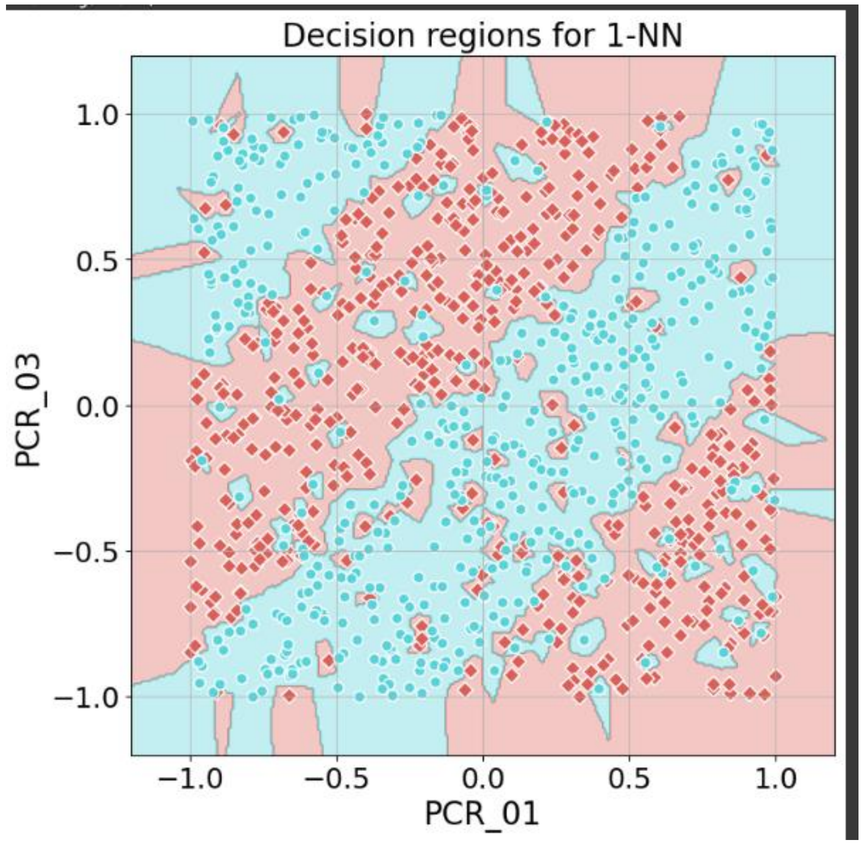
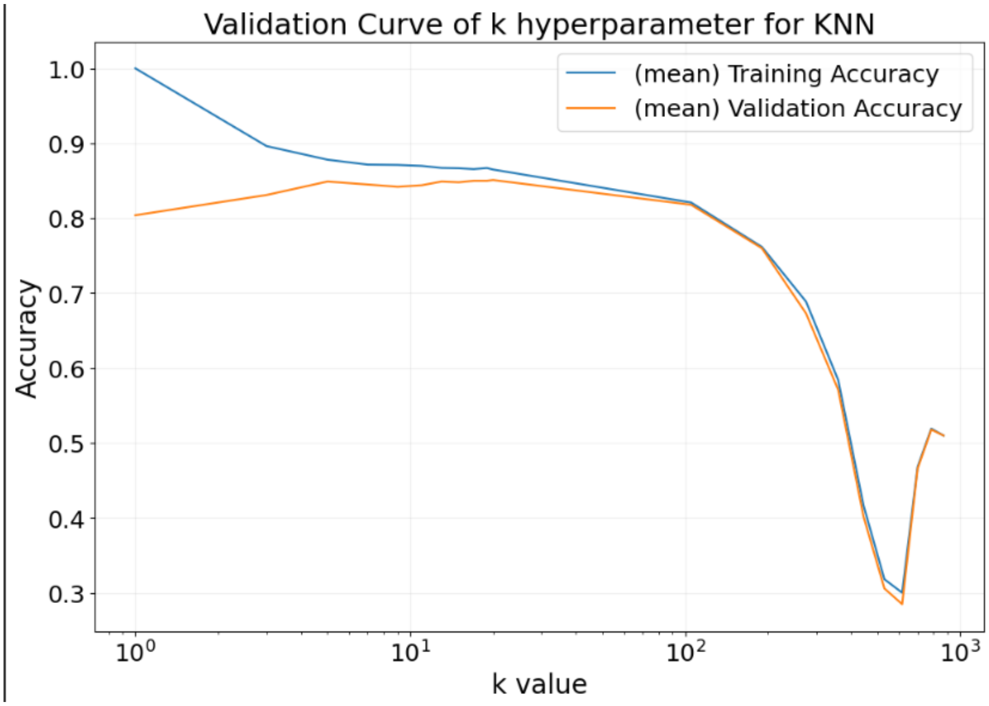
מבוא למערכות לומדות- תרגיל בית 2- דו״ח עבודה

(Q1)



Decision regions for 1-NN
PCR_03
PCR_01
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0

(Q2)



שתי העקומות מציגות לנו לכל k את מידת הדיוק הממוצעת מתהליך הcross-validation שהתקבלה מאימון של מודלי knn עם k שכנים.

אנחנו רוצים לבחור את הk שנתן לנו את הדיוק הכי טוב על הvalidation set (כיוון הvalidation set נותן לנו הערכה על ביצועי המודל על מידע שלא ראה). לכן, ניקח את הנקודה המקסימלית על העקומה הכתומה ונמצא את הk שנותן לנו את הדיוק המקסימלי.

עשינו את זה כמובן בעזרת קוד פייתון וקיבלנו :

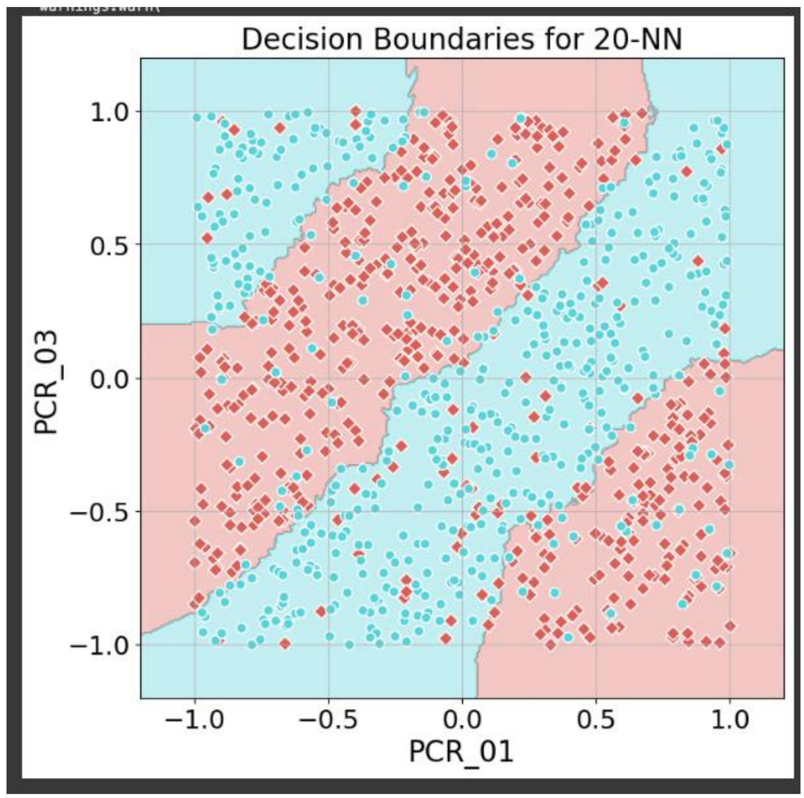
```
highest validation accuracy : 0.851) and
training accuracy when the highest validation accuracy : 0.865
for k : 20
```

<u>כלומר הk הטוב ביותר שמצאנו הוא k=20.</u>

ערכי k נמוכים גורמים לתופעת הoverfitting. ניתן לראות בגרף הvalidation curve כי עבור ערכי k נמוכים מידת הדיוק על קבוצת האימון היא גבוהה ומידת הדיוק על קבוצת הוולידציה נמוכה יותר.
בתרגול ראינו כי מצב כזה מקושר לoverfit, כלומר דיוק טוב מדי על הtraining set הוא לא בהכרח טוב ואף יכול לפגוע ביכולת ההכללה על הtest set.
ראינו בשיעורי בית הקודמים שכאשר מס׳ השכנים ב KNN הוא קטן , מתרחשת תופעת הoverfitting מאחר ויש משקל גדול יותר למספר קטן של השכנים הקרובים ביותר ולכן ההשפעה של כל שכן קרוב גדולה יותר. כתוצאה מכך האלגוריתם רגיש לרעש, ובמידה ויש נקודות חריגות בtraining set עלול להיווצר אזור סיווג שגוי מסביב לנקודות הרעש, כפי שרואים גם בplot בשאלה 1.

ערכי k גבוהים גורמים לתופעת הunderfitting. ניתן לראות בעקומה שלנו כי ככל שk גדל אנחנו מקבלים שהדיוק הולך וקטן והעקומות מתכנסות לאותה נקודה של דיוק נמוך מאוד.
תופעה זו קורית מאחר וכאשר מספר השכנים גדול מדי התיוג של הנקודות מושפע גם משכנים רחוקים שלא רלוונטים לסיווג של אותה הנקודה. כתוצאה מכך, נוצר מצב שההשפעה של הנקודות הרלוונטיות (הקרובות יותר) יורדת ונקודות חסרות חשיבות נלקחות גם כן בחשבון. אם כמותן גדולה אז השפעתן יכולה להעפיל על המשקל של הנקודות הקרובות יותר ולייצר סיווג שגוי.
כתוצאה מכך, נוצר מצב של חוסר התאמה של המודל לtraining set וגם לtest set.
ככל שk גדל, המודל נעשה כללי מדי ואינו מסתגל היטב לניואנסים של קבוצת האימון. מכיוון שהוא לא מתאים, התחזיות של המודל אינן מושפעות במידה רבה מהמאפיינים של נתוני האימון, מה שמוביל לציוני דיוק דומים ונמוכים גם באימון וגם במבחן.

(Q3)




```
Training Accuracy: 0.866

Test Accuracy: 0.82
```

(Q4)

התוצאות של המודל 1NN בסעיף 1 הן :

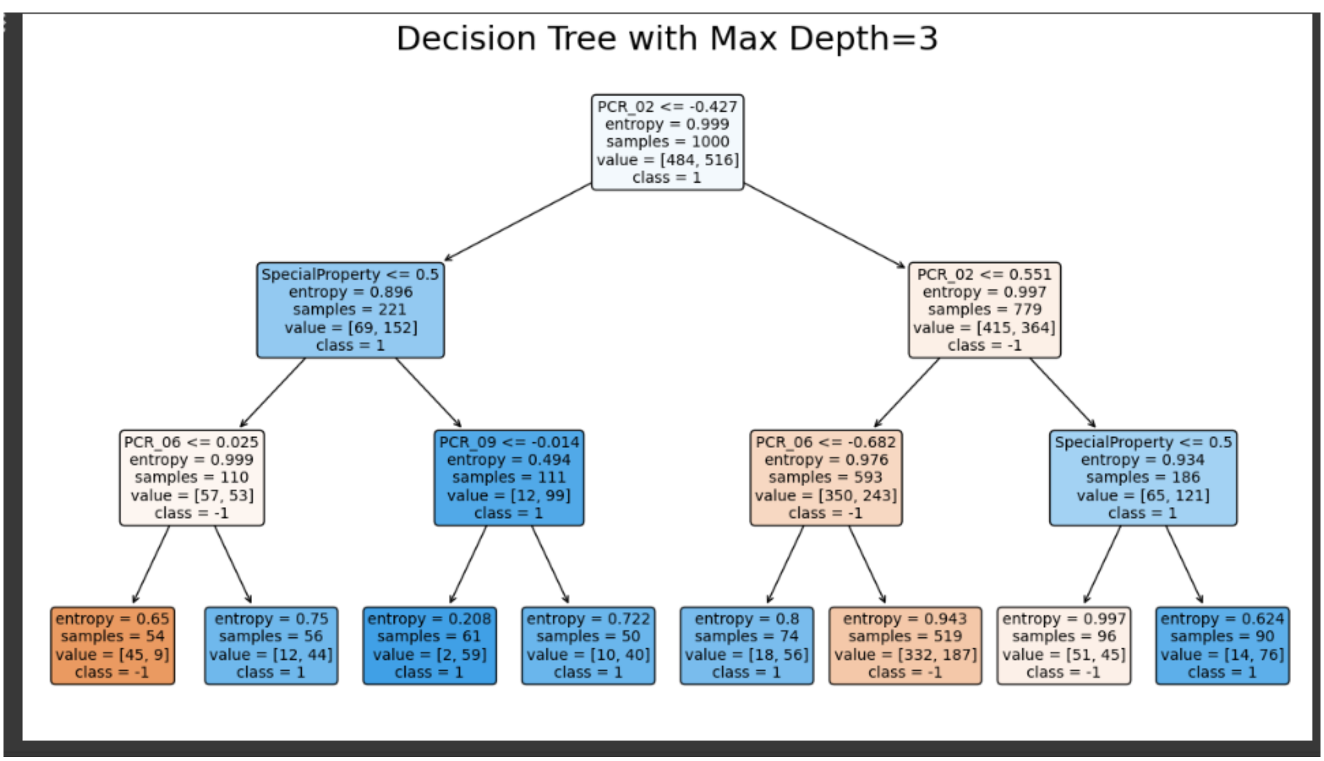
```
Training Accuracy: 1.0

Test Accuracy: 0.756
```

ניתן לראות שהדיוק של קבוצת האימון הוא גבוה בעוד שהדיוק של קבוצת המבחן נמוך. בנוסף גבולות ההחלטה של k=1 מתאימות באופן מושלם לקבוצת האימון. לכן, עבור k=1 מתרחשת תופעת הoverfitting, כלומר למודל אין יכולת הכללה מספיק טובה על קבוצות שונות מקבוצת האימון, והוא מתאים את עצמו לקבוצת האימון (מה שיצור שונות גבוהה של המודל על דאטה סטים שונים).

גבולות ההחלטה של המודל שלנו עם k=20 הם "חלקים יותר" ומצליחים לתפוס טוב יותר את ארבע אזורי ההחלטה העיקריים של הדאטה. בנוסף, אומנם מידת הדיוק על קבוצת האימון של המודל קטנה יותר אבל מידת הדיוק על קבוצת המבחן גבוהה יותר. ממצאים אלו מחזקים את הבחירה שלנו בפרמטר שמצאנו בתהליך הtuning ומצביעים על כך שמצאנו k באזור ה"sweet spot" שמאזן בין bias לvariance .

(Q5)

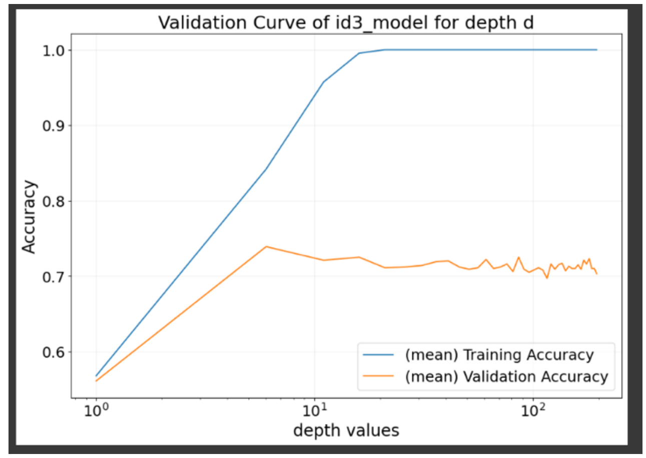
```
Training Accuracy: 0.703
```

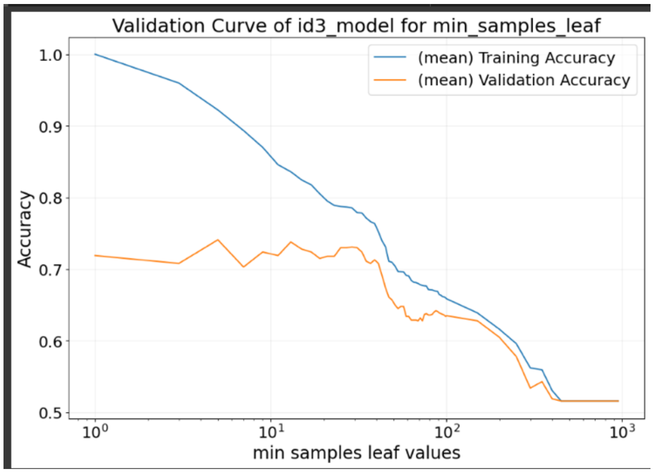


(Q6)

תחילה התייחסנו לפרמטרים כבלתי תלויים אחד בשני ובדקנו עבור כל פרמטר מתי מתקבל דיוק מקסימלי כפי שעשינו בשאלה 2 וזאת כדי להבין פחות או יותר הטווחים שכדאי להשתמש בהם.

התוצאה עבור max_depth, והתוצאה עבור min_samples_leaf:






```
highest validation accuracy : 0.739) and
training accuracy when the highest validation accuracy : 0.84175
for d : 6
```

```
highest validation accuracy : 0.741) and
training accuracy when the highest validation accuracy : 0.92225
for m : 5
```

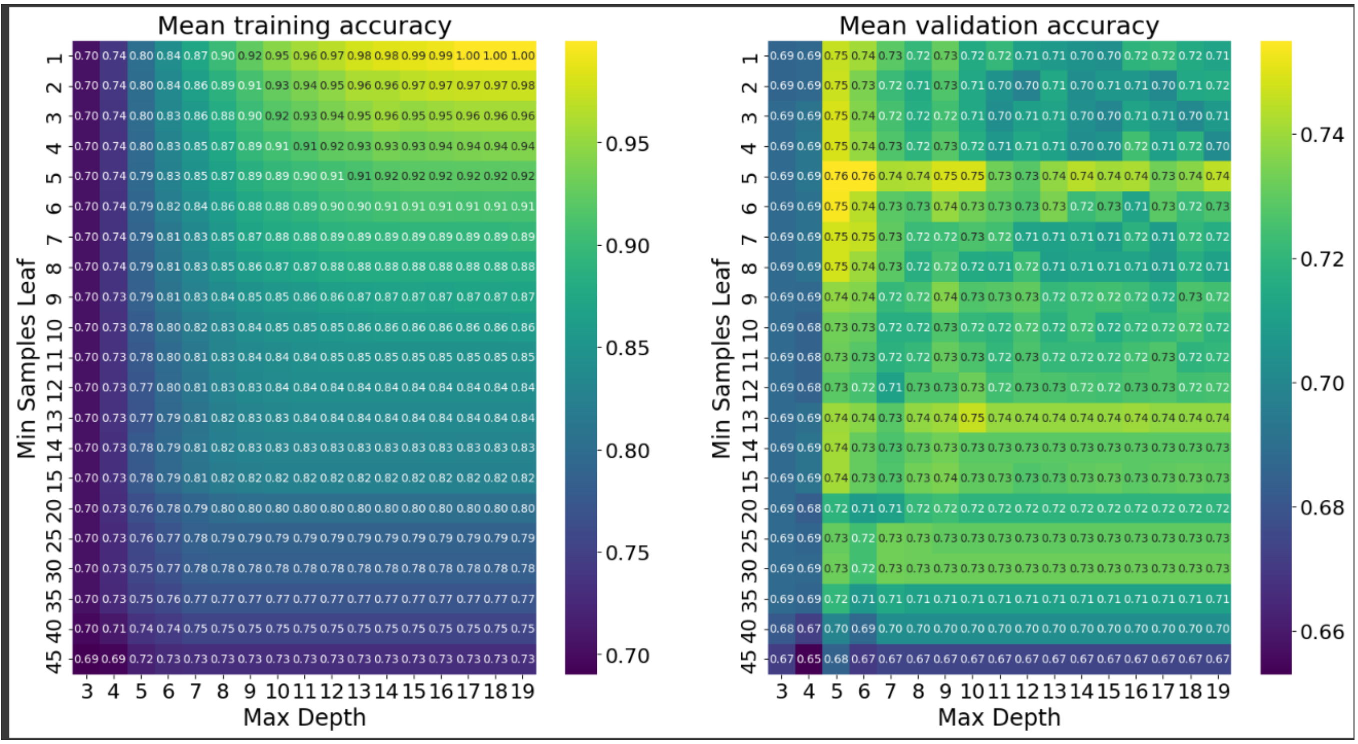
ניתן לראות כי עבור min_samples_leaf – תחילה יש הפרש גדול בין מידת הדיוק על קבוצת האימון לקבוצת המבחן (overfit) , לאחר מכן יש התכנסות של מידת הדיוק על האימון לכיוון מידת הוולידציה ובאזור של כמות דגימות גדולה מ-40 מתחיל להיות underfit (ירידה הן של הדיוק על האימון והן על הוולידציה).
עוד נשים לב כי עם העלייה בכמות הדגימות בעלה באזור 0-40 אין עליה משמעותית ביכולת ההכללה של המודל ובמידת הדיוק על הוולידיציה (למרות שמידת הדיוק על האימון יורדת ויוצאים מאזור הoverfit). עם זאת, נראה כי באזור של 5-6 דגימות בעלה ובאזור 13-14 דגימות בעלה , מקבלים מידת דיוק מעט גדולה יותר על קבוצת הולידיציה.

ניתן לראות עבור max_depth כי בעומקים נמוכים יכולת ההכללה של המודל היא נמוכה (מידת דיוק נמוכה הן על האימון והן על הוולידציה). ככל שעולים עד עומקים של כ5-7 מידת הדיוק על האימון והולידיציה עולות, כאשר לאחר עומקים אלו מגיעים לאזור של overfit- מידת הדיוק על האימון ממשיכה לעלות עד למידת דיוק מושלמת באזור ה-20 ואילו מידת הדיוק על הוולידיציה נשארת פחות או יותר יציבה.

א. לאחר החיפוש הראשוני בחרנו <u>בשני טווחים:</u>
עבור הmax-depth נראה שאנחנו יכולים לתפוס את ה"sweet spot" באזור 3-20 (אזורים של עומק 1-2 הם נמוכים מדי ואזורים מעל 20 הם כבר ממש באזור הoverfit).
עבור ה-min_samples leaf - בחרנו לעשות חיפוש מדויק יותר על הטווח של 1-15 וחיפוש מעט פחות מדויק על הטווח 15-45.

```
param_grid = {
    'max_depth': list(range(3, 20, 1)),
    'min_samples_leaf': list(range(1, 15, 1))+list(range(15, 50, 5))
    }
```

ב+ג.



קומבינציה אופטימלית:

```
{'max_depth': 6, 'min_samples_leaf': 5}
0.755
```

ד+ו.

<u>קומבינציה שגורמת לunder fitting:</u>

Max_depth=3

Min_samples_leaf=45

עבור קומבינציה זו קיבלנו מידת דיוק נמוכה הן על קבוצת אימון והן על קבוצת המבחן.

<u>סיבה לתופעת הunderfitting על קומבינציה זו-</u> התופעה מתרחשת כאשר למודל אין יכולת לתפוס את קשרים מורכבים בדאטה. ערכים מסוימים של היפר-פרמטרים מסוימים עשויים למנוע ממנו ללמוד ניואנסים מסוימים שקיימים בקבוצת האימון ולכן מקבלים מודל פשוט מדי שלא מתאים לדאטה. כאשר עומק העץ קטן, המודל לא יכול לבצע מספיק החלטות כדי לבצע פרדיקציה ולכן לא יכול לחלק את הדאטה למספיק קבוצות (למשל עץ בגובה 3 יחלק את הדאטה לכל היותר ל8 קבוצות). באותו אופן, כאשר המספר המינמלי שעלה יכול להחזיק הוא גדול , תהליך יצירת העץ יעצור גם כאשר האנטרופיה בעלה גבוהה יחסית וזה לא יאפשר לעץ לבצע התאמות נוספות הדרושות על מנת ללמוד קשרים עדינים בדאטה.

ה+ו.

<u>קומבינציה שגורמת לOverfit fitting:</u>

Max_depth=19

Min_samples_leaf=1

עבור קומבינציה זו קיבלנו מידת דיוק גבוהה על קבוצת אימון ומידת דיוק נמוכה על קבוצת המבחן.

<u>סיבה לתופעת הOverfitting על קומבינציה זו-</u>

תופעה זו מתרחשת כאשר המודל מתאים את עצמו בצורה יותר מדי טובה לקבוצת האימון, מה שפוגע ביכולת ההכללה שלו על חיזוי נקודות שהוא לא התאמן עליהם.

max_depth מאפשר לחלק את הדאטה לקבוצות יותר מדי קטנות (למשל עבור גובה עץ 19 ניתן לחלק את הדאטה ל2^{19} קבוצות. כתוצאה מכך, יש רגישות -יתר לרעש בדאטה והתאמה יותר מדי מושלמת לדאטה סט אחד , כך שיכולת ההכללה כבר נפגעת.

באופן דומה, מס׳ דגימות קטן בעלה, מאפשר למודל לייצר קבוצות אפילו עבור דגימות בודדות, שעשויות להיות רעש, ולכן גם במקרה נוצרת התאמת- יתר ופגיעה נוספת ביכולת ההכללה.

(Q7)

מספר הקומבינציות שנבדקו שווה למכפלת מספר הערכים האפשריים לכל פרמטר.

לפרמטר Max_depth הטווח הוא מ-3 ל-20 (לא כולל)ולכן יש 17 ערכים אפשריים.

לפרמטר Min_samples_leaf הטווח הוא מ1 ל15 בהפרשים של 1 ומ15-45 בהפרשים של 5 ולכן יש 21 ערכים אפשריים.

סה"כ יש <u>357 קומבינציות אפשריות</u> (21*17).

במידה והיינו מכווננות פרמטר נוסף, מספר הקומבינציות היה גדל פי מס׳ הערכים שהיינו בוחנות עבור פרמטר זה. למשל אם היינו בוחנות 10 ערכים אפשריים עבור הפרמטר השלישי , היינו מקבלות 3570 קומבינציות.

באופן כללי, בכל פעם שמוסיפים היפר -פרמטר נוסף למרחב החיפוש , מס׳ הקומבינציות שאנו בוחנים גדל באופן מעריכי (עם כל היפר-פרמטר מכפילים את מס׳ הקומבינציות במס׳ האפשרויות עבור ההיפר פרמטר החדש). לכן, הוספת פרמטר נוסף תגדיל את העלות החישובית ואת הזמן הנדרש למציאת השילוב האופטימלי.

(Q8)

לאחר אימון המודל עם ההיפר פרמטר האופטימליים שמצאנו (,min samples leaf=5 ,(max_depth=6

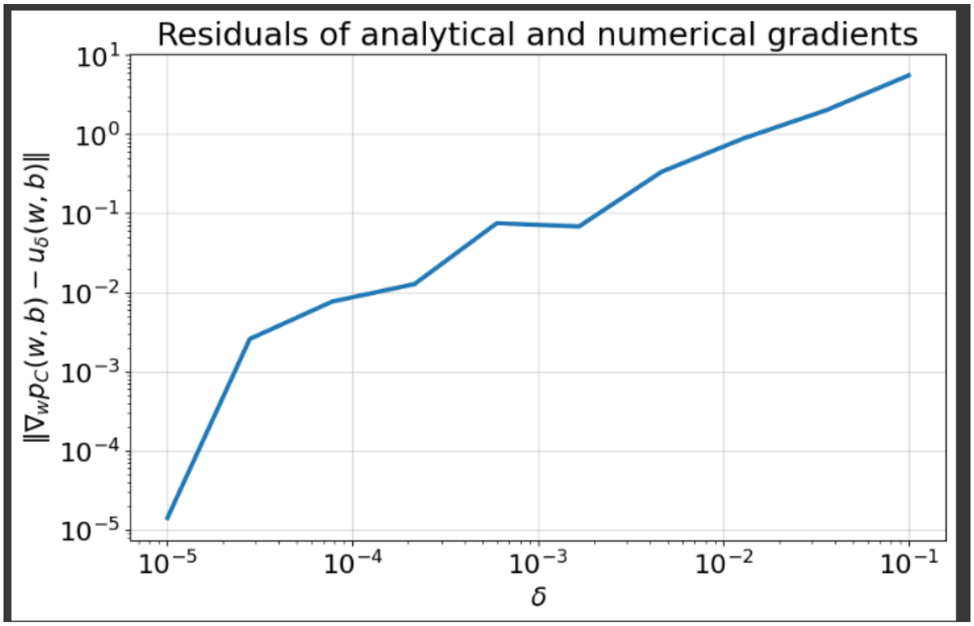
קיבלנו את התוצאות הבאות:

```
Test Accuracy: 0.772
```

(Q9)

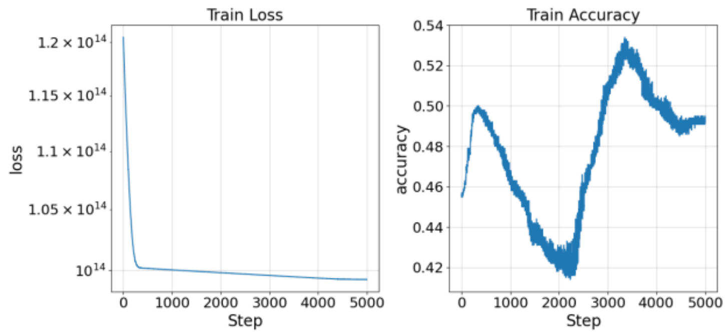
הגדרת הנגזרת כפי שלמדנו בקורסים קודמים משאיפה את δ לאפס היא $\lim_{\delta\to 0}\frac{f(x_0+\delta)-f(x_0)}{\delta}$.
החישוב הנומרי שהוצג לנו הוא קירוב לנגזרת החלקית לפי w_i . ככל ש δ קטן נצפה שהחישוב הנומרי יהיה יותר מדויק ויתקרב לנגזרת האמיתית ולכן החישוב האנליטי והנומרי צפויים להיות קרובים יותר.

העקומה הבאה מראה ששתי השיטות של החישוב מתקרבות ככל ש δ קטן, מה שמצביע על נכונות המימוש של הגראדיאנט שלנו.



ככל ש-δ גדל אנחנו רואים שההפרש בין הגראדיאנט האנליטי והנומרי גדל . תופעה זאת קורית, כיוון ככל ש-δ גדל -> הקירוב הנומרי מתרחק מהגרדיאנט האמיתי ולכן התוצאה דומה למה שהיינו מצפים.
נשים לב שאנחנו מקבלים בכל הרצה תשובות שונות כי אנחנו מחשבים את הגרדיאנט על w,b שנבחרים באופן רנדומלי ובודקים את הערכים שמתקבלים ולכן בכל פעם יכולים להתקבל מעט ערכים שונים.
מה שחשוב שבכל ההרצות אנו מקבלים שככל δ קטן כך שהרנדומליות של w, b לא משפיעה ואנחנו מקבלים קירוב טוב בין הגרדיאנט שלנו האנליטי לבין הנומרי.
ייתכן גם שיש רעשים מהחילוק במספרים אינפינטימליים במהלך תהליך הקירוב אבל אנחנו מסתכלים על ההתנהגות הכללית של העקומות שקיבלנו בשביל לאמת את החישוב האנליטי שלנו לגרדיאנט - ככל δ קטן הresiduals קטן.

(Q10)



נשים לב כי אנו מנסים לפתור בעיית soft svm תחת התנאים הבאים:

- ע״פ הגרף המצורף לשאלה, הדאטה לא נראה פריד לינארית ולא נראה כי ניתן להשיג מידת דיוק גבוהה על הדאטה עם מפריד לינארי.
- ההיפר פרמטרים המצורפים לשאלה הם C שהוא יחסית גדול מאד, וlearning rate יחסית קטן וקבוע.

ניתן לראות כי במהלך תהליך האימון, ע״פ התרשים, ערך פונקציית הloss הלך וירד ,כאשר בהתחלה הייתה ירידה חדה ולאחר מכן הירידה איטית יותר ומתכנסת לערך מסויים.

ציפינו שירידה זו תקרה במהלך האימון כיוון שלבעיית הsoft-svm קיים מינימום גלובאלי כפי שראינו בהרצאה ובתרגולים, ולכן ע״י שימוש באלגוריתם sgd ערך הפונקציה צפוי לנוע לכיוון הערך המינימלי במהלך האימון (עם זאת, זה לא מובטח כי פונק׳ המטרה לא גזירה וlr קבוע). ה learning rate שאנו משתמשים בו יחסית קטן, ולכן הירידה איטית יחסית וגם אין התבדרות.

מאחר שC גדול מאוד, ע״פ הגדרת בעיית האופטימיזציה, תינתן עדיפות לצמצום פונק׳ הhinge על פני מציאת מפריד עם margin גדול. עם זאת, מאחר והדאטה אינו פריד לינארית, ניתן לראות כי ערך פונ׳ הloss נשאר יחסית גבוה בסוף האימון, כלומר ערך פונקציית הhinge לא הצליח להגיע לערך מאד נמוך ולכן גם הloss נשאר גבוה.

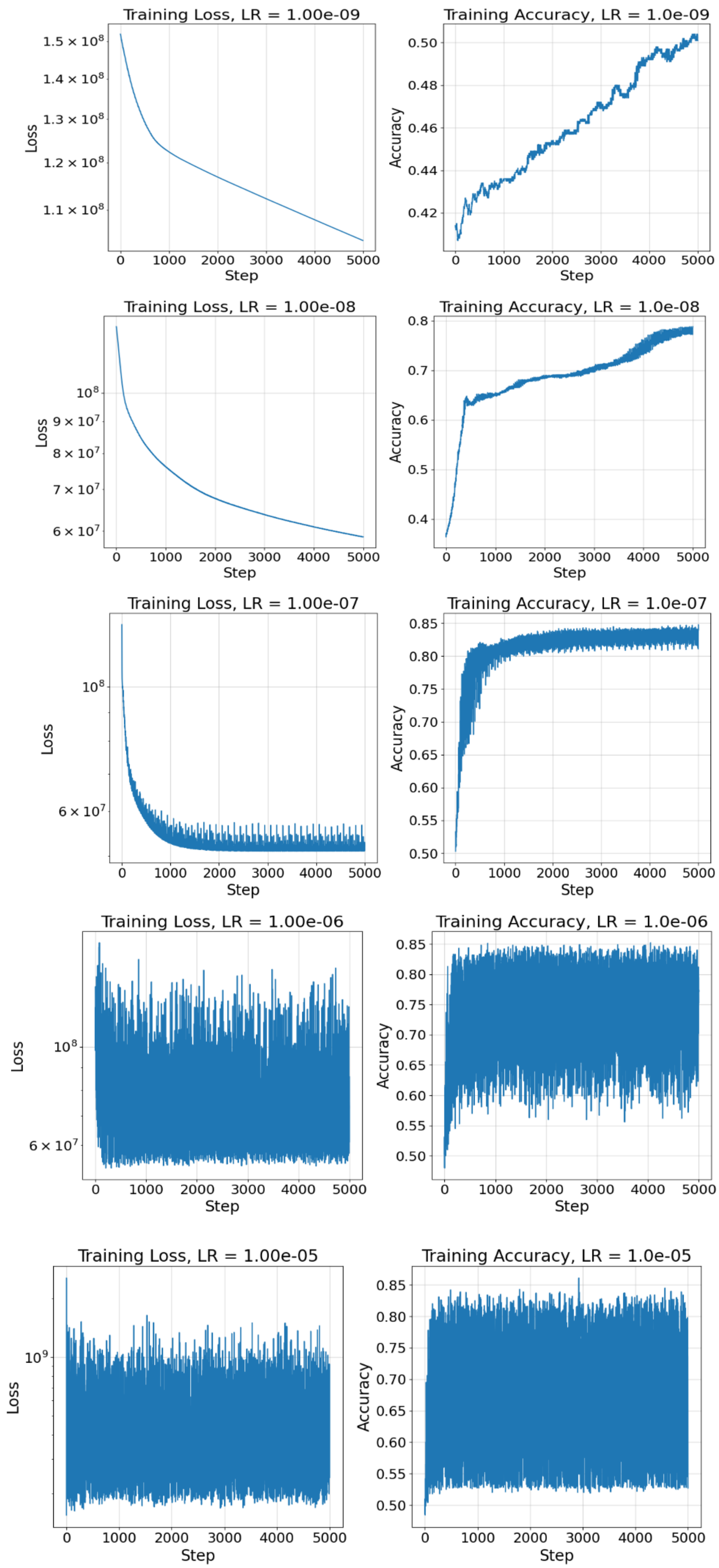
מבחינת הaccuracy , ניתן לראות כי במהלך האימון היו עליות וירידות ולא הייתה התכנסות למידת דיוק כלשהי. נזכיר כי הaccuracy הוא המשלים של loss 0-1 .

כאשר אנו פותרים בעיית soft-svm אנו כן מקווים שפתרון לבעיית svm יצליח להביא גם לשיפור במידת הדיוק. עם זאת, פונקציית המטרה בבעיית הsoft-svm מכילה את פונקציית הhinge ולא את פונקציית הloss 0-1. לכן, פתרון של בעיית soft-svm אינו מבטיח פתרון אופטימלי עבור loss 0-1.

עוד נזכיר כי ראינו בהרצאה, שפונק׳ הhinge היא קירוב של פונק׳ הloss 0-1 , אך כאשר אנו מסתכלים על אזורים בהם הנקודות כבר בצד הלא נכון, הפונקציות האלה כבר לא מאוד קרובות כיוון שפונק׳ הhinge מענישה את הנקודות שלא מתויגות נכון בעונש כבד יותר ככל שהן רחוקות יותר מהmargin.

כאשר אנו מחפשים את נק׳ המינימום של בעיית הsoft-svm ייתכן מצב שבו כדי להקטין את הloss , עדיף פשוט לנסות למצוא מפריד שיותר קרוב גם לנקודות שאינן מתויגות נכון מאשר לנסות להגדיל את כמות הנקודות באימון שמתויגות נכון. לכן, במצב של הדאטה שלנו, שהוא אינו פריד, ויש הרבה דגימות שיכולות להיות רחוקות מכל מפריד אופציונאלי, פונק׳ הhinge היא למעשה קירוב כבר לא כ״כ טוב ל-0 loss 1 . לכן ,תופעת התנודות של הaccuracy שראינו בתרשים היא הגיונית, תחת התנאים שהזכרנו במהלך הדיון.

(Q11)


Training Loss, LR = 1.00e-09
Loss
Step
Training Accuracy, LR = 1.0e-09
Accuracy
Step
Training Loss, LR = 1.00e-08
Loss
Step
Training Accuracy, LR = 1.0e-08
Accuracy
Step
Training Loss, LR = 1.00e-07
Loss
Step
Training Accuracy, LR = 1.0e-07
Accuracy
Step
Training Loss, LR = 1.00e-06
Loss
Step
Training Accuracy, LR = 1.0e-06
Accuracy
Step
Training Loss, LR = 1.00e-05
Loss
Step
Training Accuracy, LR = 1.0e-05
Accuracy
Step

על בסיס. התרשימים החלטנו לבחור ב$learning\ rate = 10^{-7}$.

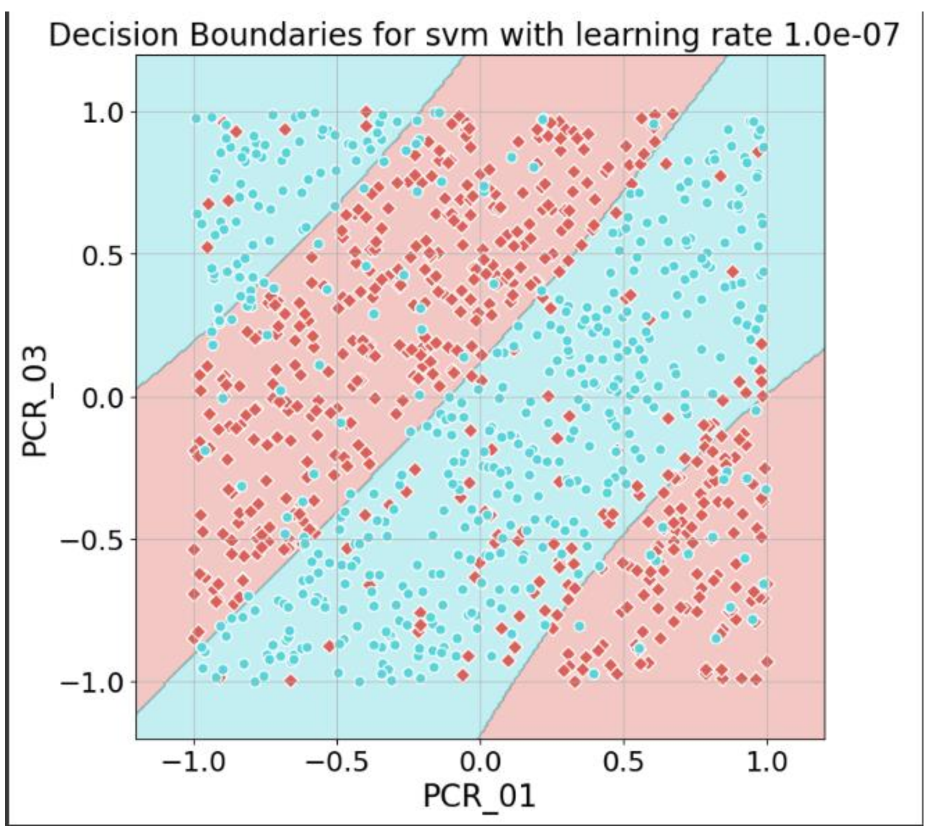
ראשית, עבור $learning\ rate = 10^{-5}, learning\ rate = 10^{-6}$ אין התכנסות בכלל של פונק׳ הloss, אלא היא מתבדרת לגמרי ולכן הlearning rate גדול מדי.

בנוסף, נראה כי $learning\ rate = 10^{-9}$ הוא קטן מדי עבור מס׳ הsteps. כן יש ירידה בפונקציית הloss אבל היא איטית וניתן לראות שבתום האימון ערכה עדיין גדול לעומת מינימומים אחרים. בנוסף, ה accuracy רק 0.5 בסוף האימון, בהשוואה לlearning rates אחרים שמידת הדיוק שם היתה יחסית גדולה (כמובן, ע״פ מה שרשמנו קודם שאין התחייבות שתהיה שיפור במידת הדיוק, כי זה לא הגורם לו אנחנו מחפשים מינימום אבל כן נרצה לבחור learning rate שמגדיל אותו במידת האפשר).

ההתלבטות שלנו הייתה בין הקצבים $learning\ rate = 10^{-8}, learning\ rate = 10^{-7}$ כיוון ששניהם מראים ירידה גדולה ביותר בloss ומראים שיפור במידת הaccuracy.

החלטנו לבחור בסופו של דבר ב$learning\ rate = 10^{-7}$, על אף שהוא מעט רועש יותר, מאחר ונראה כי הוא מביא את הloss לערך מינמלי יותר ב5000 צעדים, וגם הaccuracy מגיע לדיוק מעט גבוה יותר.

(Q12)




```
training accuracy: 0.833
test accuracy: 0.784
```

(Q13)

א. הוכחה:

- $\forall x$: $||x||_2 \le c_1$ לפי ההנחה בשאלה.

אם $||x_i||_2$ ו- $||x||_2$ חסומים אז $||x-x_i||_2$ חסום.

$\exists c_1 : ||x_i||_2 \le c_1, ||x||_2 \le c_1,$

לפי אי שויון המשולש מתקיים :

$$||x-x_i||_2 \le ||x||_2 + ||x_i||_2 \le 2c_1$$

לכן $\lim_{\gamma\to 0} \exp(-\gamma||x-x_i||_2^2) = 1,$.

- לפי ההנחה בשאלה, ניתן לבצע $\lim(sign) = sign(\lim)$ (ניתן לבצע זאת כי הפונקציה בתוך החסם sign היא רציפה פרט לנקודה 0).

לכן נקבל,

$$\lim_{\gamma\to 0} sign\left(\sum_{i\in[m],\alpha_i>0} \alpha_i y_i \exp(-\gamma||x-x_i||_2^2)\right)$$
$$= sign\left(\lim_{\gamma\to 0} \sum_{i\in[m],\alpha_i>0} \alpha_i y_i \exp(-\gamma||x-x_i||_2^2)\right)$$
$$\overset{(*)}{\triangleq} sign\left(\sum_{i\in[m],\alpha_i>0} \alpha_i y_i\right) \overset{\alpha_i>0}{\triangleq} sign\left(\sum_{\{i|y_i=1\}} \alpha_i - \sum_{\{i|y_i=-1\}} \alpha_i\right)$$
$$\overset{\text{הגדרת } sign}{\triangleq} \begin{cases} +1, & \sum_{\{i|y_i=1\}} \alpha_i > \sum_{\{i|y_i=-1\}} \alpha_i \\ 0, & \sum_{\{i|y_i=1\}} \alpha_i = \sum_{\{i|y_i=-1\}} \alpha_i \\ -1, & \sum_{\{i|y_i=1\}} \alpha_i < \sum_{\{i|y_i=-1\}} \alpha_i \end{cases}$$
$$\overset{(**)}{\triangleq} argmax_{y\in\{-1,1\}} \sum_{\{i|y_i=y\}} \alpha_i$$

(*) הנורמה של α חסומה, לכן הגבול קיים.

(**) אם sign מחזיר 1+ זה אומר ש $\sum_{\{i|y_i=1\}} \alpha_i > \sum_{\{i|y_i=-1\}} \alpha_i$ כלומר המקסימום על הביטוי $\sum_{\{i|y_i=y\}} \alpha_i$ יתקבל כאשר $y=1$ ולכן $argmax_{y\in\{-1,1\}} \sum_{\{i|y_i=y\}} \alpha_i$ יחזיר 1.

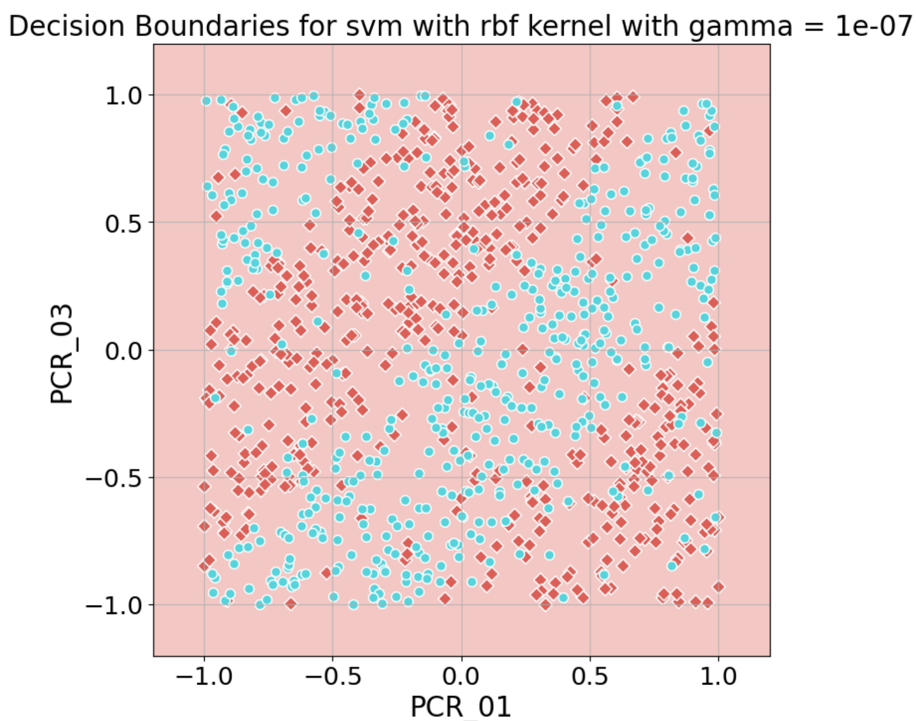
באותו אופן, sign יחזיר 1- אם $\sum_{\{i|y_i=1\}} \alpha_i < \sum_{\{i|y_i=-1\}} \alpha_i$ כלומר המקסימום על הביטוי $\sum_{\{i|y_i=y\}} \alpha_i$ יתקבל כאשר $y=-1$ ולכן $argmax_{y\in\{-1,1\}} \sum_{\{i|y_i=y\}} \alpha_i$ יחזיר 1-.

(ניתן להניח, לפי הפיאצה, ש(sign(0 הוא מחזיר את הערך 1 ולא אפס ובנוסף שאם הסכום של המקדמים השליליים שווה לסכום המקדמים החיוביים הסיווג יהיה חיובי ומכאן שכן יש שוויון)

ב. הוכחנו קודם כי כאשר $\gamma \to 0$ אנחנו מקבלים שכלל הפרדיקציה הוא $argmax_{y\in\{1,-1\}} \sum_{\{i|y_i=y\}} a_i$ ובמקרה שלנו כלל ההחלטה הוא $argmax_{y\in\{1,-1\}} \sum_{\{i|y_i=y\}} 1$ הפרדיקציה של כל נקודה נקבעת על פי תיוג של רוב הנקודות בדאטה סט. נשים לב, שככל שגמא שואף ל0 כמעט ואין חשיבות למרחק של הנקודה x משאר הדוגמאות ולכן ככל שגמא קטן לכיוון האפס אנחנו מקבלים מודל שמחזיר פרדיקציה קבועה (1- או 1 על פי תיוג הרוב שיש בדאטה סט). אפשר גם להגיד שככל שגמא שואף ל0 אנחנו מקבלים מודל knn שההחלטה שלו מתקבלת על פי m שכנים (m דוגמאות בקבוצת האימון), כאשר שככל שגמא קטן ככה "מס השכנים" מתקרב לm.

(Q14)

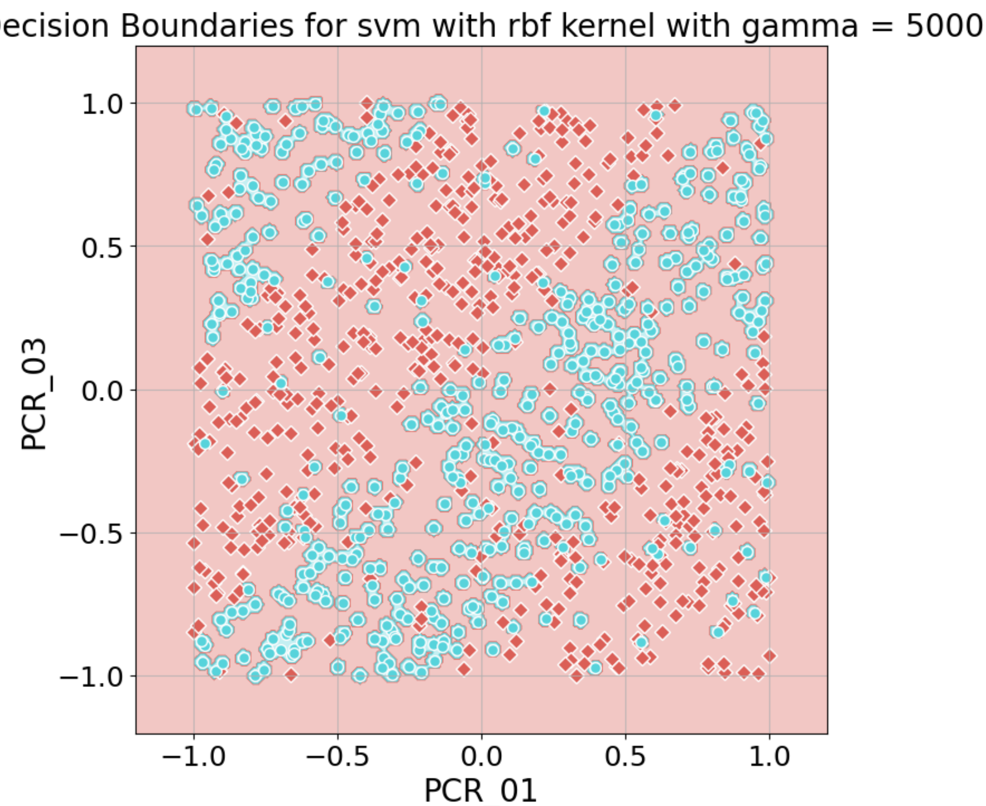
נשים לב שγ מאוד קטן ואכן גבולות ההחלטה של המודל מתאימים לכלל ההחלטה הנידון בסעיף 13 ב, כפי שניתן לראות קיבלנו את המסווג ה"טריוואלי" שפולט פרידקציה קבועה לכל הנקודות.(1-)




```
training accuracy: 0.51
test accuracy: 0.544
```

(Q15)

```
training accuracy: 0.999
test accuracy: 0.688
```



נשים לב כי המודל שקיבלנו דומה למודל חח-1 בכך שהוא מתאים את עצמו לקבוצת האימון בצורה מושלמת (יוצר סביבות קטנות סביב נק׳ חריגות). על פי כלל הפרידקציה, מאחר וגמא גדול מאד, נקבל שההשפעה של נק׳ support vector מקבוצת האימון תבוא לידי ביטוי רק במידה והנקודה שרוצים לחזות את התיוג שלה קרובה אליה מאד. לכן, במובן הזה, המודל אמור להיות דומה מאד למודל knn.

עם זאת, ניתן לראות כי לעומת מודל החknn, המודל הזה נותן פרדיקציה קבועה (במקרה שלנו הצבע האדום מייצג פרדיקציה 1-) לרוב הנקודות, ויודע לתת פרדיקציה 1 רק שהנקודה נמצאת ממש קרוב לנקודה עם תיוג 1 . כלומר המודל לא יודע לתת פרדיקציה 1 לנקודה שהשכן הקרוב שלה הוא כחול כאשר המרחק בין הנקודות לא מאד מאד קרוב.

אנחנו חושבות שתופעה זו קורית מאחר וכאשר גם גמא גדול וגם המרחק גדול נקבל כי $e^{-\gamma||x-x_i||}$ הוא כמעט 0, ומבחינה נומרית -> המחשב עשוי להתייחס אליו כאפס.

במצב כזה, כאשר לנקודה אין אף שכן מספיק קרוב מבחינה נומרית, הפרידקציה שלה נקבעת לפי גורם הb. כפי שניתן לראות כלל החלטה של המודל שאימנו לפי הדוקומנטציה של sklearn הוא:

Once the optimization problem is solved, the output of decision_function for a given sample x becomes:

$$\sum_{i \in SV} y_i \alpha_i K(x_i, x) + b,$$

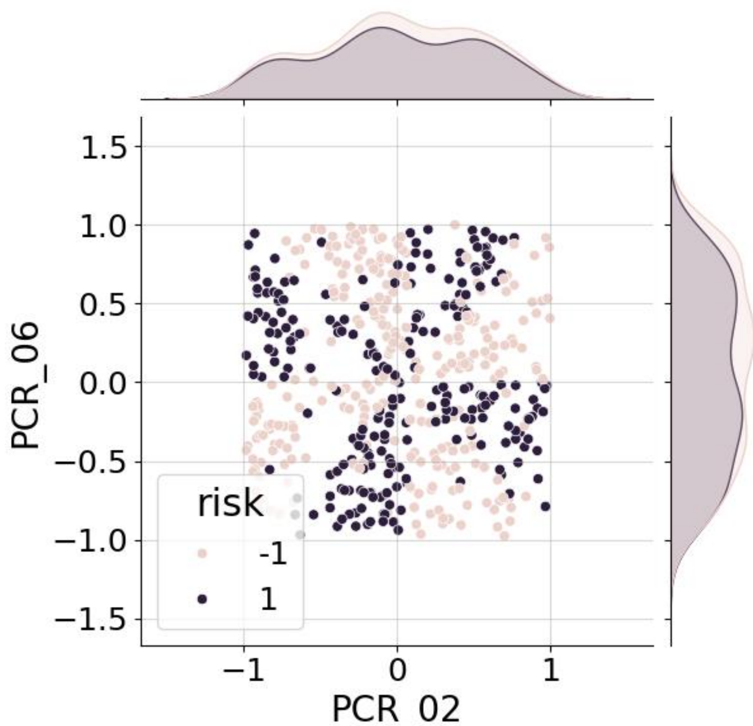
and the predicted class correspond to its sign. We only need to sum over the support vectors (i.e. the sam-

במקרה שלנו הדפסנו וראינו כי : `b=[-0.03094057]`

ולכן הגיוני שהחיזוי שלנו במקרים של נקודות ללא שכנים מספיק קרובים יוצא שלילי (שכן החיזוי נקבע ע״פ הסימן של b שהוא שלילי).

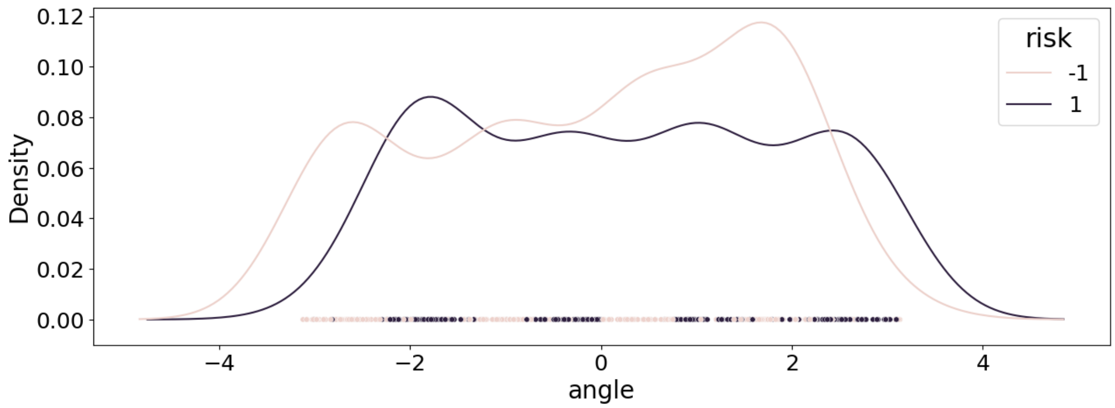
(Q16)

PCR_02 vs. PCR_06 With special property=0 colored by risk



על בסיס התרשים, ניתן לראות שבדאטה יש תבנית של 4 קבוצות (הקבוצות נוצרות הרביעים שנגזרים מהצירים x=0, y=0) כך שכל קבוצה היא כמעט פרידה לינארית. ניתן לראות שעבור כל רביע, בערך בזווית 45 מעלות התיוג של הדאטה משתנה.

(Q17)



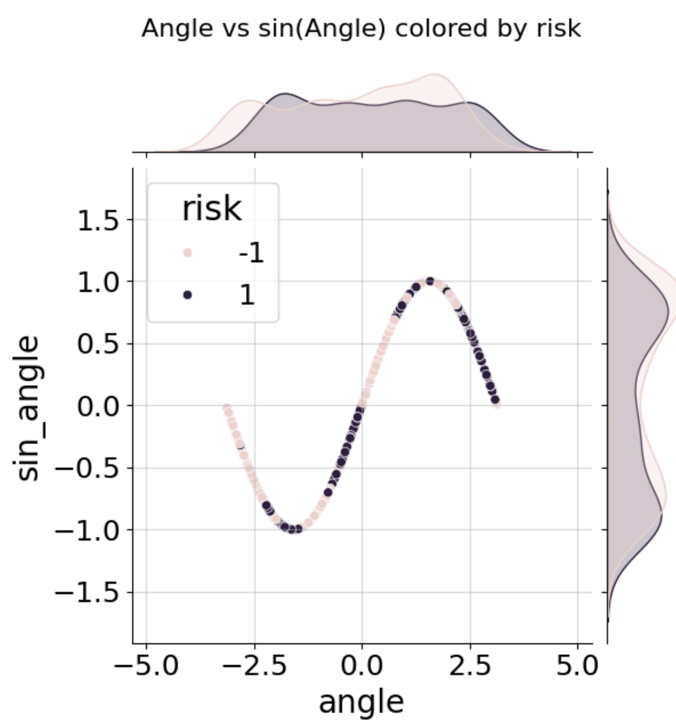
בתרשים ניתן לראות כי יצירת הפיצ׳ר החדש של הזווית , מחלקת את הדאטה (לא בצורה מושלמת) כך שניתן לחלקו לטווחים של תיוגים מסוימים במרווחים של $\frac{\pi}{4}$ רדיאנים (כפי שראינו ב plot הראשוני).

<u>הדאטה אינו פריד לינארית (לא ניתן להפרידו ל2 קבוצות בלבד באמצעות מפריד לינארי).</u>

עוד ניתן לראות ע״פ פונקציות הצפיפות, שסה״כ משתנה הזווית מתפלג יוניפורמית על פני הטווח פאי למינוס פאי, כלומר הזוויות שהווקטורים שלנו יוצרים נעים סביב כל הטווח בין [0,360] . בנוסף ניתן לראות שבטווחים מסוימים יש אכן יותר סיכוי לקבל מחלקה -1 על פי פונקציית הצפיפות (בהתאם לטווח) וכן גם עבור המחלקה 1.

(Q18)

על פי התרשים הפיצ׳רים עדיין אינם פרידים לינארית:



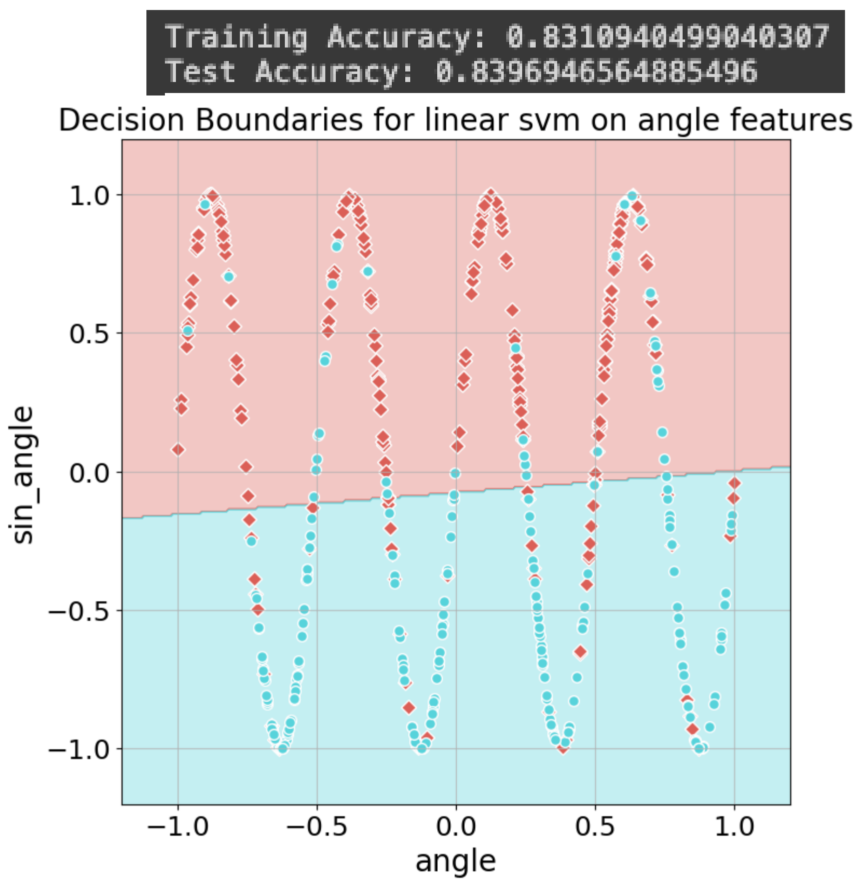
(Q19)

ראינו קודם שניתן להפריד את הדאטה לחלוקה לטווחים, כל שבערך כל $\frac{\pi}{4}$ רדיאנים מקבלים מחלקה אחרת. אנחנו יודעים שפונ׳ הסינוס היא פונ׳ מחזורית הנעה בין 1 ל-1-, לכן נרצה לצמצם את המחזור שלה כך שנקבל שהפונקציה משלימה מחזור שלם ב-$\frac{\pi}{2}$ רדיאנים. כך אכן נקבל שכל $\frac{\pi}{4}$ רדיאנים יש נק׳ קיצון והסימן של הפונקציה משתנה. כך נוכל לדאוג שבטווחים של מחלקה risk=1 ערך הסינוס יהיה בעל סימן מסוים ובטווחים של מחלקה risk=-1 ערך הסינוס יהיה בעל ערך אחר, מה שיצור הפרדה לינארית כפי שאנו רוצים.

לצורך כך חשבנו להגדיר למשל $\beta = 4$ ואז נקבל שהמחזור של פונק׳ סינוס משתנה כפי שרצינו:

$$sin(4x) = sin(4x + 2\pi) = sin(4\left(x + \frac{\pi}{2}\right))$$

אימנו מודל על הדאטה וקיבלנו את המודל הבא:



נציין כי ביצענו נרמול באמצעות min max scaling בין 1 ל 1- למשתנה angle כי ראינו קודם שמתפלג יוניפורמית , ואת המשתנה sin(angle) לא נרמלנו כי ערכיו של סינוס כבר חסומים

ניתן לראות כי למודל שלנו ביצועים טובים ביחס למודל הקודם. למודל קודם היה אחוז דיוק נמוך על קבוצת האימון וקבוצת הטסט, מה שמצביע על underfit ועל כך שההיפותזה של מודל לינארי לא התאימה לפיצ׳רים הקודמים. כעת, לאחר שיישמנו את הפיצ׳רים החדשים, ניתן לראות שהדאטה יחסית

יותר פריד לינארית, ואכן קיבלנו אחוז דיוק של כ0.83 שזו עלייה משמעותית ביחס לקודם (ואפילו לא כווננו היפר פרמטרים במקרה זה).