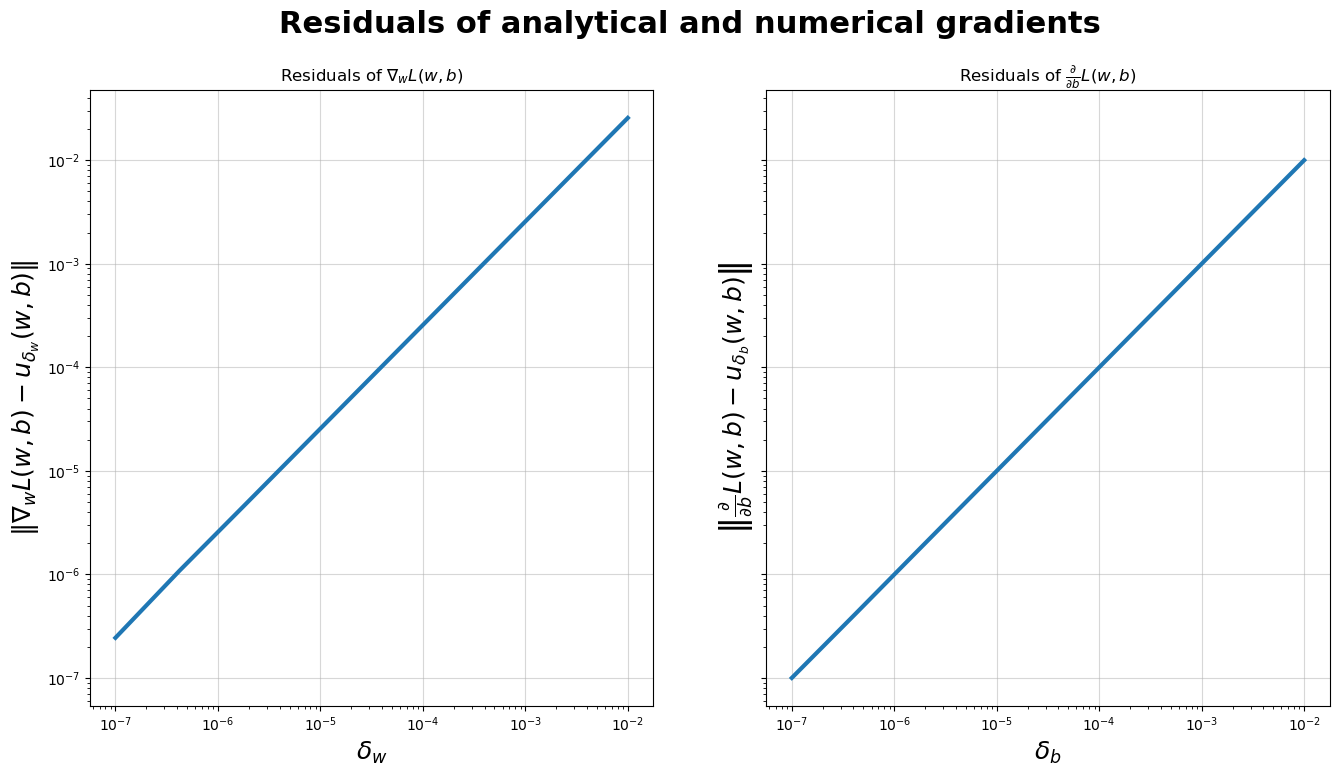
**תרגיל בית רטוב 2 – מבוא למערכות לומדות (02360766)**

*בן הייטנר – 213930175*

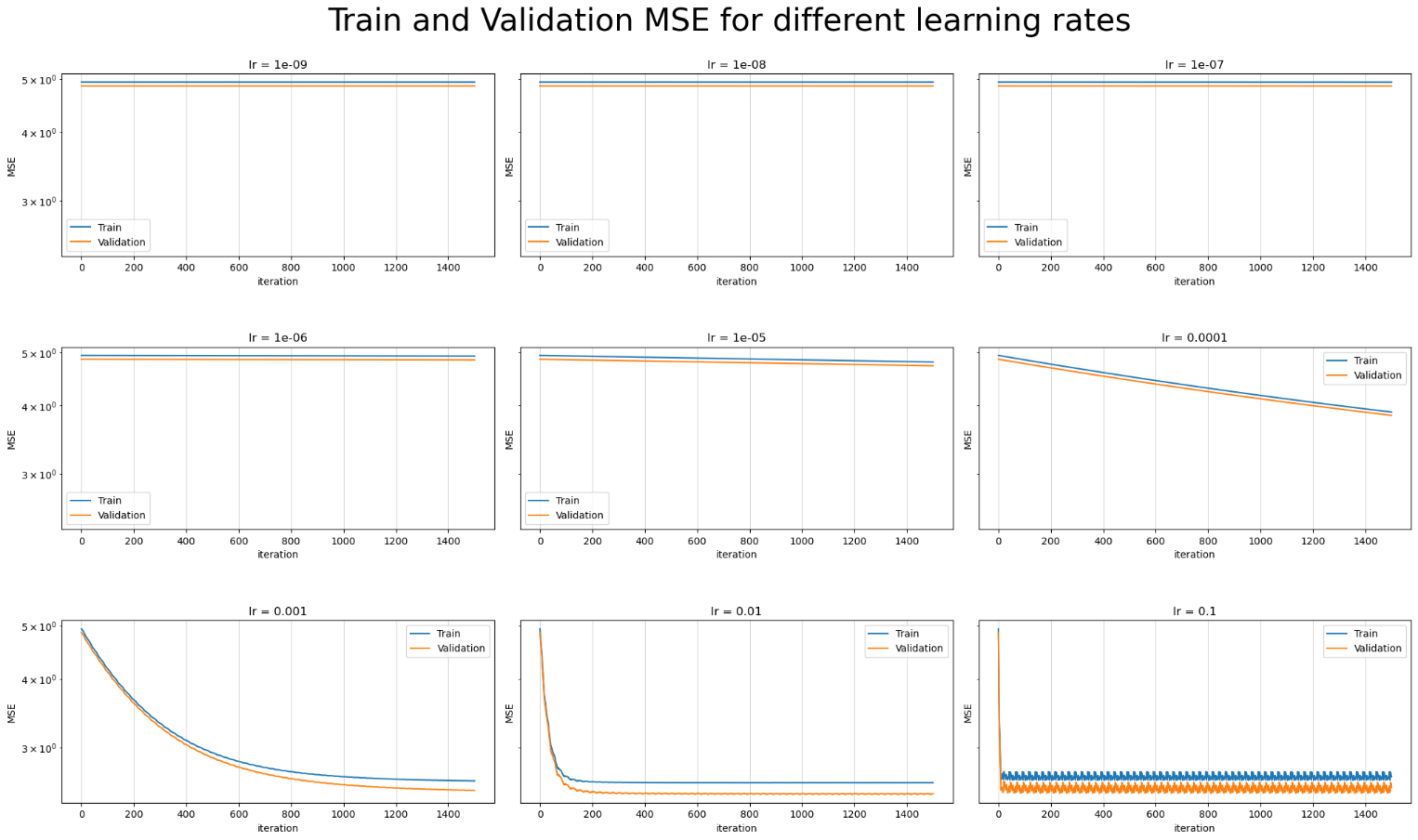
*לילך ביטון – 205764517*

(Q1)

(Q2)



(Q3)



כשקצב הלמידה קטן מאוד אז, כצפוי דרושות המון איטרציות כדי שנתחיל לראות שינוי בשגיאה של המודל (בפרט עבור כאלה קטנים במיוחד אנחנו לא רואים אף שינוי תוך 1500 צעדים). כאשר מגדילים את קצב הלמידה, השגיאה קטנה לאורך האיטרציות ומתכנסת לערכים קטנים מאוד. אבל, עבור קצב למידה גדול מתחיל להיות רעש בהתכנסות כתוצאה מקפיצות גדולות מדי באיטרציות מאוחרות.  
ניתן להסיק שקצב הלמידה האופטימלי הוא 0.01 שכן הוא נותן את הטעות הקטנה ביותר על קבוצת הולידציה עם רעש שאינו גדול מספיק כדי לפגוע בתוצאות המוצלחות שלו.  
נציין שאין טעם להגדיל את מספר הצעדים שנבצע עבור מודל עם קצב למידה של 0.01 כיוון שהוא כבר הגיע להתכנסות במספר הצעדים הנתון ולא יהיה שום שינוי מהותי בטעות שלו ככל שנמשיך להריצו.

(Q4)

|  |  |  |  |
| --- | --- | --- | --- |
| Valid MSE | Train MSE | Section | Model |
| Cross validated | |  |  |
| 3.793 | 3.787 | 2 | Dummy |
| 2.629 | 2.549 | 2 | Linear |
| 2.624 | 2.551 | 3 | Lasso Linear |
| 3.452 | 3.414 | 4 | Polynomial Lasso |
| 0.977 | 0.677 | 5 | GBM Regressor |

(Q5)

A graph with lines and numbers

Description automatically generated

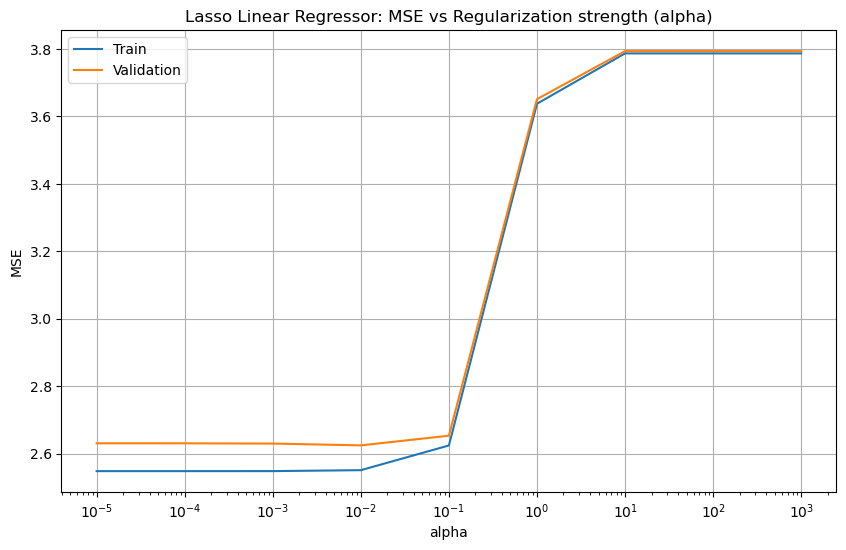
קצב למידה אופטימלי: 0.01  
שגיאת MSE על קבוצת אימון: 2.549  
שגיאת MSE על קבוצת ולידציה: 2.629

|  |  |  |  |
| --- | --- | --- | --- |
| Valid MSE | Train MSE | Section | Model |
| Cross validated | |  |  |
| 3.793 | 3.787 | 2 | Dummy |
| 2.629 | 2.549 | 2 | Linear |

(Q6)

הנירמול של הנתונים מתבצע לפי נתוני האימון. על כן, בהיעדר שגיאות נומריות, אי נירמול הנתונים לא היה פוגע בביצועי המודל על נתוני האימון.

(Q7)



חוזק רגולריזציה אופטמילי: 0.01  
שגיאת MSE על קבוצת אימון: 2.551  
שגיאת MSE על קבוצת ולידציה: 2.624

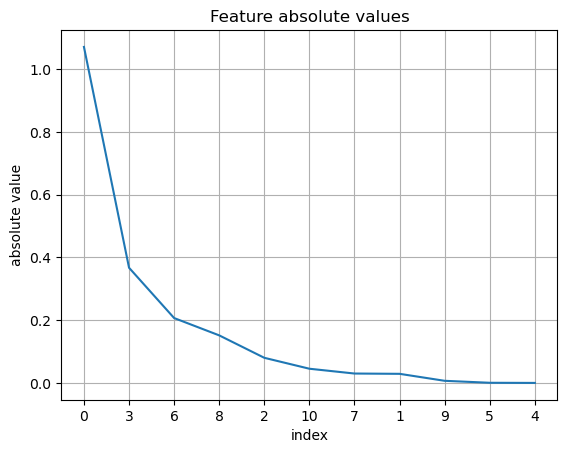
(Q8)

|  |  |  |  |
| --- | --- | --- | --- |
| Valid MSE | Train MSE | Section | Model |
| Cross validated | |  |  |
| 3.793 | 3.787 | 2 | Dummy |
| 2.629 | 2.549 | 2 | Linear |
| 2.624 | 2.551 | 3 | Lasso Linear |

(Q9)

חמשת ה-features בעלי המקדמים הגדולים ביותר בערך מוחלט הם (בסדר יורד):  
happines\_score  
PCR\_03  
PCR\_06  
PCR\_08  
PCR\_01

(Q10)



(Q11)

כפי שאפשר לראות, הרבה מקדמים מקבלים ערכים קרובים מאוד ל-0. זו התנהגות סטדרטית למודל Lasso שכן נורמליזציה על בסיס נורמת L1 יוצרת התמקדות במספר מצומצם של features רבי השפעה (מקבלים ערכים גדולים יחסית) והתעלמות מהשאר (מקבלים ערכים קטנים יחסית).

(Q12)

כמו בשאלה שש, הנירמול של הנתונים מתבצע לפי נתוני האימון. על כן, בהיעדר שגיאות נומריות, אי נירמול הנתונים לא היה פוגע בביצועי המודל על נתוני האימון.

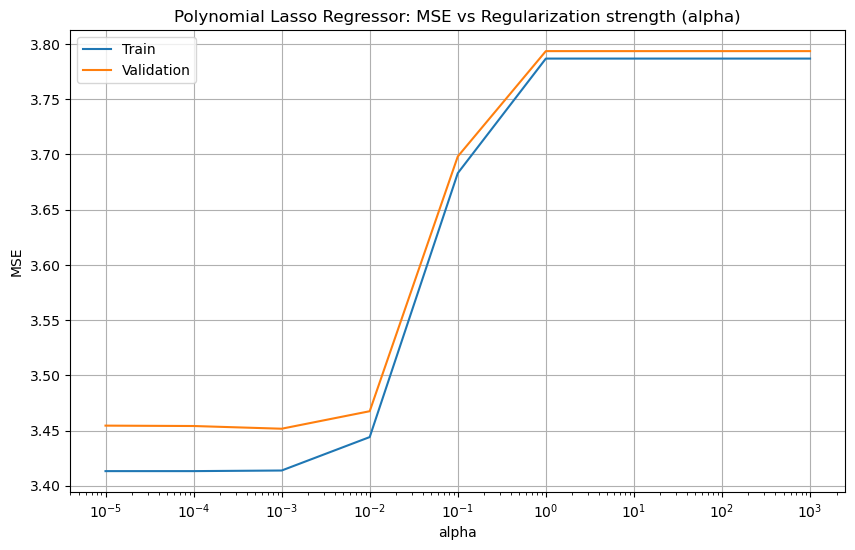
(Q13)

היינו מצפים לפילוג יותר אחיד של המקדמים (במקום כמה עם כרך משמעותי וחלק גדול שואפים לאפס). כתוצאה מכך שבניגוד ל-Lasso, Ridge משתמש ברגולריזצית L2. זו לרוב לא נוטנת ערכים מסדר גודל 0 לאף אחד מה-features (עדיין קיימת העדפה ל-features מסוימים).

(Q14)

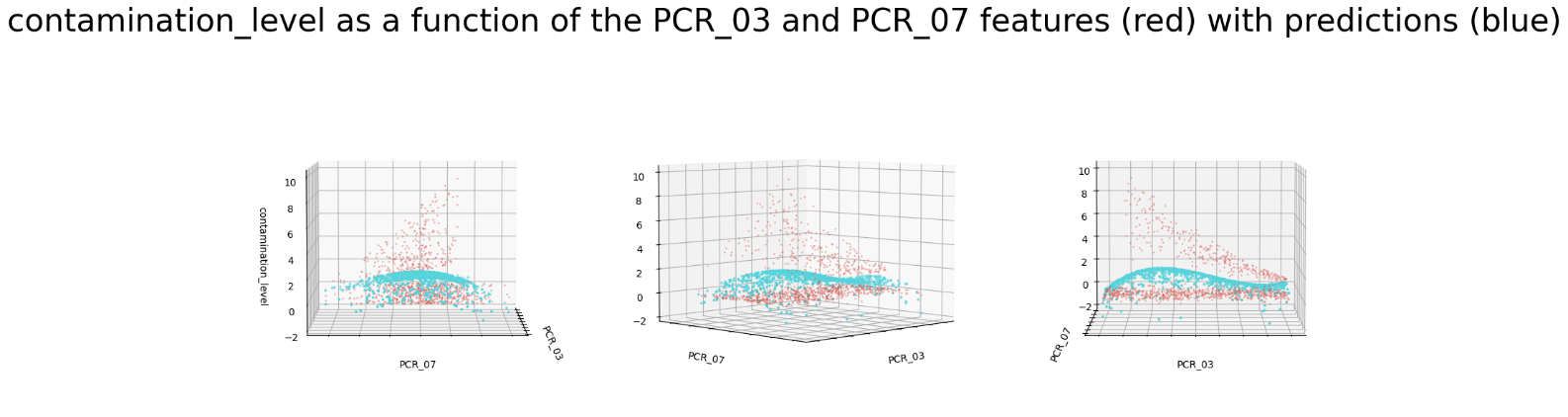
מיפוי פולינומי משנה את מרחב הfeatures באופן שסביר שיצור אי-איזון בקצב השינוי של כל feature. לכן, בהיעדר נורמליזציה, קצב השינוי ישפיע באופן שונה על כל ציר. נירמול מחדש מאפשר לנו לשמר את היחס המקורי (יחסית אחיד) בין השפעת קצב השינוי על כל feature.

(Q15)



חוזק רגולריזציה אופטמילי: 0.001  
שגיאת MSE על קבוצת אימון: 3.414  
שגיאת MSE על קבוצת ולידציה: 3.452

(Q16)



(Q17)

|  |  |  |  |
| --- | --- | --- | --- |
| Valid MSE | Train MSE | Section | Model |
| Cross validated | |  |  |
| 3.793 | 3.787 | 2 | Dummy |
| 2.629 | 2.549 | 2 | Linear |
| 2.624 | 2.551 | 3 | Lasso Linear |
| 3.452 | 3.414 | 4 | Polynomial Lasso |

(Q18)

A chart of different colors

Description automatically generated

צמד (min\_sample\_leaf, learning\_rate) אופטימלי: (19, 0.1)  
שגיאת MSE על קבוצת אימון: 0.677  
שגיאת MSE על קבוצת ולידציה: 0.977

(Q19)

|  |  |  |  |
| --- | --- | --- | --- |
| Valid MSE | Train MSE | Section | Model |
| Cross validated | |  |  |
| 3.793 | 3.787 | 2 | Dummy |
| 2.629 | 2.549 | 2 | Linear |
| 2.624 | 2.551 | 3 | Lasso Linear |
| 3.452 | 3.414 | 4 | Polynomial Lasso |
| 0.977 | 0.677 | 5 | GBM Regressor |

(Q20)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Test MSE | Valid MSE | Train MSE | Section | Model |
| Retrained | Cross validated | |  |  |
| 4.963 | 3.793 | 3.787 | 2 | Dummy |
| 3.432 | 2.629 | 2.549 | 2 | Linear |
| 3.428 | 2.624 | 2.551 | 3 | Lasso Linear |
| 4.265 | 3.452 | 3.414 | 4 | Polynomial Lasso |
| 1.124 | 0.977 | 0.677 | 5 | GBM Regressor |

המודל עם הביצועים הטובים ביותר הוא ה-GBM Regressor.  
Dummy מבצע ממוצע טריוויאלי ועל כן השגיאה המוגדלת שלו על המבחן נובעת מהבדל קל בהתפלגות של המבחן מזו של קבוצות ה-cross validation.  
המודל הלינארי ומודל הלאסו הלינארי שניהם ביצעו מעט overfitting (השגיאה על המבחן גדלה ביחס לזו ב-cross validation כתוצאה מהתאמת יתר לנתוני האימון במעט).  
מודל הלאסו הפולינומי חטא פעמיים – גם נתן תוצאות פחות טובות מהמודלים הלינארים שבאו לפניו וגם ביצע overfitting דומה לשלהם.  
מודל ה-GBM Regressor לעומתם הוא הצדיק בסדום – נתן תוצאות טובות מאוד ב-cross validation שטובות כמעט באותה מידה על נתוני המבחן.