{m}{2}\right)^{1-\delta} > 0$ נקבל:

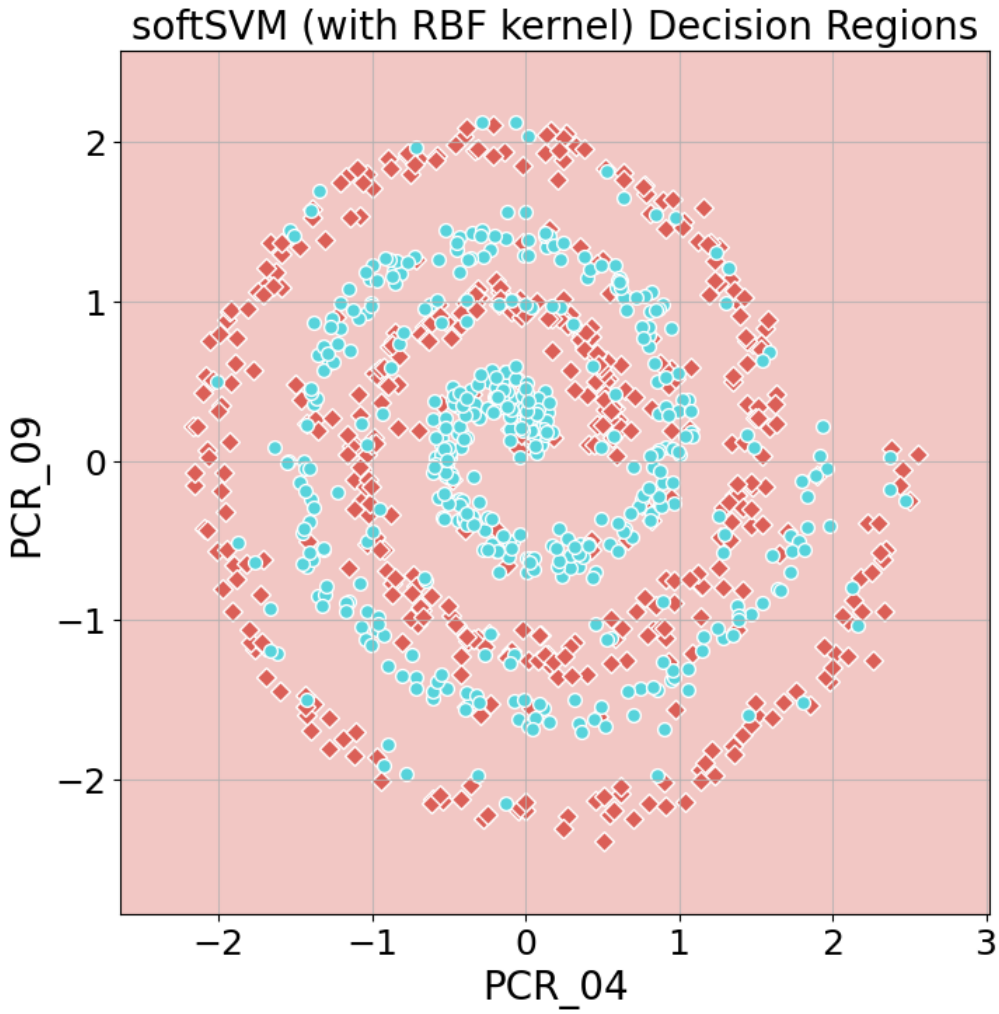
$$\sum_{j\in\{i\in[m]|y_i=1\}} K\left(x,x_j\right) - \sum_{j\in\{i\in[m]|y_i=-1\}} K\left(x,x_j\right) < 0$$

כלומר:

$$h\left(x\right) = sign\left(\sum_{j\in\{i\in[m]|y_i=1\}} K\left(x,x_j\right) - \sum_{j\in\{i\in[m]|y_i=-1\}} K\left(x,x_j\right)\right) = -1 = y$$

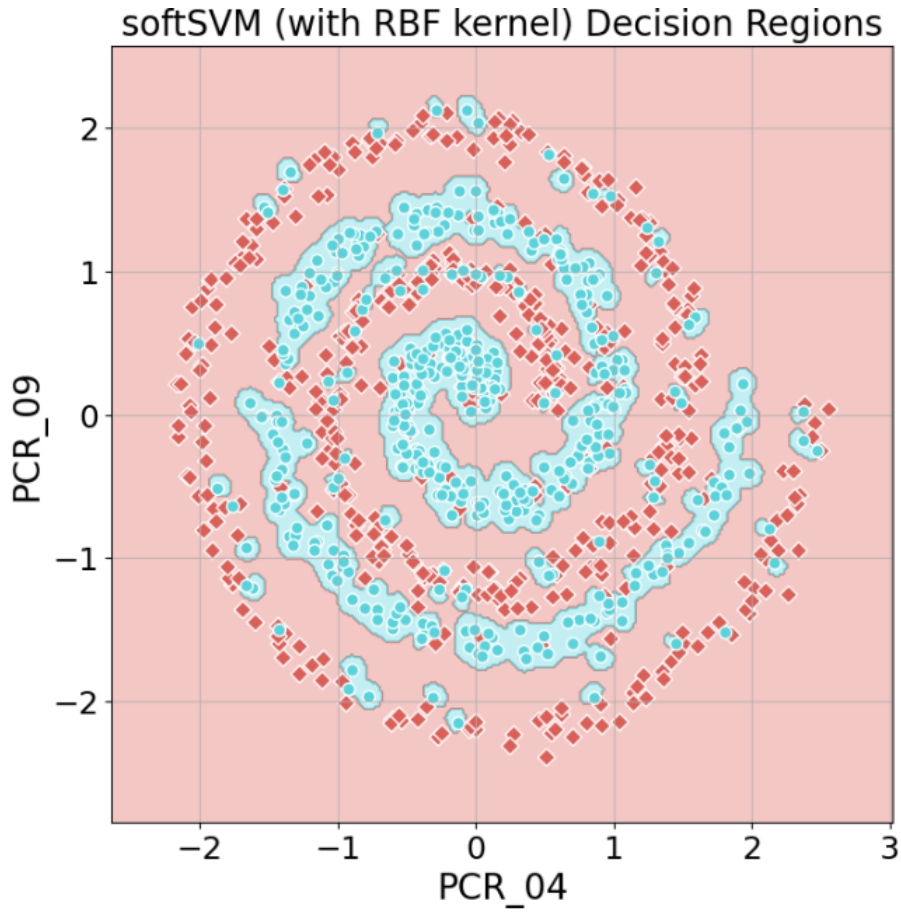
לכן, בכל מקרה $h\left(x\right)=y$ ו- h תמיד צודקת.

שאלה 14



כפי שניתן לראות קיבלנו מסווג גרוע מאוד שמהווה $under fitting$. זאת משום שקיבלנו $Train\ accuracy = 0.502$ ו- $Test\ accuracy = 0.492$ (ערכים נמוכים מאוד). כמו כן, נראה שהמסווג שהתקבל מסווג את הרוב המוחלט של נקודות המבחן כאדומות - פעולה לא רצויה בהחלט שכן ישנן גם דוגמאות כחולות.

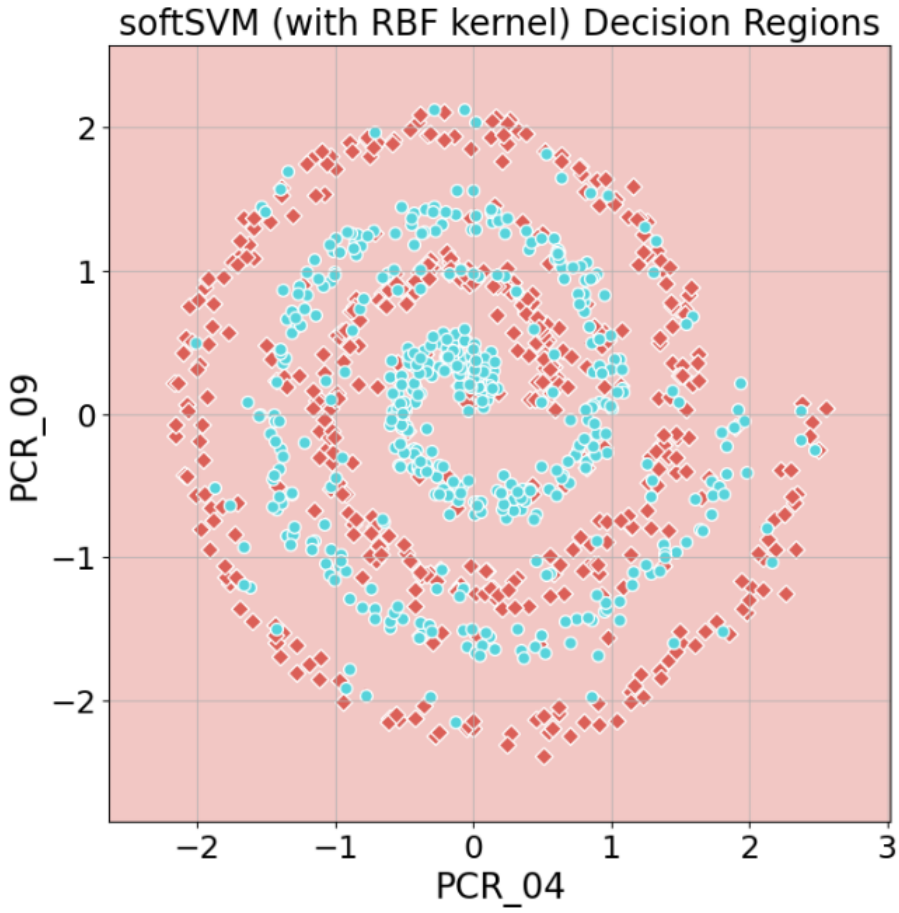
שאלה 15



כפי שניתן לראות קיבלנו מסווג הרבה יותר טוב מבשאלה 14. מסווג זה משיג $Train\ accuracy = 0.923$ ו- $Test\ accuracy = 0.752$. כלומר, הוא מתאים עצמו באופן נהדר לקבוצת האימון אך ההתאמה חזקה מידי (כלומר המסווג מהווה $overfitting$) וכתוצאה מכך אחוז ה- $Test\ accuracy$ שלו לא כל כך גבוה. בשאלה 3 מודל ה- $15-NN$ השיג $Test\ accuracy = 0.82 > 0.752$. כלומר, הוא טוב יותר עבור חיזוי קבוצות מבחן מאשר מודל ה- $SoftSVM$ שהשתמשנו בו בשאלה זו. הסיבה לכך היא שהמודל שלנו מבצע $overfitting$ חזק יותר לקבוצת האימון מאשר מודל ה- $15-NN$ וזאת כתוצאה מהפרמטר $\gamma = 200$ הגדול:

ככל ש- γ גדול יותר כך המודל דומה ל- kNN בעל k קטן יותר (כפי שראינו בתחילת חלק 4). כלומר, אנו מקבלים מודל "שמתחשב במעט שכנים (דוגמאות אימון שכנות) בסיווג דוגמת מבחן". היות שהראנו בשאלה 2 שה- k האופטימלי (שמשיג $Test\ accuracy$ מקסימלי) הוא 15 נצפה ש- $k < 15$ יוביל ל- $Test\ accuracy$ לא מקסימלי. כלומר, γ מספיק גדול (200 אכן מספיק גדול כפי שהתוצאות מראות) " מוביל ל- $k < 15$" <u>וכתוצאה המודל שלנו דומה למודל kNN לא אופטימלי</u>.

שאלה 16



כפי שניתן לראות קיבלנו מסווג הרבה פחות טוב מבשאלה 14. מסווג זה משיג $Train\ accuracy = 0.998$ ו- : $Test\ accuracy$ 0.552. כלומר, הוא מתאים עצמו באופן אידיאלי (כמעט) לקבוצת האימון אך ההתאמה חזקה מידי (כלומר המסווג מהווה $overfitting$) וכתוצאה מכך אחוז ה- $Test\ accuracy$ שלו נמוך מידי (ואחוז ה- $Train\ accuracy$ שלו כמעט מקסימלי). כפי שהסברנו בשאלה 15, <u>עקב השימוש ב- $\gamma = 5000$ המודל שלנו דומה ביותר למודל kNN עם k קטן מאוד</u> (ככל הנראה $k \in \{1, 2\}$). כפי שהראנו בשאלה 2, k כזה כלל אינו אופטימלי ומוביל ל- $Test\ accuracy$ לא טוב ובפרט לא מקסימלי.