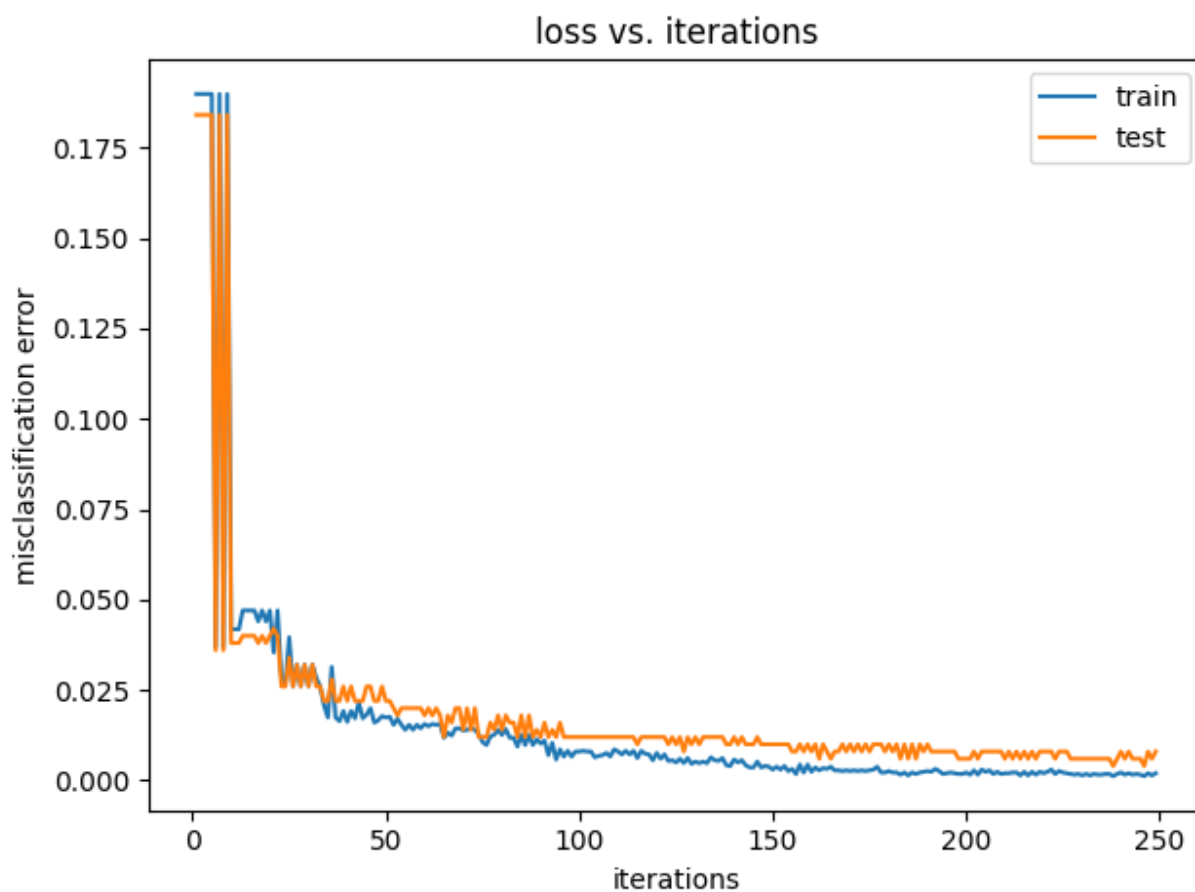


מערכות לומדות תרגיל 4 חלק מעשי:

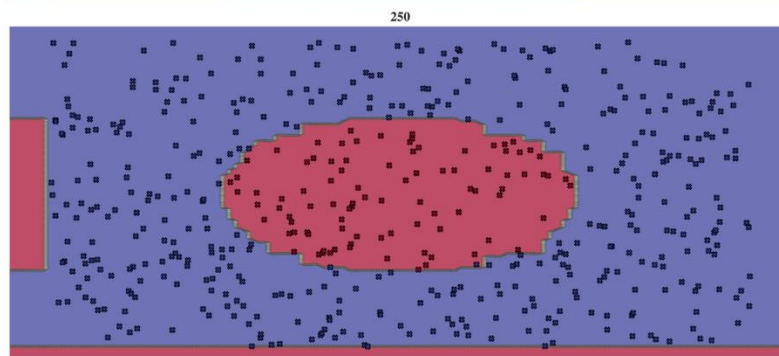
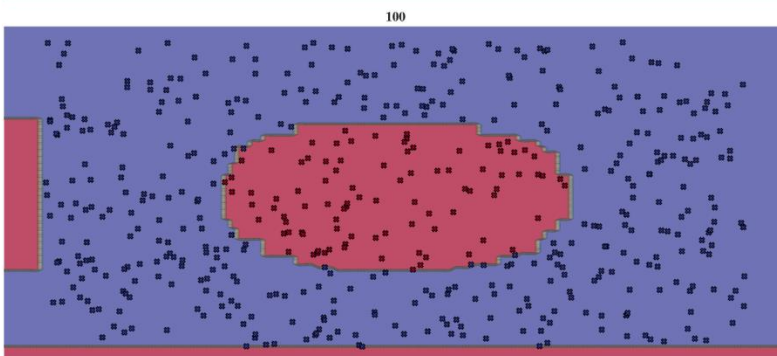
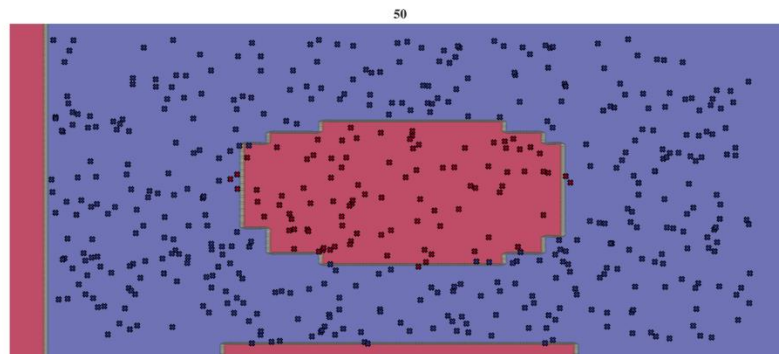
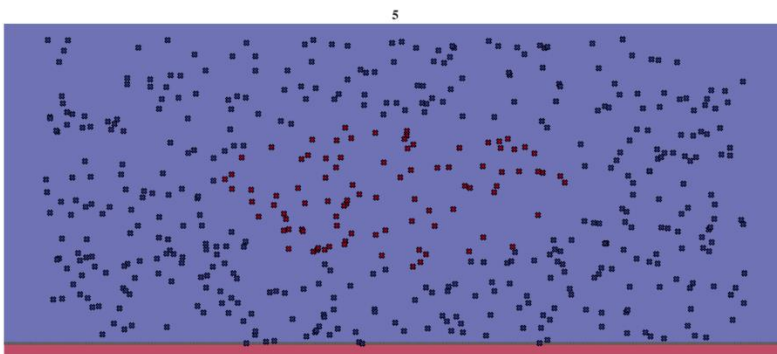
שאלה 1



קיבלנו שעבור רעש מועט test והtrain מתנהגים דומה והשגיאה שלהם נהיית אפסית. אסיק שהמודל חזק ומשתפר עם עליית כמות האיטרציות. בנוסף המודל לא עושה overfit כי הtrain והtest משתפרים ביחד.

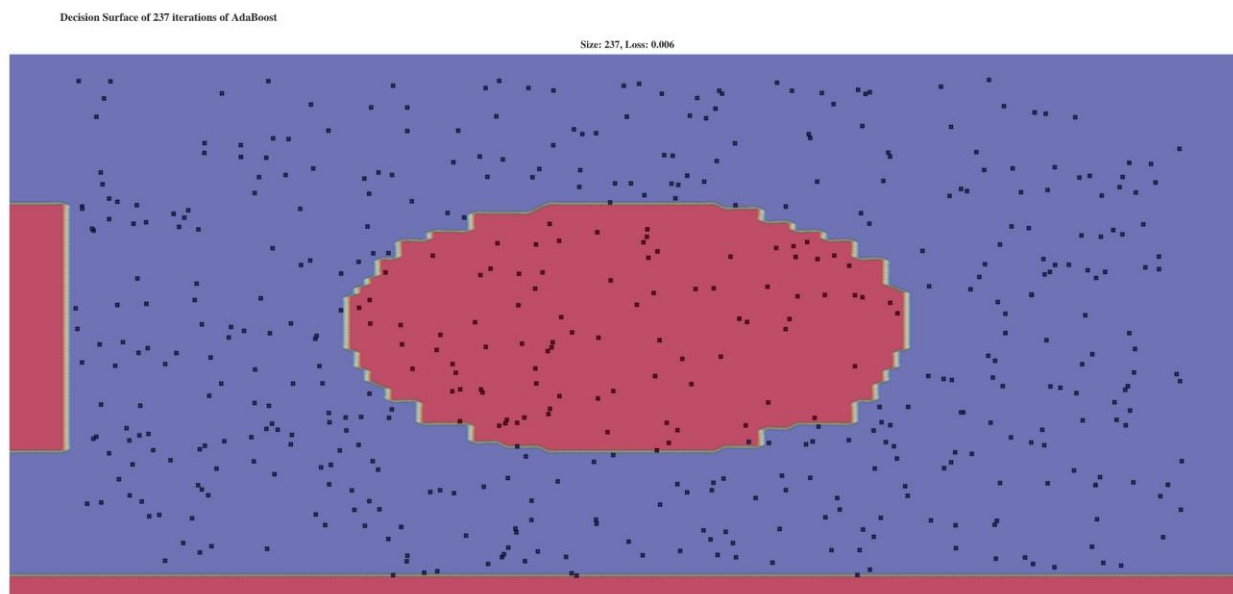
שאלה 2

Decision Surface of 5,50,100,250 iterations of AdaBoost



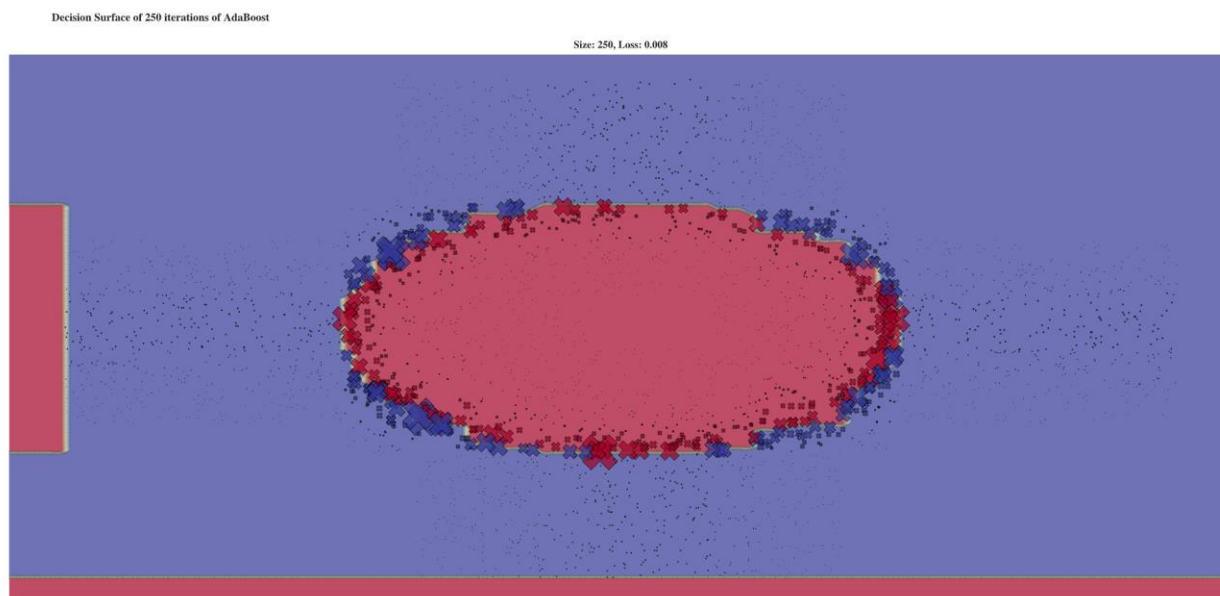
מגרפים אלו ניתן לראות שהביצועים הולכים ומשתפרים עם עליית מספר האיטרציות. המודל מצליח לחזות את הפונקצייה המקורית של המידע.

שאלה 3

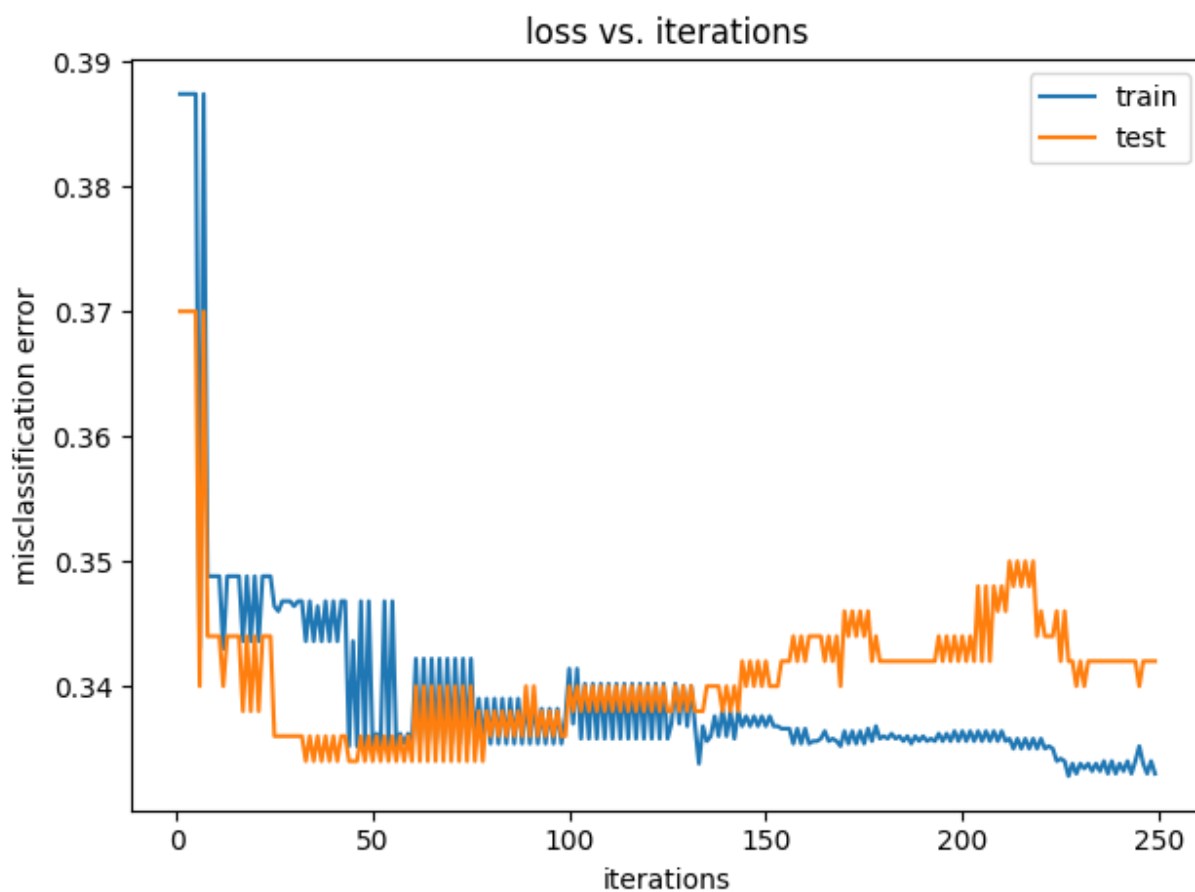


בגרף זה ניתן לראות שעבור 237 איטרציות השגיאה היא 0.006! מאוד קטנה.

שאלה 4



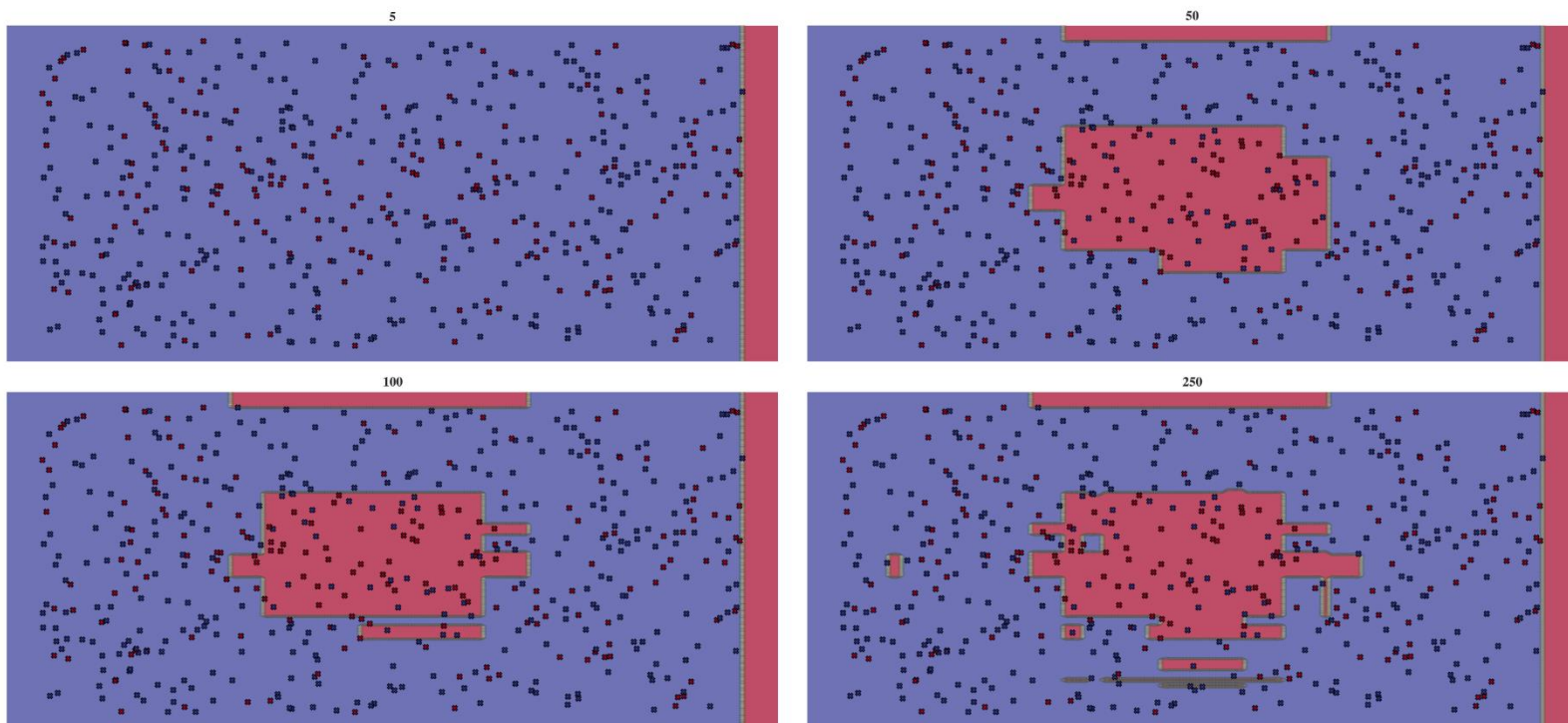
ניתן לראות לפי גודל הנקודות שהנקודות הבעייתיות הן אלו שקרובות לקצה התחום האדום והכחול, ז"א נקודות שבאמת קשה להגיד לאיזה תחום הן משתייכות.



במקרה זה ניתן לראות שהרעש גורם לכך שהעלאה נוספת של מספר האיטרציות גורמת לoverfit במסגרתו הtrain משתפר והtest נעשה גרוע יותר. הביאס יורד אבל הוואריאנס עולה עבור ערכים גדולים מ~37.

2.

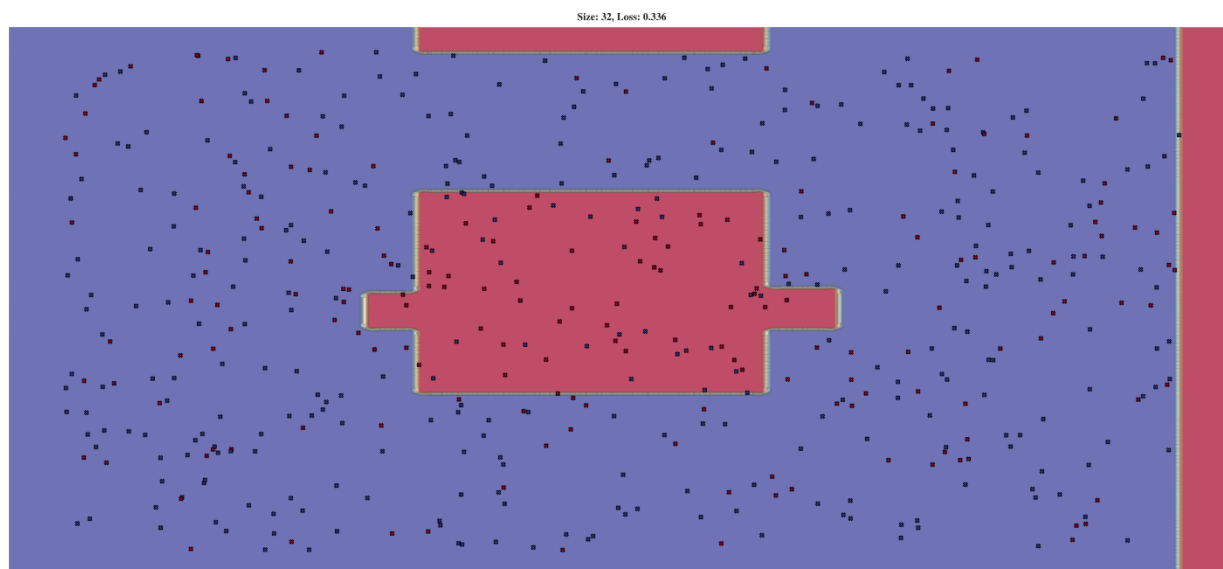
Decision Surface of 5,50,100,250 iterations of AdaBoost



גם במקרה זה הגרפים מתארים יחסית טוב את ההתפלגות המקורית על אף שיש הרבה רעש. עבור מספר איטרציות גבוה מדי (250) יש מקטעים שמרמזים שהתבצע overfit.

3.

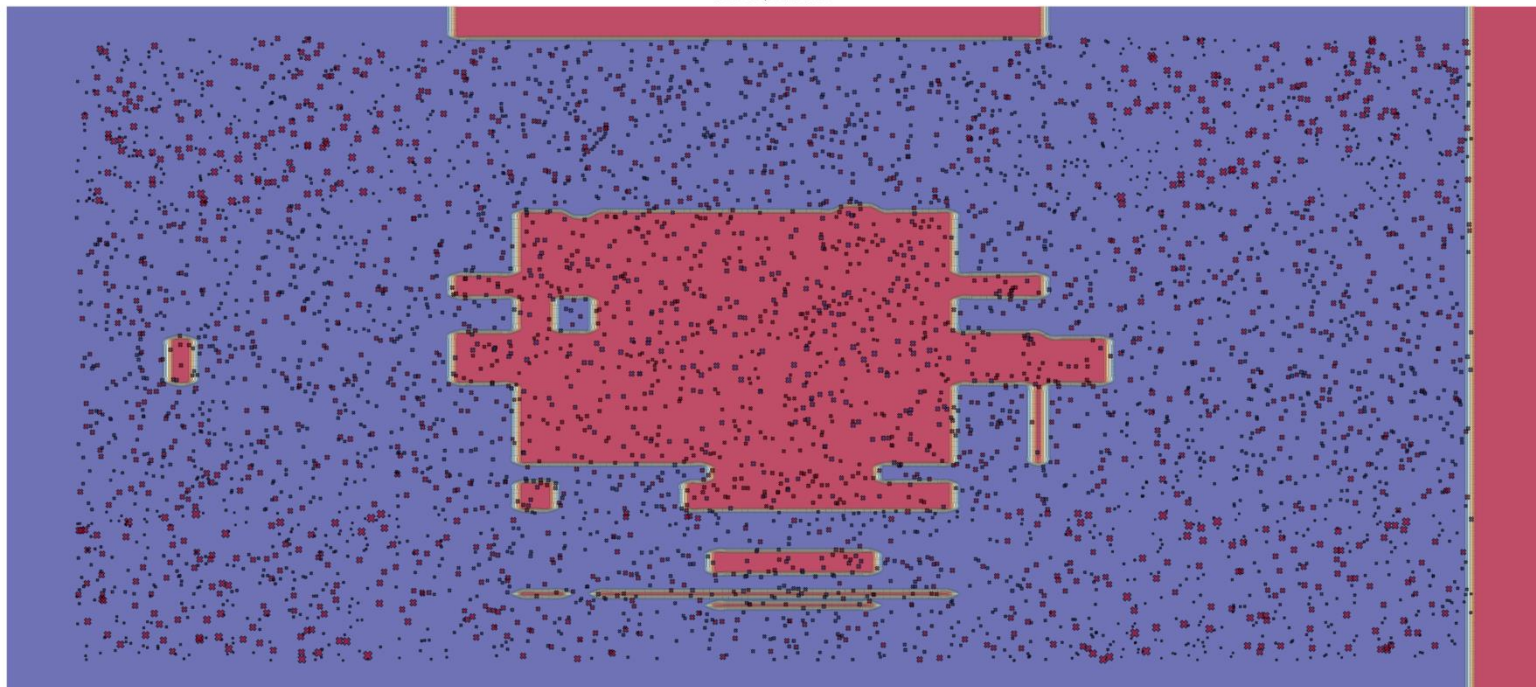
Decision Surface of 32 iterations of AdaBoost



הפעם הפיט המיטבי הוא עבור 32 איטרציות והשגיאה היא 0.336. מכך שהאיטרציות נמוכות הפעם אסיק שמודל זה התאים יותר טוב מאחר שהוא ריאנס שלו נמוך יותר והוא פחות עושה overfit על המידע.

Decision Surface of 250 iterations of AdaBoost

Size: 250, Loss: 0.342



בגרף זה ניתן לראות שהנקודות הבעייתיות הן אלו שנמצאות מחוץ לתחום ההגיוני עבורן, ז"א נקודות שנובעות מרעש.