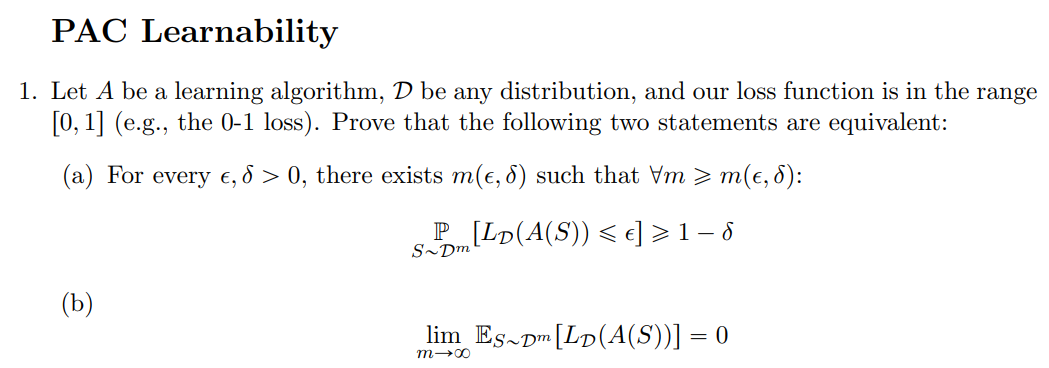
מבוא למערכות לומדות – תרגיל 4



1. ראשית נראה כי :

נניח כי b נכון, נסדר מחדש את הביטוי ב-a ונפעיל עליו א"ש מרקוב:

*ומכיוון שהנחנו את מתקיים:*

מכיוון שההסתברות יכולה להיות בין 0 ל-1 בלבד.  
נהפוך את אי השוויון בתוך ההסתברות ונקבל:

*כעת אנו יודעים כי ולכן ונקבל:*

*כנדרש.*

*נראה כעת את הגרירה* :

מהנחת אנו יודעים כי לכל קיים כך שלכל מתקיים

ומכך שיכלנו לבחור כל נוכל לבחור את ואז מתקיים:

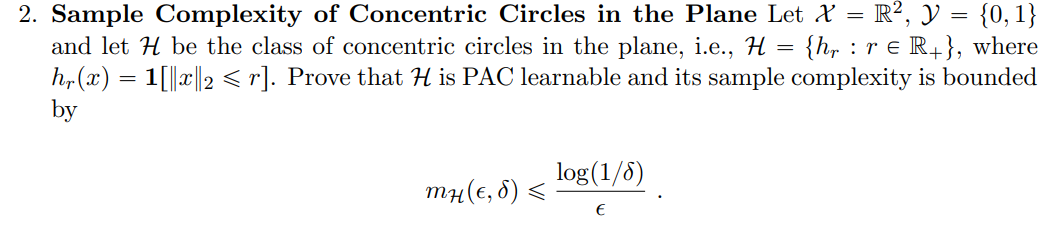
*כעת נתחשב בעובדה כי הוא מ"מ חסום בין 0 ל-1 ונתחם את האינטגרל בהתאם:*

נציב את ההנחה ב- ונקבל:

*כעת נסתכל על גבול התוחלת כאשר :*

*ניזכר בעובדה כי מדור על תוחלת של מ"מ אי-שלילי ולכן תוחלת זו חסומה ע"י 0, ומכלל הסנדוויץ' נקבל:*

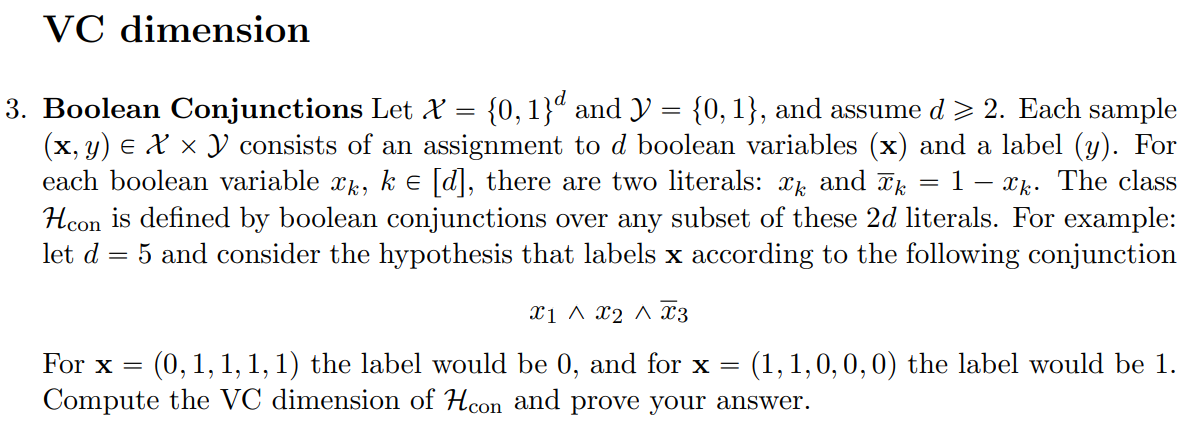
*כנדרש.*



1. ראשית נשים לב בשאלה כי מחלקת ההיפותזות היא כל המעגלים הקנוניים סביב הראשית, וככזאת יש מעגל סביב הראשית שתופס בתוכו את כל הנקודות החיוביות, ונראה שנוכל תמיד למצוא מעגל כזה עבור m דגימות כך שהוא יטעה לכל היותר ב- בסבירות של לכל הפחות.  
   נניח שרדיוס המעגל האמיתי (ההיפותזה הנכונה) הוא , כעת, בקבלינו m דגימות נשרטט רדיוס חדש שיכיל בתוכו את כל הדגימות החיוביות מתוך m הדגימות שלנו, כלומר הנקודה "הרחוקה" ביותר מהראשית תתווה את היקף המעגל הפנימי (בעל רדיוס ) שהוא ההיפותזה אותה אנו מציעים. נראה איזה מספר של דגימות נצטרך בשביל להבטיח שהמעגל שלנו יטעה לכל היותר ב-.

נבחר מעגל כזה שהרצועה בינו לבין המעגל האמיתי היא בשטח , נשים לב שכל הנקודות שבתוך המעגל הזה הן בהכרח חיוביות כי הן בתוך מעגל ההיפותזה , וכל מה שמחוץ למעגל בהכרח מתויג כשלילי, לכן הטעויות שיכולות לקרות לנו הן אך ורק ברצועה בין שני המעגלים.   
ההסתברות שנקודה תיפול מחוץ לרצועה היא , ומכיוון שכל m הנקודות ב"ת ההסתברות שכולן יפלו בתוך המעגל הפנימי היא .  
נשתמש בזהות ונקבל:

*ומכאן:*



1. ראשית נראה ש-   
   עבור היפותזה עם ליטרים בהכרח יהיה ליטרל כלשהוא כך שבהיפותזה הוא יופיע פעמיים

כעת בצורה כזאת, בהכרח לא משנה איזו השמה של הליטרלים נבחר תמיד נקבל שהתגית שלנו 0, מכיוון ש- ובעצם לא נוכל לקבל אף פעם תגית 1 ולא הצלחנו להגיע לכל האפשרויות ולנפץ את הדוגמא.

נראה שכל סט דוגמאות מגודל d ומטה נוכל לנתץ.

נניח שיש לנו d דגימות, ונניח שכל הדוגמאות הן וקטור היחידה המתאים למיקום שלהן (כלומר הווקטור יהיה הדגימה ה-i)

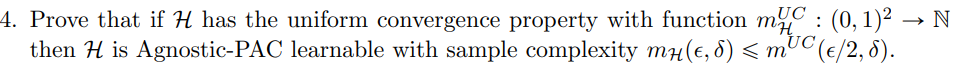
כעת נרצה להראות שנוכל להגיע לכל התגיות האפשריות נסדר אותן לפי סדר, לדוגמא פה כאשר :

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *y* | | | | | | | | *X* | | |
| *1* | *0* | *1* | *1* | *0* | *0* | *1* | *0* | *1* | *0* | *0* |
| *1* | *1* | *0* | *1* | *0* | *1* | *0* | *0* | *0* | *1* | *0* |
| *1* | *1* | *1* | *0* | *1* | *0* | *0* | *0* | *0* | *0* | *1* |

*כעת, נוכל להראות שאין אף תגית y שלא נוכל להגיע אליה עם היפותזה באמצעות וקטורי היחידה שלנו.*

* *נניח כי כל התגיות שוות ל-0, נוכל להגיע לזאת באמצעות בחירת ההיפותזה "נבחר תמיד 0"*
* *אם יש לנו תגית אחת בלבד שהיא "1", נגיד תגית i, נוכל לבחור בהיפותזה , כל שאר עמודות יסתדרו גם כן כי בהן וגם התגית בהן שווה 0*
* *נניח יש לנו שתי תגיות "1", נוכל לבחור (בדוגמא שלנו) את ההיפותזה עבור y המסומן באדום*
* *באופן כללי אם יש לנו תגיות מסומנות ב-1, נניח בה"כ כי אלה תגיות , נוכל תמיד לבחור את ההיפותזה*
* *ואם כל התגיות הן 1 נבחר בהיפותזה "תמיד 1"*

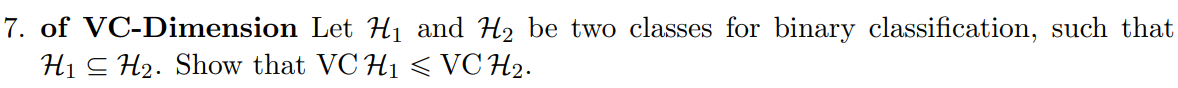
*ובעצם הראינו שנוכל להגיע לכל תוצאה עם d וקטורים וניפצנו כל קבוצה מגודל d.  
עבור כל קבוצה C קטנה מגודל d, נוכל לנתץ אותה באותו אופן, כאשר נתעלם כליל מ- ונפעל על השאר באותה דרך.*



1. נניח כי ל- יש את התכונה של התכנסות במידה שווה כמתואר, אזי, לכל , ו-S קבוצת דגימות בגודל m, כך ש- , מתקיים שבהסתברות שלכל הפחות S היא  -מייצגת, כלומר, לכל מתקיים:

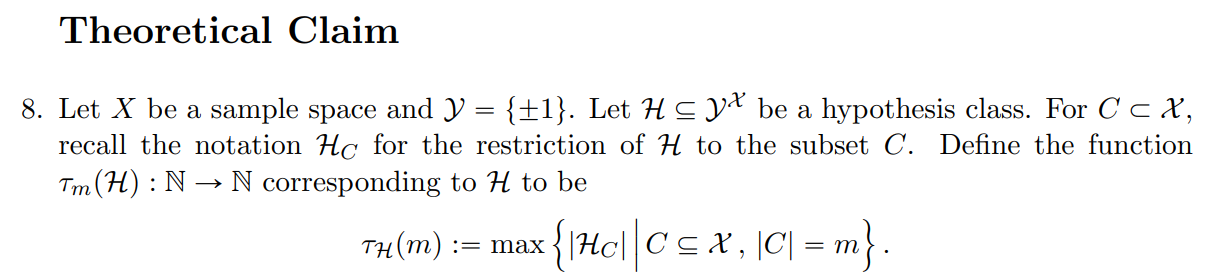
כלומר

ולכן היא עם .



1. נסתכל על ההגדרה:

כלומר המימד VC נקבע ע"י גודל תת-הקבוצה המקסימלית שמחלקת ההיפותזות מצליחה לנפץ.  
כעת, בהנחה ש- נניח בשלילה כי , אבל מכיוון ש- , ומחלקת ההיפותזות מנתצת תת-קבוצה בגודל C שכזאת, על אותו מרחב דגימות גם תצליח לנתץ קבוצה באותו הגודל, לפחות ע"י צמצום שלה למחלקה בלבד, וזו בסתירה ל- .

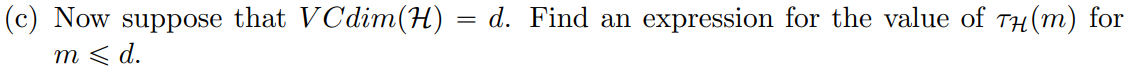




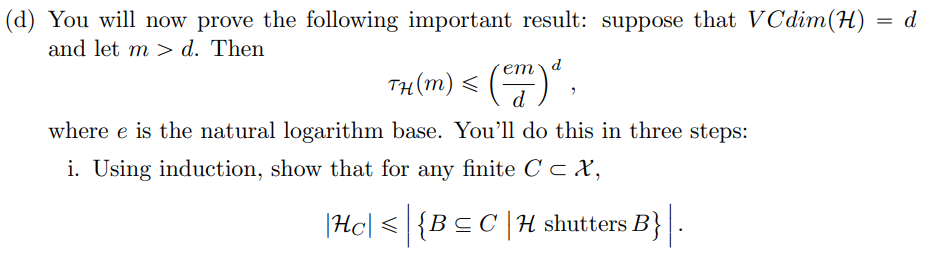
1. טאו היא הגודל המקסימלי של מחלקת ההיפותזות המצומצמת לתת-קבוצה C בגודל m, שמצליחה לנפץ את אותה תת-קבוצה. כלומר, נניח והצטמצמנו לקבוצה C כלשהי בגודל m, וישנן קבוצות כאלה, טאו מהווה את הגודל המקסימלי של מחלקת ההיפותזות שמצליחה לנפץ מי מהקבוצות האלה.



1. *כאשר יש לנו מחלקת היפותזות בעלת זה אומר שכל צמצום לקבוצה C בגודל m נקבל כי היא מכילה את כל האפשרויות לפונקציות , כלומר עבור כל m , ומכיוון שטאו הוא הגודל המקסימלי של קבוצת היפותזות כזאת, נקבל כי , ומכיוון ש- נקבל כי .*



1. *באותו אופן טאו מסמן את הגודל המקסימלי של קבוצת ההיפותזות, ומכיוון שעבור כל מחלקת ההיפותזות מנפצת את קבוצה C נקבל כי .*



* 1. *נרצה להשתמש באינדוקציה להוכיח טענה זו.*

*ראשית, עבור , תמיד יוכל לנפץ כל קבוצה שנבחר, וגודל שני צדדי א"ש תמיד יהיה 1.  
כעת, נניח עבור ונוכיח עבור סט דגימות בגודל .  
נקבע את מחלקת ההיפותזות שלנו ואת תת הקבוצה של הדגימות , ונגדיר תת קבוצה חדשה של דגימות ללא הדגימה הראשונה .  
כעת נוכל לקבוע את הקבוצות הבאות של תוצאות אפשריות להיפותזות:*

*במילים, קבוצה אלה כל ההיפותזות האפשריות ל- שהיינו יכולים להתאים להן קורדינאטה ראשונה 0 או 1 ועדיין הייתה היפותזה כלשהי ב- שהייתה מנבאת את התוצאה הזאת בכללותה.  
קבוצה אלה רק ההיפותזות האפשריות ל-, כך שעבור מסוים גם הקורדינאטה הראשונה 1 וגם 0 מתקיימות והיו מנובאות ע"י היפותזה כלשהי ב-.*

*כלומר, נוכל להבין כי , וגם . נפעיל את הנחת האינדוקציה שלנו על ו- ונקבל*

*כעת נגדיר את מחלקת ההיפותזות :*

*כלומר, כל ההיפותזות שמסכימות על כל הקורדינאטות מלבד הראשונה, ניתן היה לכתוב גם כך: .  
כעת קל לראות כי כל היפותזה ב- שהייתה מנפצת תת קבוצה הייתה מנפצת גם את תת הקבוצה , וזאת מכיוון שההיפותזה לא משתמשת בקורדינאטה של לצורך הניתוץ, ולכן מה שיהיה רשום בא לא משנה.*

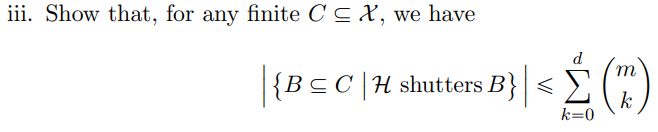
*כעת נפעיל את הנחת האינדוקציה שלנו על ו- , ונתחשב בעובדה כי ונקבל:*

*סה"כ קיבלנו:*

*כנדרש.*

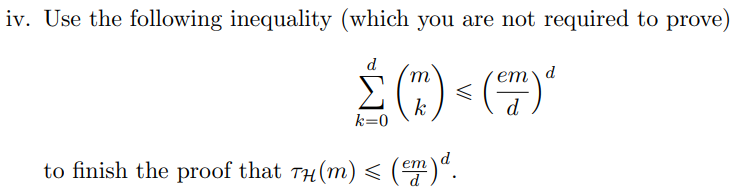


1. *אי השוויון גורס כי מחלקת ההיפותזות תמיד תהיה קטנה ממספר תתי-הקבוצות B אותה היא מסוגלת לנפץ, כלומר, תמיד תהיה היפותזה שיכולה לנתץ מספר תתי קבוצות שונות, שזו תוצאה מעניינת כי היא מרמזת שאימון על תת קבוצה B מסוימת היה יכול לתת לנו לחזות במדויקת את התוצאות על תת קבוצה אחרת. ומנגד, ישנן תתי קבוצות ב-C שאינן מוסיפות לנו מידע חדש על גבי מידע שכבר קיבלנו מתת קבוצה אחרת.*



1. *אי השווין הנ"ל מציג מצידו הימיני את מספר כל תתי הקבוצות האפשריות של , בכל גודל אפשרי מ-0 ועד כל*

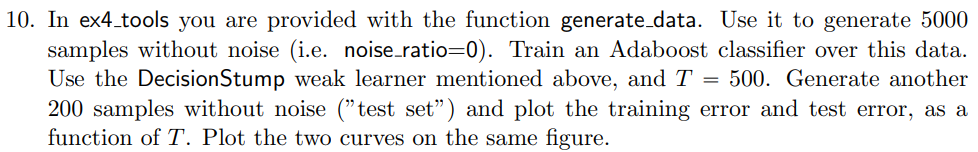
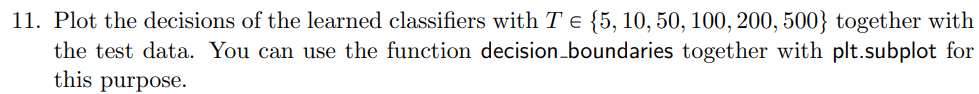
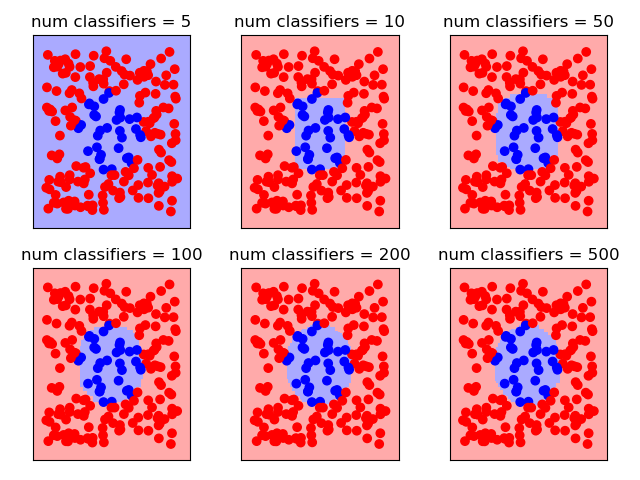
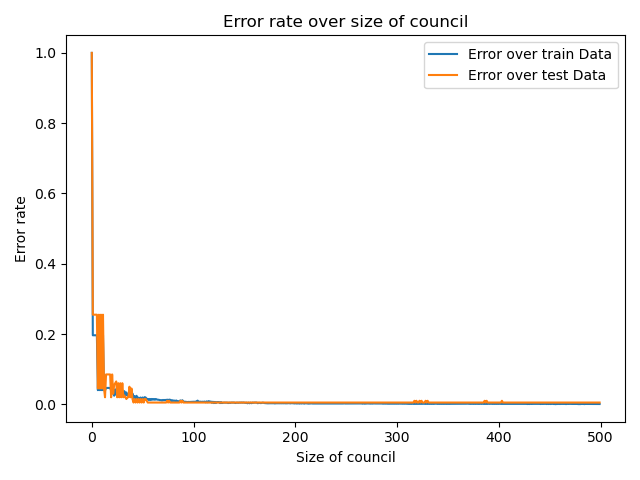
*הקבוצה כולה, מספר זה מבוטא בבינום כפי שהוא רשום. לעומת זאת, בצד השמאלי של הביטוי יש לנו את מספר כל תתי הקבוצות ש- מצליח לנתץ, בהנחה והוא מצליח לנתץ את כל תתי הקבוצות לא משנה מאיזה גודל נקבל שוויון, אחרת, אם יש תתי קבוצות שהוא לא מצליח לנפץ נקבל את אי השוויון.*

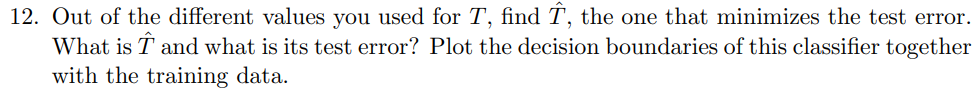


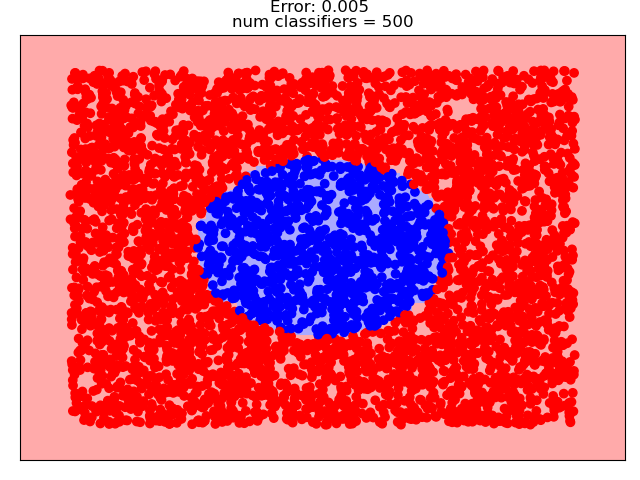
1. *לפי ההגדרה של טאו:*

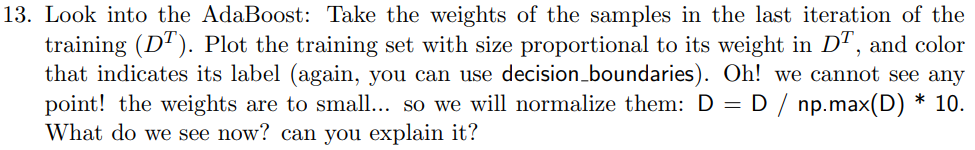


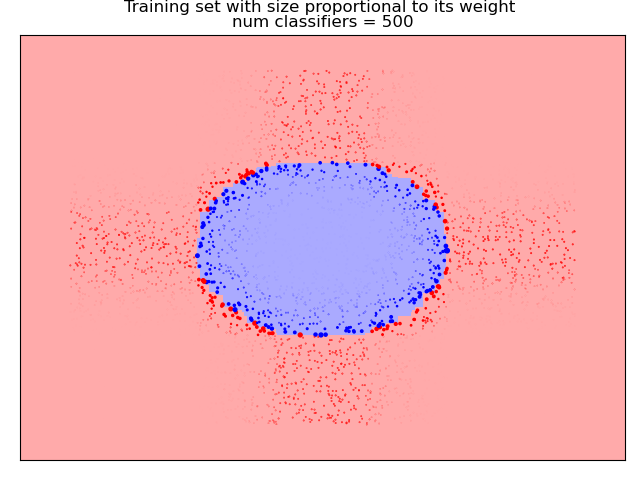
1. *נניח כי , אזי הביטוי באי שוויון ניתן לכתיבה כ- , כעת, ניתן להבין שכאשר ככל ש-d יהיה יותר גדול ככה החסם יהיה יותר הדוק, אך כאשר , כלומר כמו בשאלה נקבל שא"ש הוא , זה אמנם עדיין חסם לטאו, אך הוא פחות הדוק.*



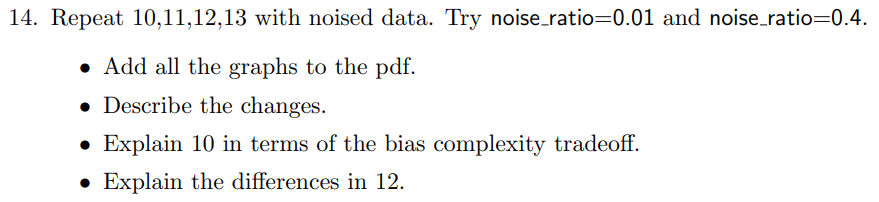


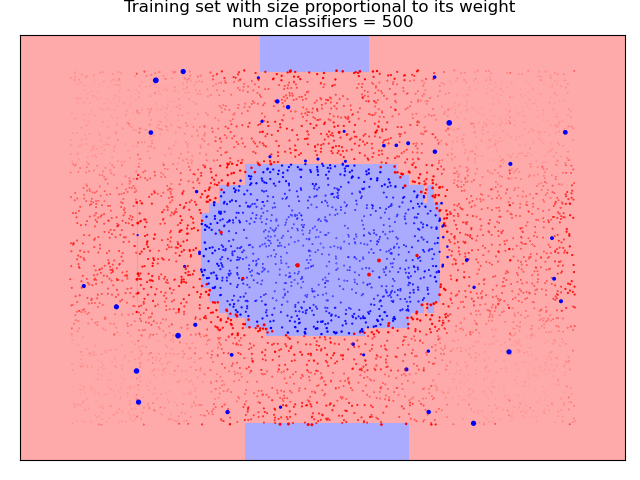
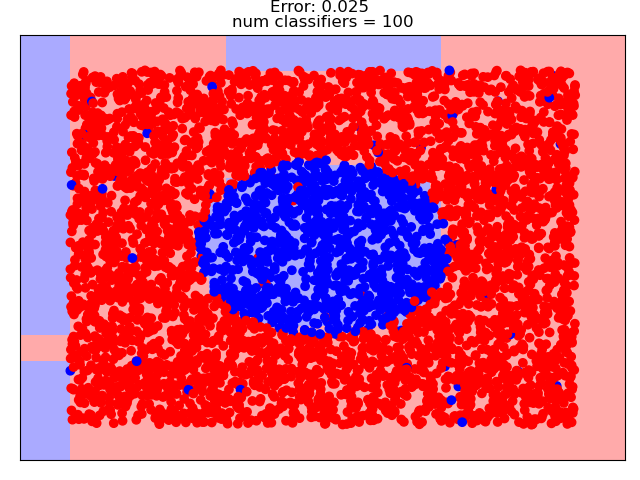
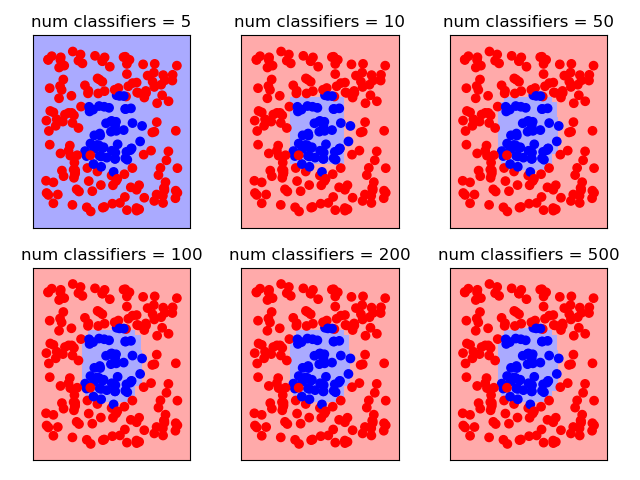
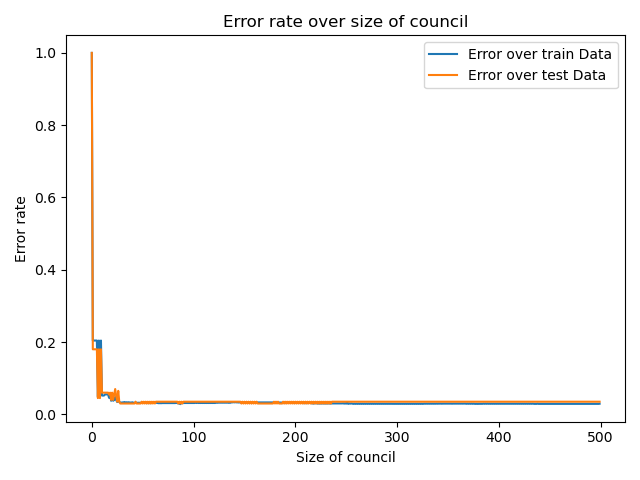




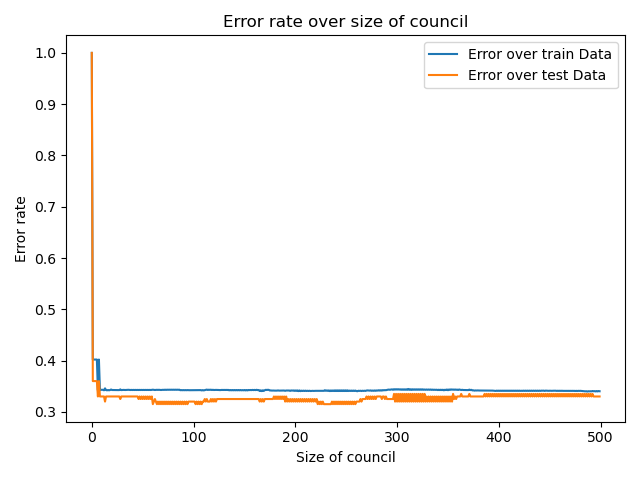
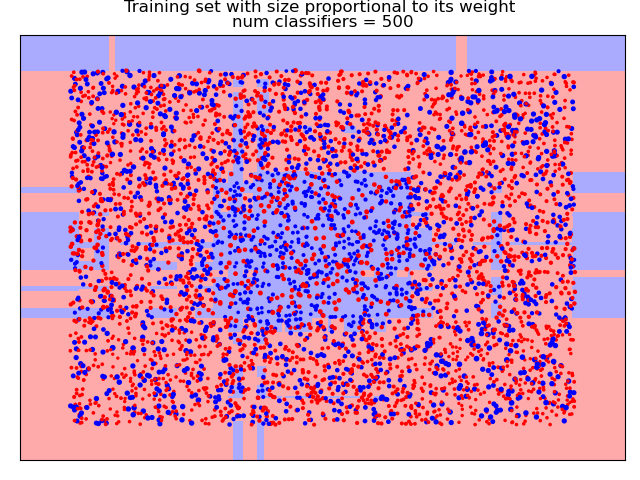
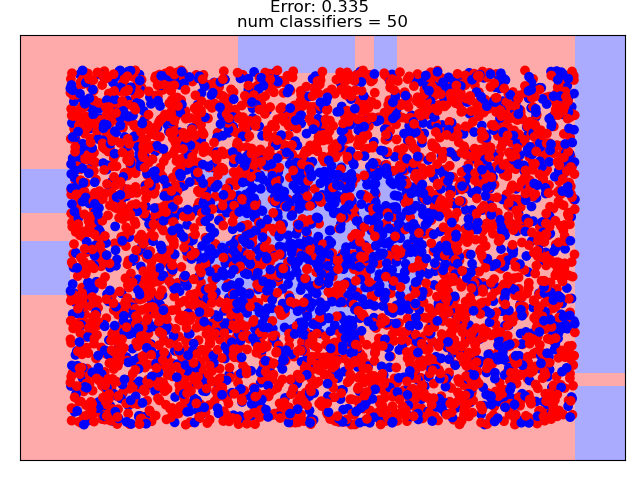
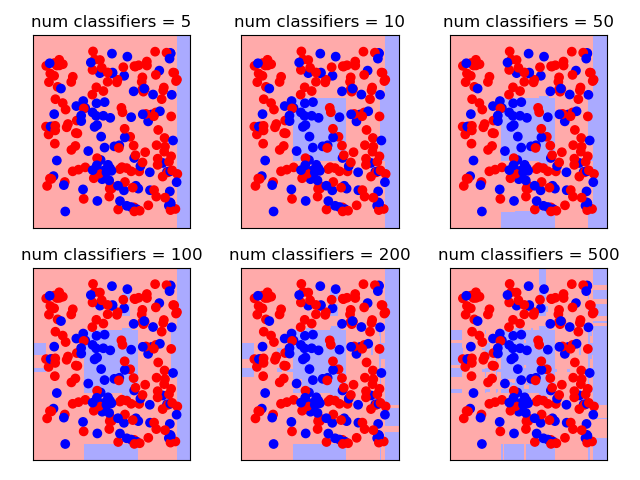


1. ניתן לראות שבשלב הזה כמעט ואין נקודות בכל ההיקף, הן כן שם אבל עם כל פעם שצדקנו בסיווג שלהן הן הלכו וקטנו, וכעת בסיבוב האחרון צדקנו בהן כל כך הרבה עד שהן כבר זניחות.  
   בנוסף ניתן לראות כי הנקודות הגדולות ביותר, גם הכחולות וגם האדומות, נמצאות בהיקף של המעגל משני צדדיו, זו תוצאה שהיינו יכולים לצפות כי זה האזורים בהם הסיווג הכי עדין, ובהם טעינו הכי הרבה פעמים, ניתן לראות שככל שמתקרבים למרכז הכחול הנקודות הולכות וקטנות, זה אזור שכמעט תמיד צדקנו בו לאורך זמן.

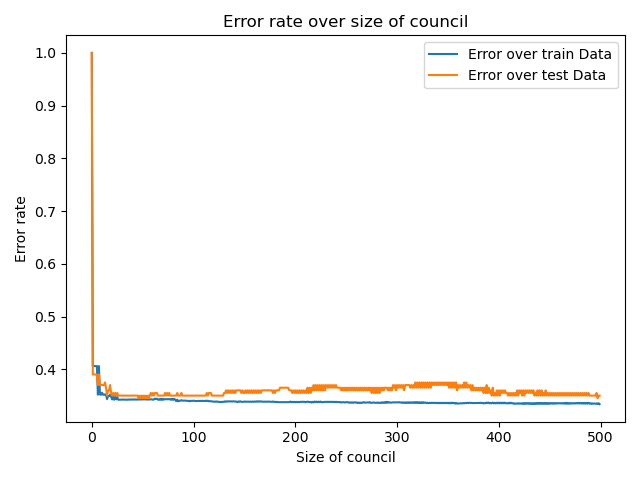




**Noise = 0.001**



**Noise = 0.4**



1. ניתן לראות שינויים רבים כתלות ברמת הרעש בדאטה, ניתן להבחין ככלל אצבע שככל שיש לנו יותר רעש יש יותר אזורי חלוקה רבים יותר אותם ה- מנסה להכליל, כמו כן כצפוי הטעות המינימלית במדגם הרועש יותר גבוהה מהטעות במדגם הרועש פחות, ואולי התוצאה המעניינת ביותר בעיניי היא שניתן לראות בגרף השמאלי התחתון, המייצג את גודל הנקודה כמספר הפעמים שסיווגנו אותה נכונה, כי הנקודות על המדגם הרועש יותר הן באופן כללי גדולות יותר מאלה שבמדגם השקט, הדגמה וויזואלית לאיך שהרעש מפריע לנו לסווג נכון נקודות לאורך זמן.

ניתן לראות כי ההבדלים ב-12 (גרף ימני תחתון) הם שכשהרעש גדול יותר יש יותר נקודות מפוזרות שהמודל מנסה לתפוס, זה גורם לשברור של אזורי הסיווג וליצירה של הרבה יותר אזורים ספורדיים כאשר הרעש גבוהה יותר. ניתן בנוסף לראות שהדגם הרועש יותר הוא בעל יחס טעות גבוהה יותר, כצפוי.