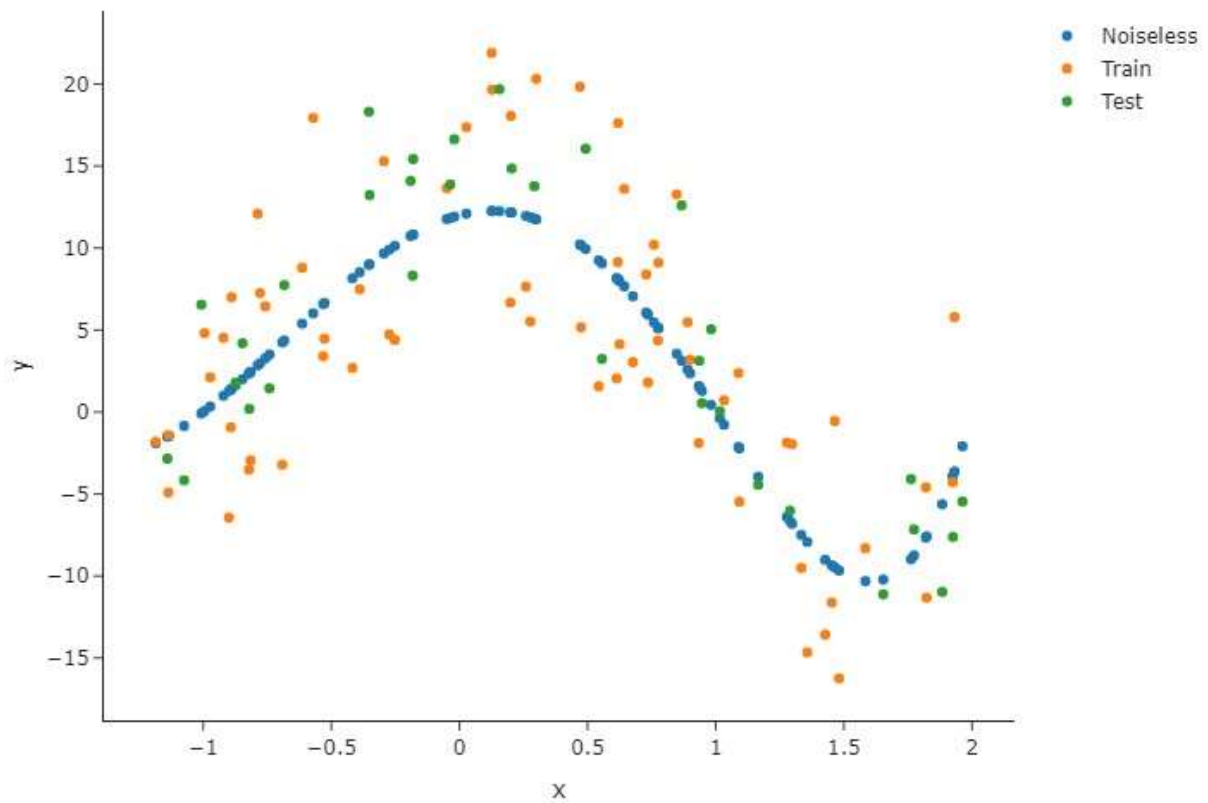
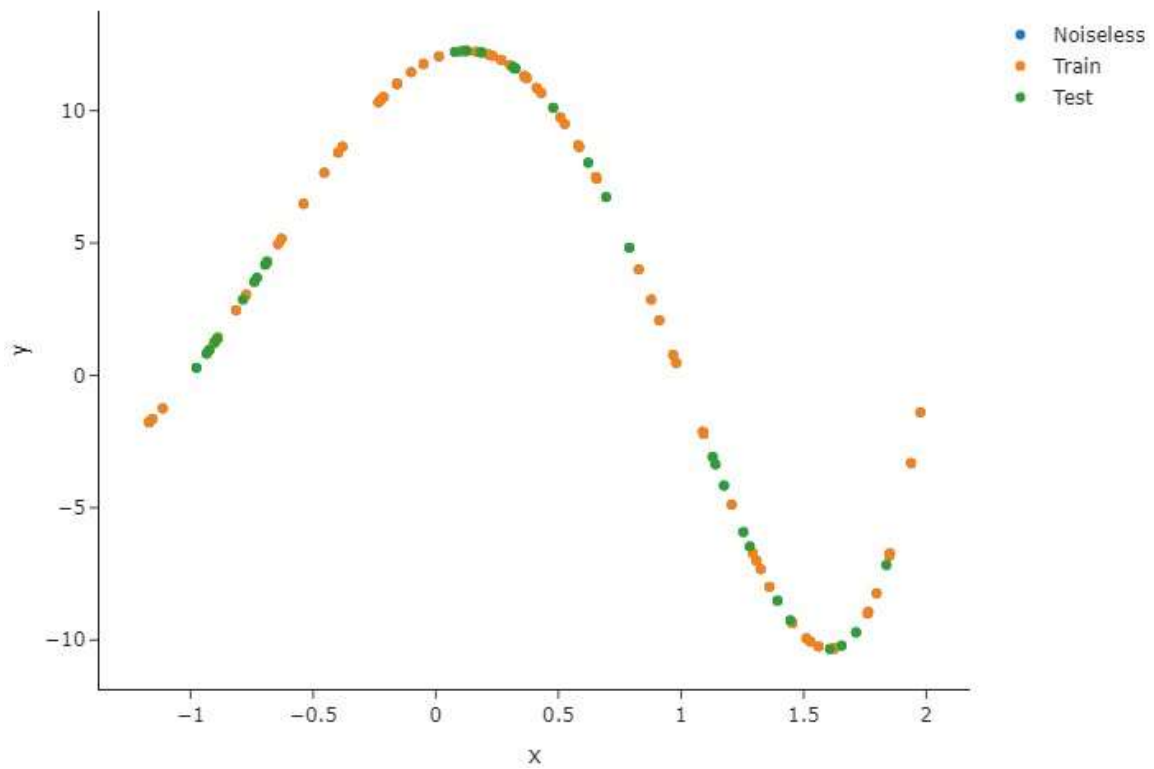


1. להלן הפלטים עבור סט הנתונים בהינתן קלטים שונים

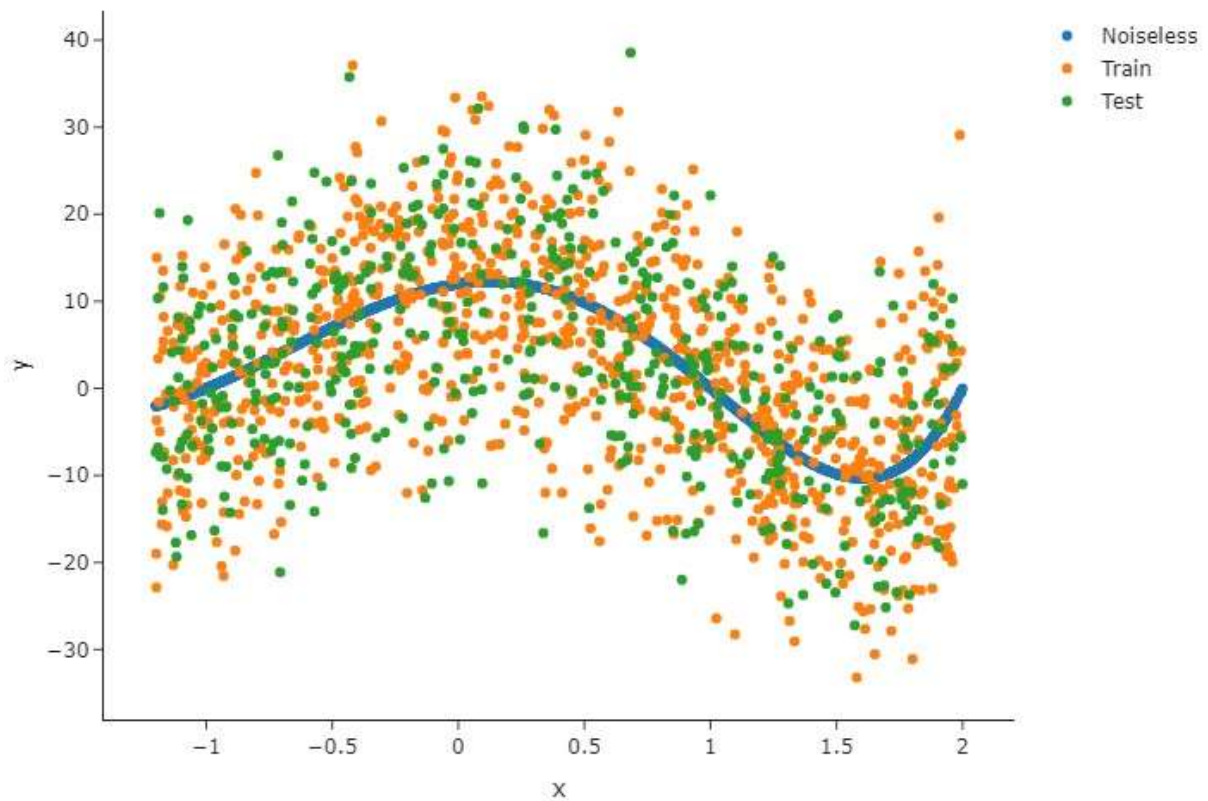
Q1, noise = 5, n=100



Q1, noise = 0, n=100

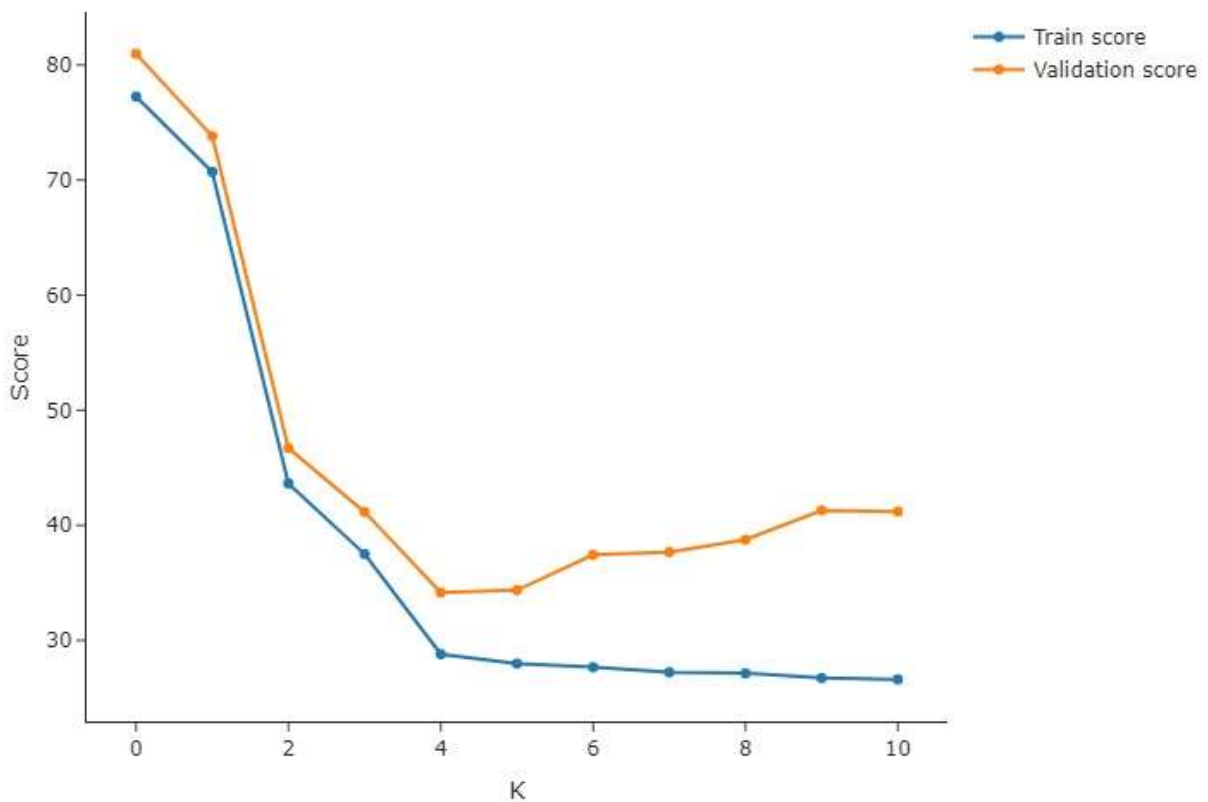


Q1, noise = 10, n=1500

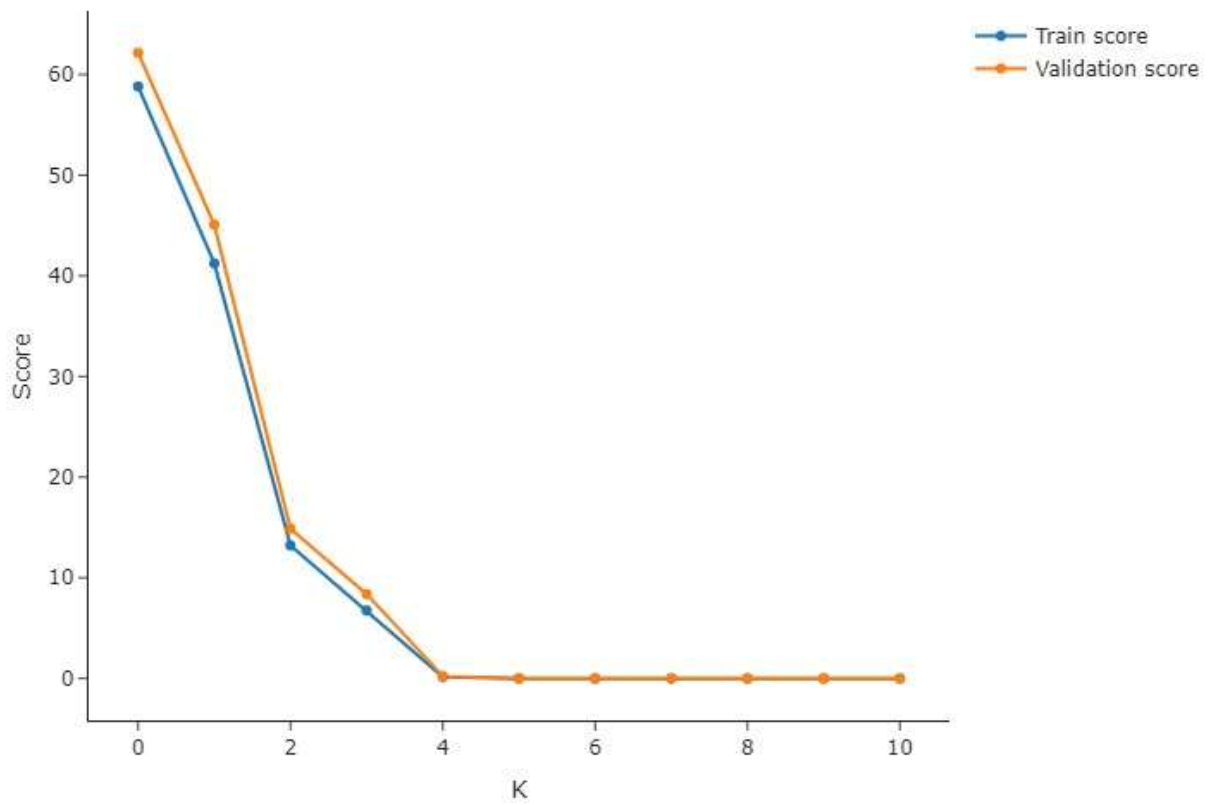


2. להלן הפלטים בהינתן התרחישים השונים (תושבה משולבת עבור סעיפים 4,5):

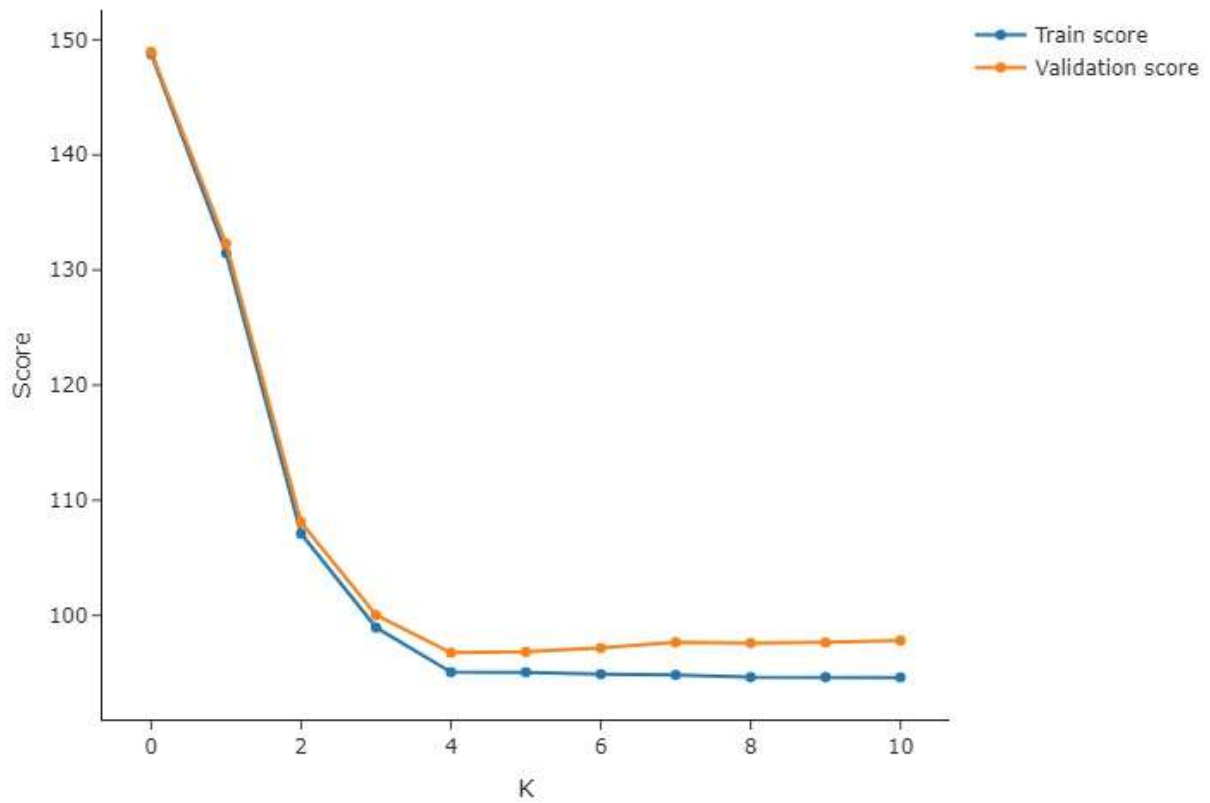
Q2, noise = 5, n=100



Q2, noise = 0, n=100



Q2, noise = 10, n=1500



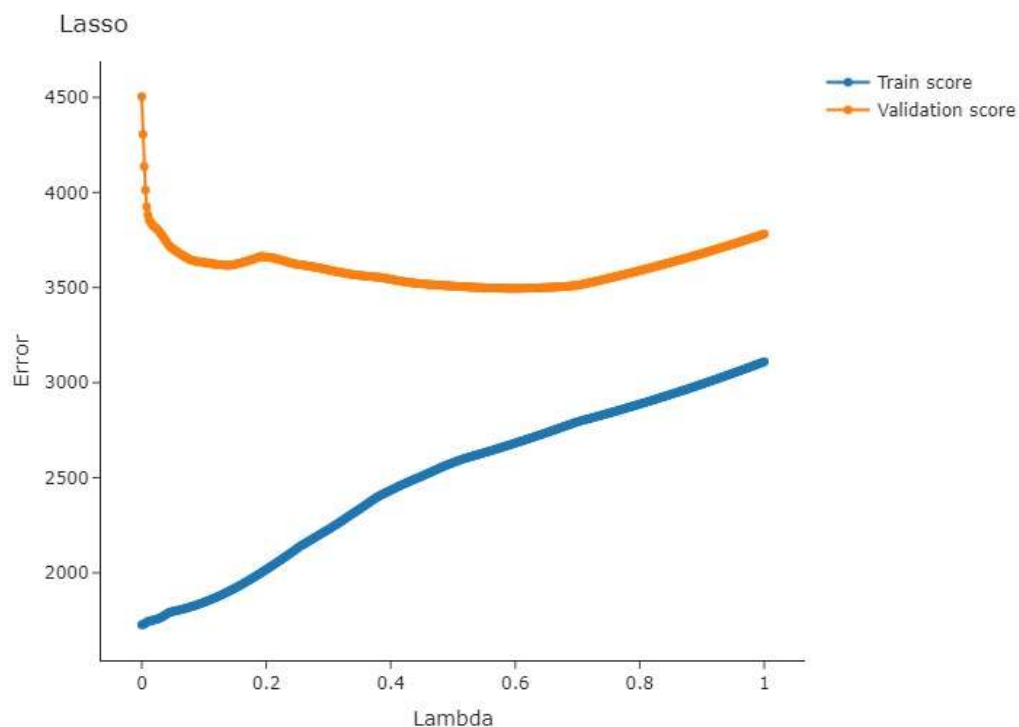
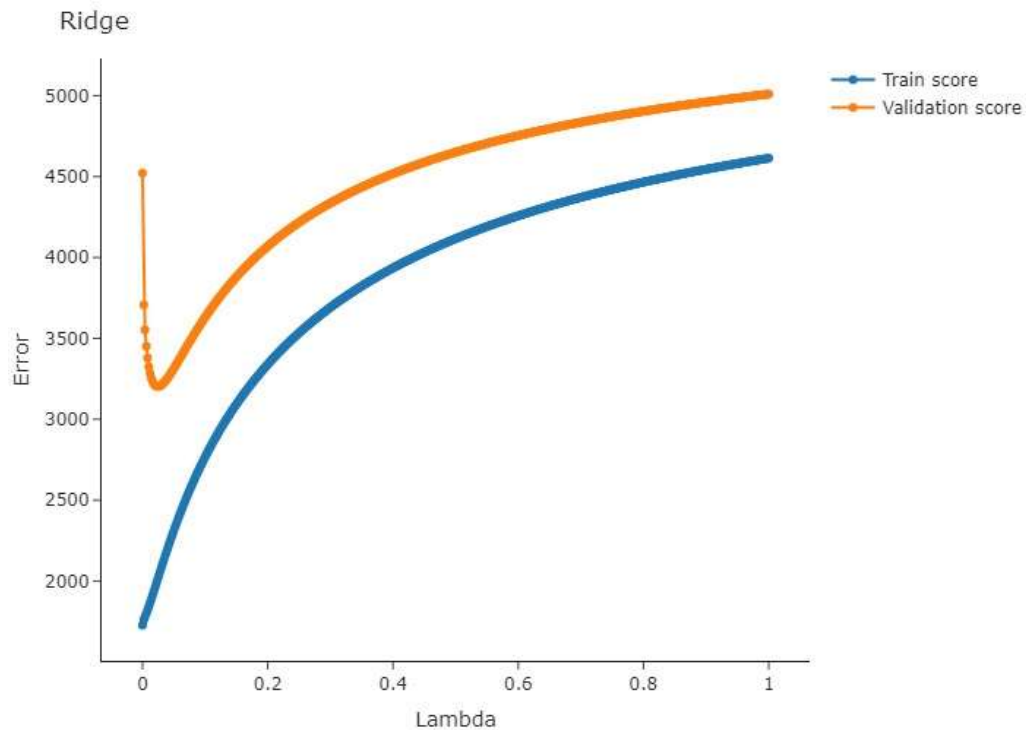
נתבונן בתוצאות שקיבלנו, נשים לב למספר אבחנות:

- ראשית כאשר כמעט ואין רעש שגיאת האימון מתלכדת עם שגיאת הולידאציה – אבחנה שנובעת ישירות מן העבודה כי כל הנקודות מייצגות את המציאות, כאשר אין משמעות אמיתית לאימון המודל על חלקים שונים (רועשים יותר או פחות) – כאשר הרעש עולה ההפך מתקבל, שוב אבחנה לה אנו מצפים.
- כמו כן, מנגנון הבקרה שלנו עובד כאשר ניתן לראות כי שגיאת הולידאציה גדולה לרוב משגיאת הסט האימון – כמו כן דבר שנרצה לראות.
- נשים לב כי תחת שינוי דרגות הפולינום השגיאות גם מתנהגות בצורה קוהרנטית, כאשר שדרגת הפולינום מתקרבת לדרגה האמיתית השגיאות הן מזעריות ביחס לקצוות האחרים – נשים לב כי שתי המדידות מקיימות גם את עיקרון *bias* – *variance* שראינו מספר פעמים עד כה בקורס (*under fit* מול *overfit*).
- לבסוף נראה כי גם כאשר אנו מעלים את מספר המדידות שלנו למרות שהרעש גדול, המודל מצליח להתמודד ולספק תחזית טובה. זאת אנו מסוגלים להפיק תחת שימוש ב-*cross validation*, בכל שלב מבצעים אימון על חלק אחר בסט האימון שלנו ובודקים שוב ושוב כך שהמודל נחשף לכמה שיותר מן המידע. באופן זה אנו מקטינים את ה-*overfit* שיכול להתקיים בהתאמת פולינומים תחת רעש גבוהה ומחזקים את השימוש בחוק המספרים הגדולים – שגם תחת הרעש מספר רב של דגימות ישאף לתצורה האמיתית (נשאף לתוחלת, אשר 0) ולכן נוכל עדין לקבל פרדיקציה טובה.

3. להלן דרגות הפולינום אשר המודל חזה עם הטעויות שקיבלנו עבור שימוש בדרגה זו:

- במקרה עבור 100 דגימות ורעש 5 -  $k = 4, MSE = 16.57$
- במקרה עבור 100 דגימות ורעש 0 -  $k = 5, MSE = 1.35$
- במקרה עבור 1500 דגימות ורעש 10 -  $k = 4, MSE = 98.89$

7. להלן הפלטים עבור  $\lambda \in [0,1]$  אשר נלקחו 500 נקודות במרחק שווה. אציין כי ע"מ להגיע לטווח זה, ראשית התחלתי עם טווח גדול  $[0,10]$  ניתן היה לראות כי השגיאה מתקבעת ב- $\sim 4$  ולכן צמצמתי את הטווח עד לקבלת מדידה שאפשרה לקבל ערך מינימלי.



שוב ניתן לראות את הבדל הניכר בין השגיאות על סט האימון וסט הוואלידציה, כאשר הפער היה מעט גדול יותר עבור השימוש ב-*Ridge*. כמו כן, ניתן לראות את מגמת השגיאה כמגמה שראינו

בכיתה – לפיה השימוש בנורה  $l_2$  מייצרת פונקציה גזירה ב-0 ולכן עקומת רידג' מעט מזכירה יותר פרבולה מאשר עקומת  $Lasso$  אשר דומה יותר לפונקצית הערך המוחלט.

כמו כן, ניתן לראות כי הוספת ערך הרגולריזציה השפיע יותר בהתחלה על  $Lasso$  מאשר על  $Ridge$  ביכולת להוריד את השגיאה תחת סט הולידאציה.

8. להלן תוצאות השגיאה שהתקבלו עבור שלושת המודלים :

*Ridge achieved 3249.69 with lambda of 0.024*

*Lasso achieved 3641.16 with lambda of 0.597*

*Basic linear regression achieved 3612.25*

ניתן לראות כי במקרה זה  $Ridge$  הוא זה שהשיג את השגיאה הקטנה ביותר, תחת שימוש בפרמטר רגולריזציה מאוד קטן, כ-0.02 אך על סט האימון הקטן הנ"ל גרר שיפור משמעותי (מעל 10%) בשגיאה.