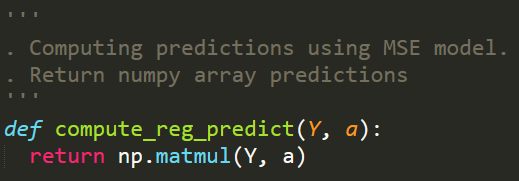
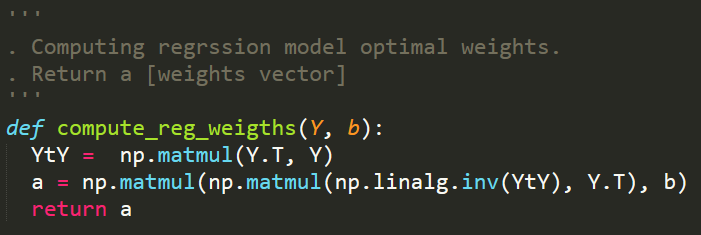
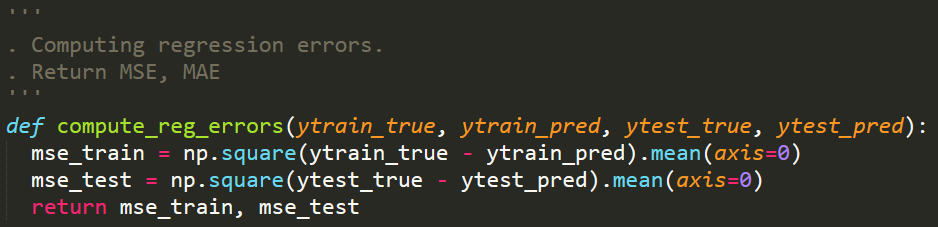
**תרגיל בית 2:**

**מגישים: איתי גיא (ת.ז. – 305104184), אורי בן-יצחק (ת.ז. – 066374737)**

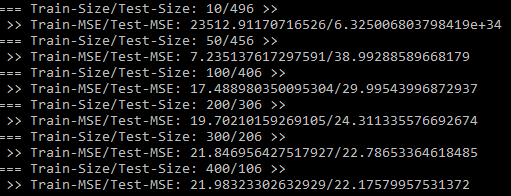
שאלה 1:

.



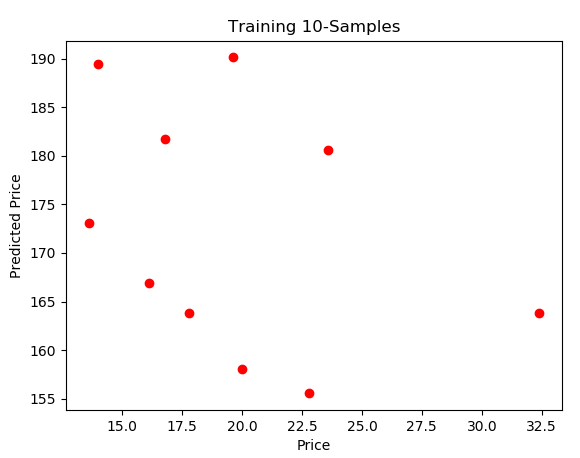
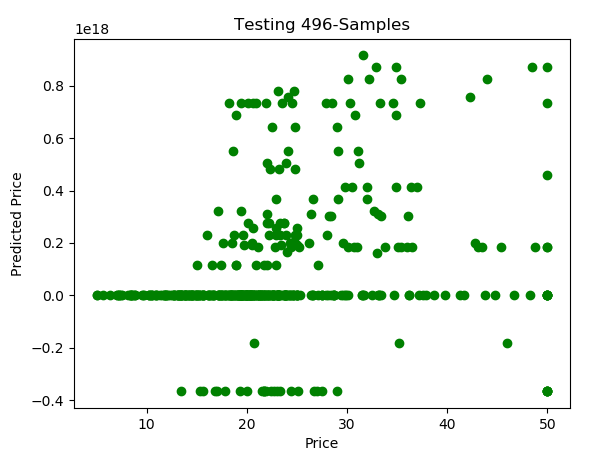
.



. כפי שניתן להבחין בהדפסה הנ"ל, שגיאת האימון יורדת ושגיאת הבדיקה עולה כתוצאה מהרצאת על הגדלים 10, 50, 100, 200, 300, 400.

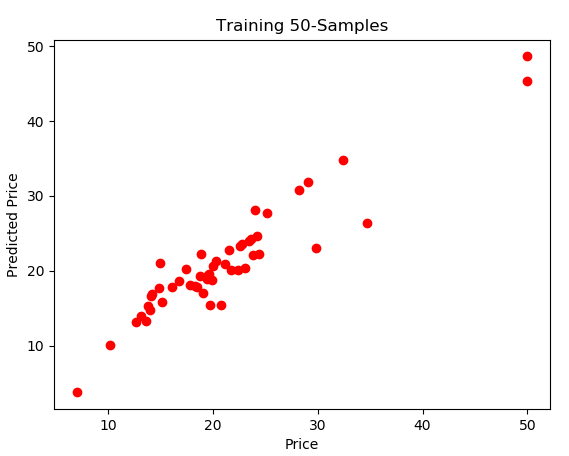
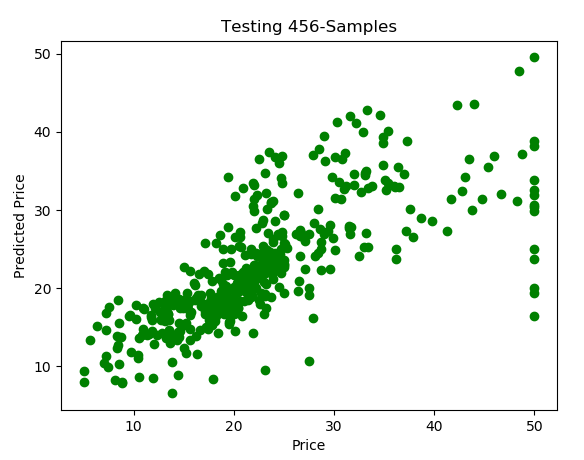
כאשר ישנם 10 דוגמאות אימון השגיאות מאוד גדולות מכיוון שדוגמאות האימון מעטות מדי ומפוזרות לכן לא ניתן למצוא מודל לינארי טוב שיכליל את דוגמאות הבדיקה (זה דומה לרעש). בנוסף לכך, ישנם הרבה דוגמאות בדיקה ביחס לדוגמאות אימון ולכן שגיאת האימון נמצאת קטנה יותר משגיאת הבדיקה.

ניתן לראות את דוגמאות האימון לעומת דוגמאות השגיאה:

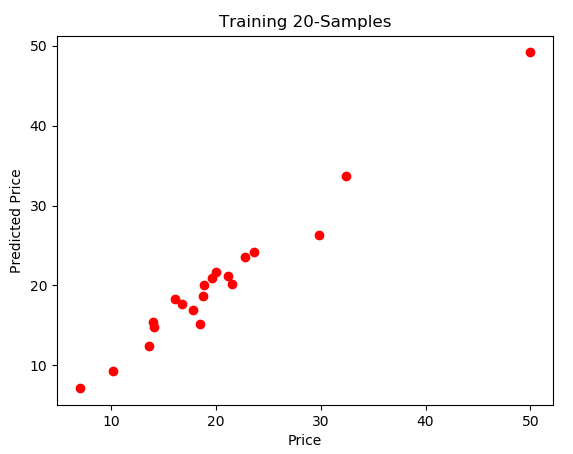
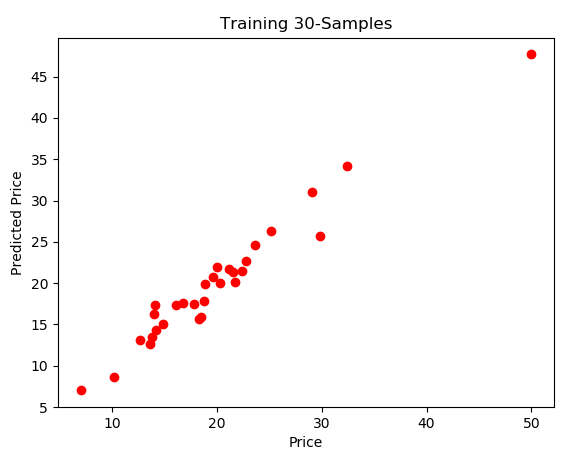
 

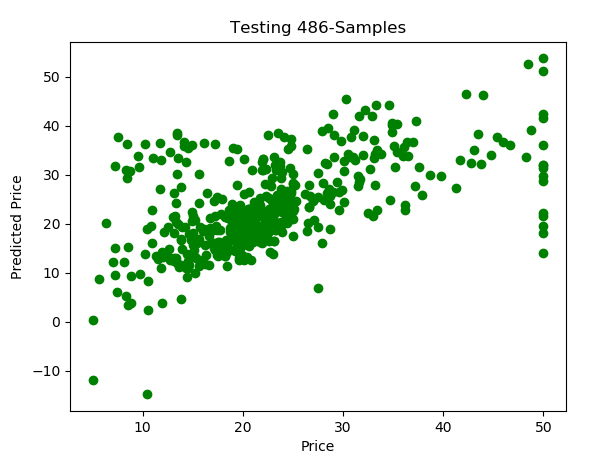
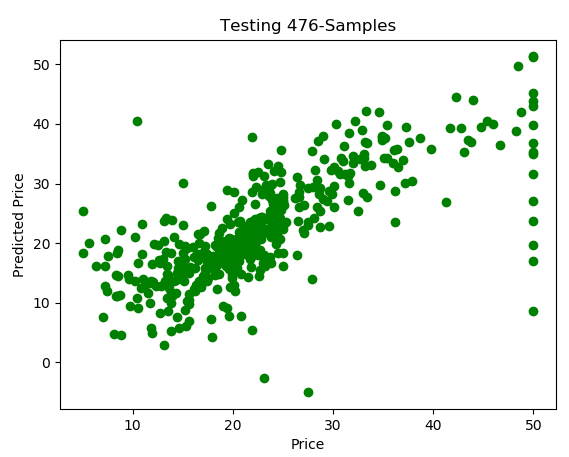
החל מ-50 דוגמאות אימון ניתן לראות שיש לדוגמאות מבנה לדוגמאות האימון והבדיקה וצופים שנמצא מודל שיוכל להקטין את השגיאות (המרחקים בין הנקודות הופכים קטנים יותר).

ניתן לראות את דוגמאות האימון לעומת דוגמאות הבדיקה:

כך, ככל שהכמות של דוגמאות האימון גדלה ההכללה הופכת טובה יותר והשגיאה תרד. בנוסף, השגיאה יורד גם כי הכמות של דוגמאות הבדיקה יורדת ולכן כמות האיברים שמחשבים את שגיאת הבדיקה קטנה.

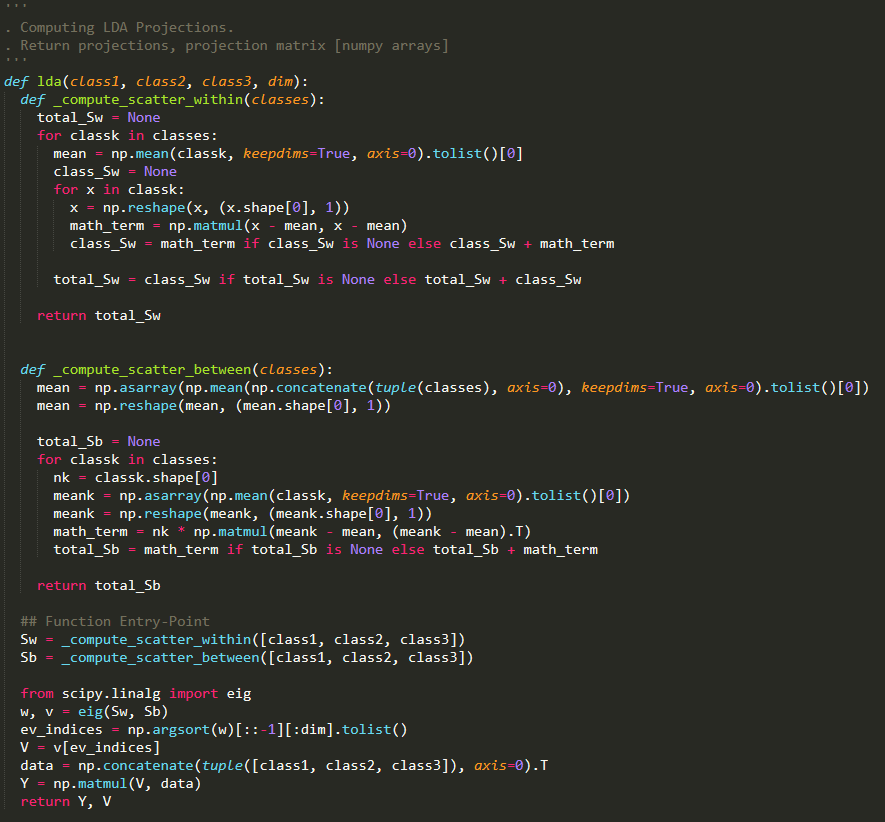
 

כפי שניתן לראות הקו הלינארי יבצע סיבוב בכדי למזער את השגיאה. הסיבוב הגדול ביותר יקרה מעל 10 דוגמאות אימון.

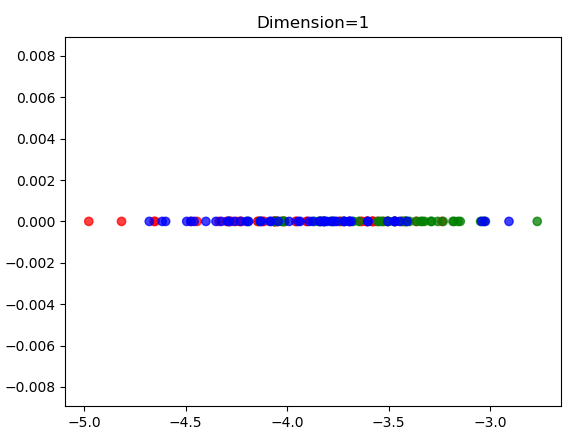
שאלה 2:

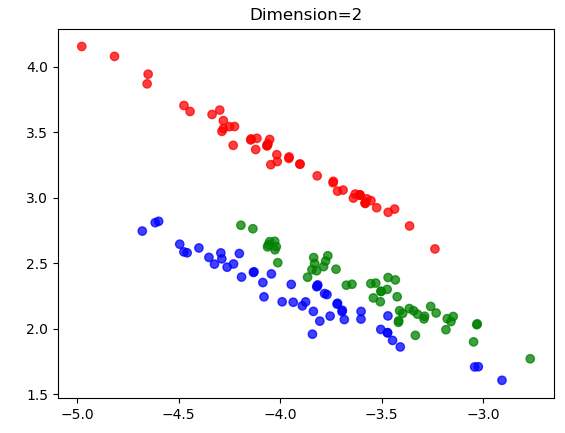
.



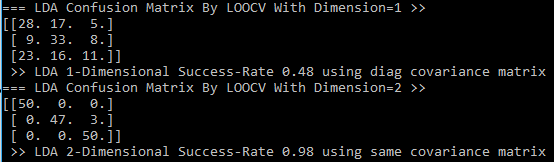
.

* דוגמאות עם מימד בודד לא ניתן להפרדה כמעט (יופרד עם המון שגיאות).
* דוגמאות עם 2 מימדים ניתנות להפרדה באופן כמעט מושלם.

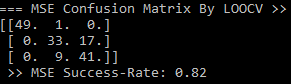




.



*.*



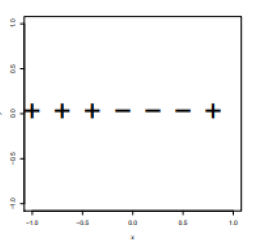
ההבדלים בין ו- הם בסוג המודל. בסעיף השתמשנו במסווג בייסיאני שמאוד טבעי להשתמש כאשר הנתונים נראה כמו בסעיף – גאוסיינים. לעומת זאת, בסעיף השתמשנו במסווג לינארי שממזער את השגיאה הריבועית – הוא מצא פתרונות לא רעים אבל בגלל הקרבה בין מחלקות 2 ו-3 (ירוק/כחול) הוא שגה בהערכת הפתרון. ניתן לראות שאפילו שהמחלקה האדומה רחוקה מהמחלקות האחרות הוא שעה בדוגמא שקרובה למחלקה הירוקה (הקירוב הלינארי זז בכדי למזער את השגיאה הריבועית והנקודה זז למחלקה הירוקה).

* *הקוד המלא מצורף לתרגיל.*

שאלה 3:

. כל עוד ניתן להשתמש רק במשפחת מסווגים לינאריים התשובה היא **לא**. הסיבה לכך היא שהנתונים בנויים בצורה כזאת שבכל איטרציה האלגוריתם יבחר מסווג לינארי **אחר** שיש לו טעות אחת לדוגמה החיובית וטעות אחת לדוגמה השלילית ומכיוון שמתחילים ממשקלים אחידים של נקבל ש- ומכאן . מנוסחת עדכון המשקלים נקבל שהמשקלים של הטעויות וההצלחות ישארו זהים ולכן גם בכל האיטרציות נקבל שמשקלי המסווגים הם ולכן לא נצליח לקבל שטוב יותר ממסווג רנדומלי.

.



1 error

+

h1

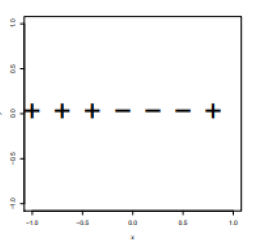
.

ה- יהיה .

.

* הדוגמאות בצד הנכון:
* הודמאות בצד הלא נכון:

.



3 errors

+

h2

. בסיום ריצה זו הדוגמאות בעלות המשקלים הנמוכים ביותר יהיו כל הדוגמאות השליליות (אותם משקלים לכל הנקודות השליליות).

. ה- **לא** השתפר מכיוון שבאיטרציה זו ולכן עדיין לא יהיה ניתן לסווג את הדוגמא החיובית הקיצונית מימין.

כי נקבל: .

שאלה 4:

. נחשב את האנטרופיה של כל אחת מהתכונות:

מכיוון ש-נותן לנו את האנטרופיה המינימאלית נבחר בו להיות ה- בעץ.

. נחשב את האנטרופיה הכללית של הנתונים:

נחשב את ה- של כל אחת מהאפשרויות:

מכיוון שמדד ה- מחפש ערך מקסימלי נבחר במקרה זה את התכונה

. כולם *(b, c, d)* **פרט** **ל-** מתארים את עץ ההחלטה.

עץ החלטה מבצע חיפוש היוריסטי במרחב הקלט ללא כמעט הנחות נוספת על הנתונים. סיבה זו מייצרת בהגדרץה אבל כתוצאה מ- נקבל (b). יתר המאפיינים *(c, d)* נובעים ממבנה העץ והאלגוריתם (כפי שנילמד והוצגו דוגמאות בהרצאה):

: מאפיין את העץ מכיוון שהוא "חותך" את המרחב לפי בחירת המאפיינים בעץ וחיתוך כזה לפי הצירים הוא לא חלק.

: מאפיין את העץ מכיוון שניתן לבצע על כמות תכונות ללא מגבלה ובכל צומת העץ יתקדם גם בלי מגבלה עד הגעתו לעלה (העץ יבחר את התכונות המסבירות ביותר עבור ההכללה).