

תרגיל בית רטוב 3 – מבוא למערכות לומדות (02360766)

בן הייטנר – 213930175

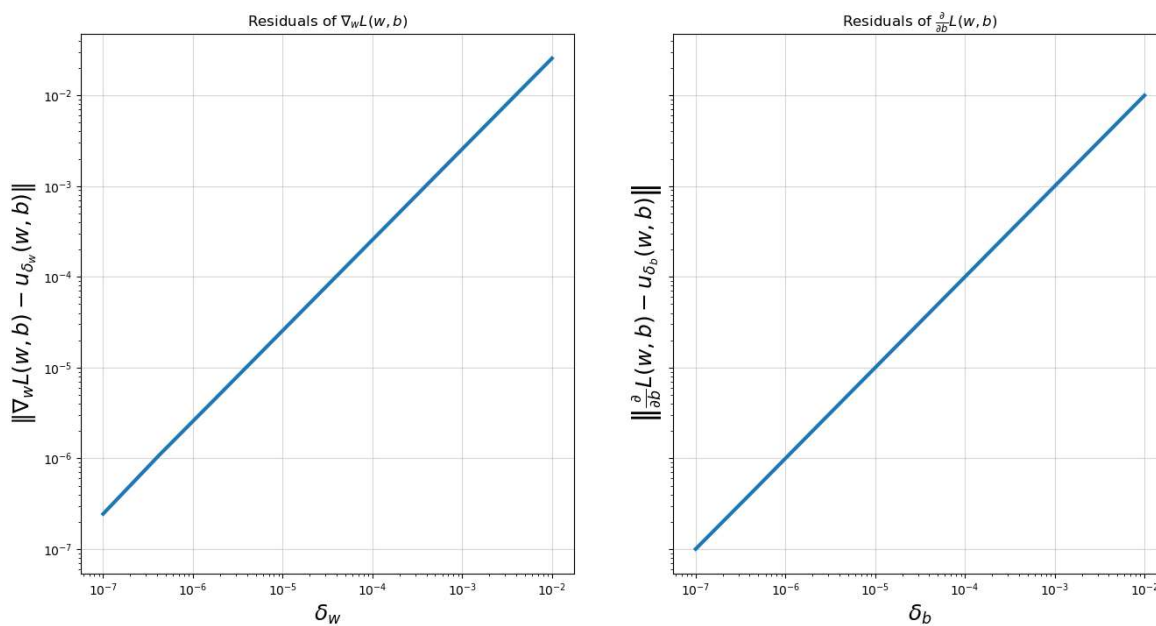
לילך ביטון – 205764517

(Q1)

$$\frac{\partial}{\partial b} L(\underline{w}, b) = \frac{\partial}{\partial b} \frac{1}{m} \sum_{i=1}^m (\underline{w}^T x_i + b - y_i)^2 = \frac{1}{m} \sum_{i=1}^m 2b + 2\underline{w}^T x_i - 2y_i = \frac{2}{m} \underline{1}_m^T (X\underline{w} + \underline{1}_m \cdot b - \underline{y})$$

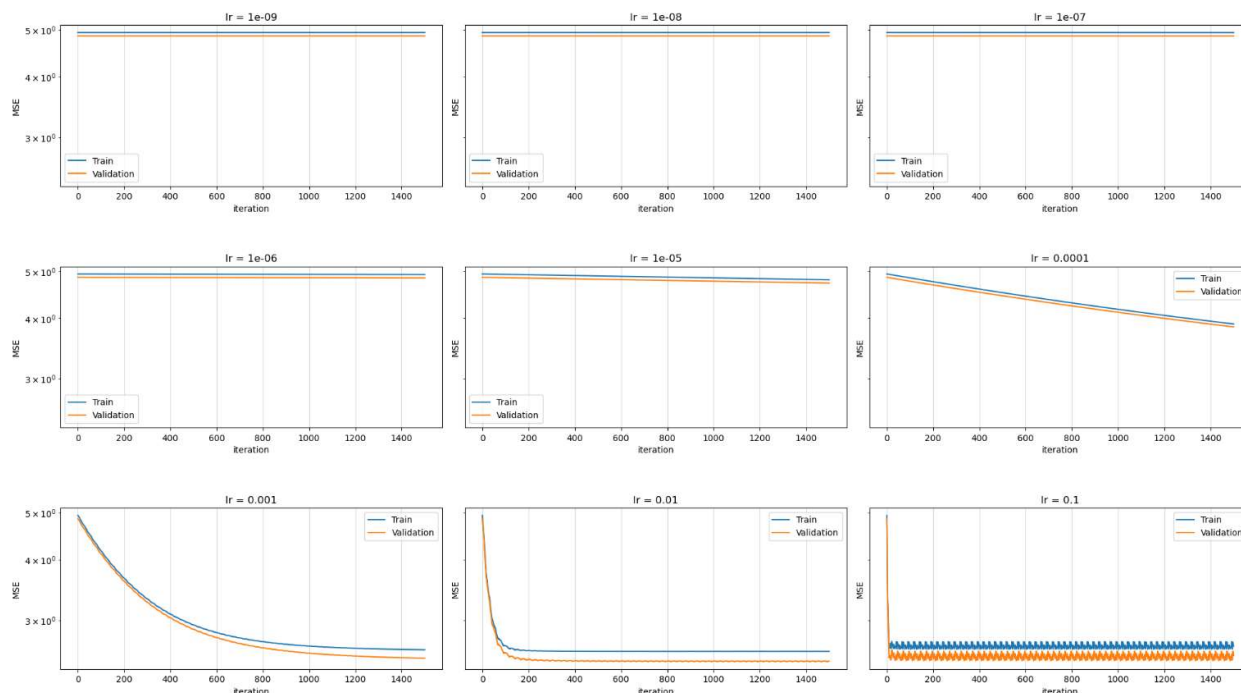
(Q2)

Residuals of analytical and numerical gradients



(Q3)

Train and Validation MSE for different learning rates



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.

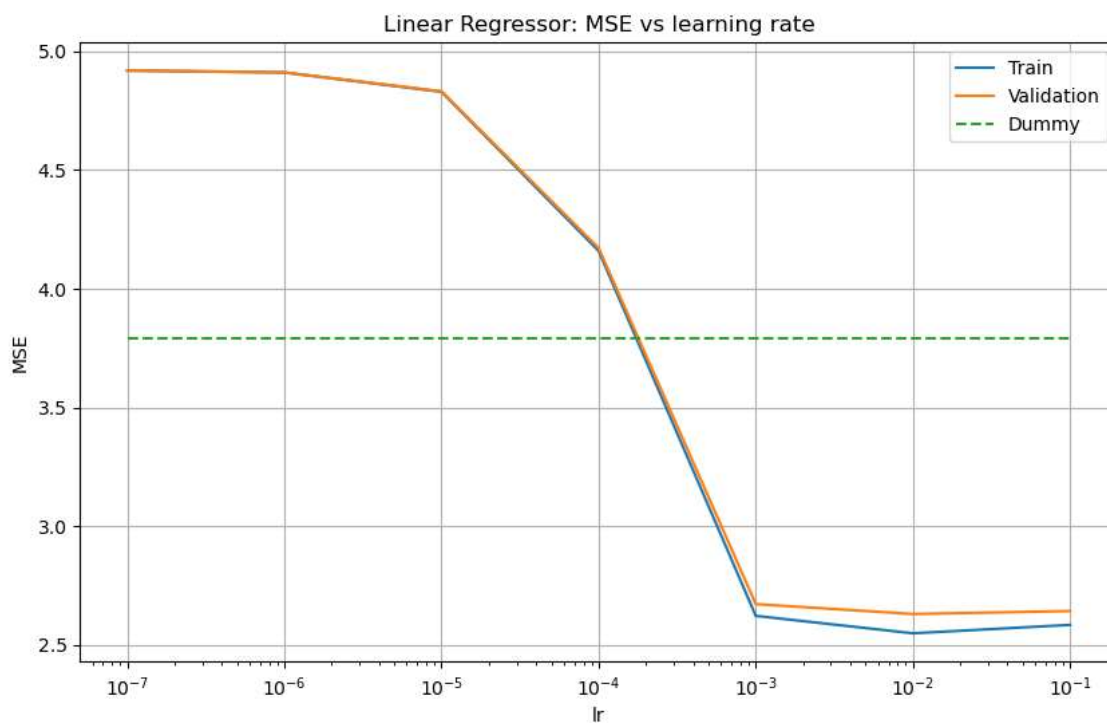
ניתן להסיק שקצב הלמידה האופטימלי הוא 0.01 שכן הוא נותן את הטעות הקטנה ביותר על קבוצת הולידציה עם רעש שאינו גדול מספיק כדי לפגוע בתוצאות המוצלחות שלו.

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)

Model	Section	Train MSE	Valid MSE
		Cross validated	
Dummy	2	3.787	3.793

(Q5)



קצב למידה אופטימלי: 0.01

שגיאת MSE על קבוצת אימון: 2.549

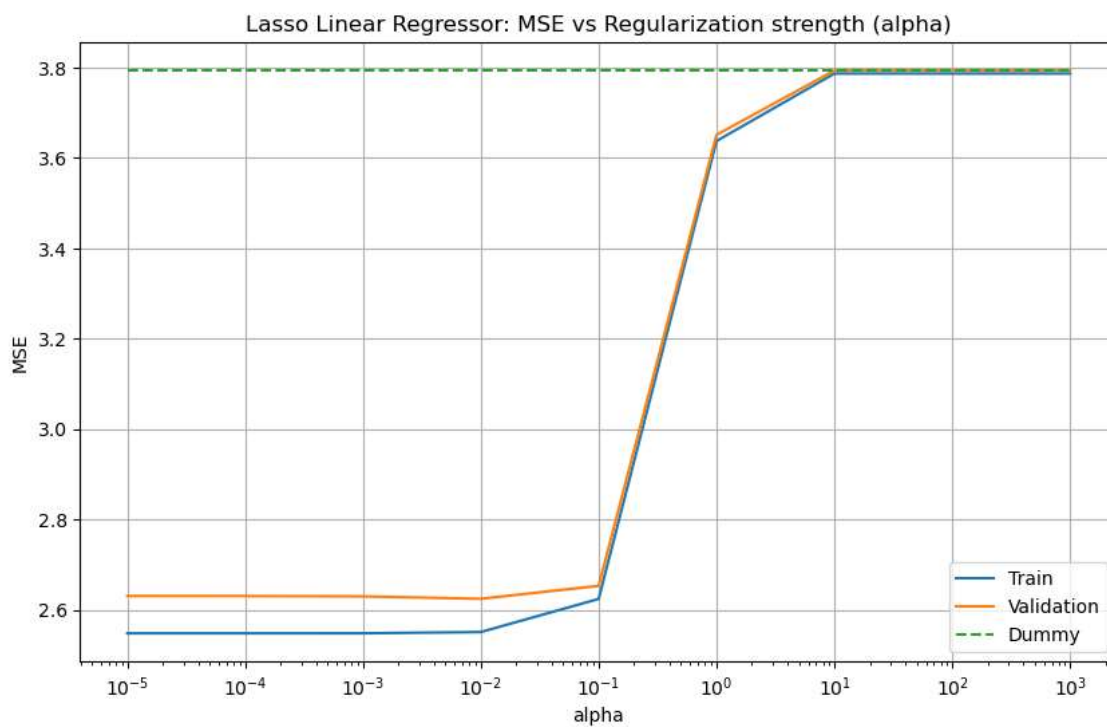
שגיאת MSE על קבוצת ולידציה: 2.629

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

(Q6)

הנרמול של הנתונים מתבצע לפי נתוני האימון. על כן, בהיעדר שגיאות נומריות, אי נרמול הנתונים לא היה פוגע בביצועי המודל הלינארי על נתוני האימון. גם הביצועים של מודל הDUMMY על נתוני האימון לא ישתנו, שכן לא מנרמלים את ערכי הcontamination_level (נתוני המטרה), וכתוצאה מכך לא יהיה שינוי בחיזוי של המודל.

(Q7)



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

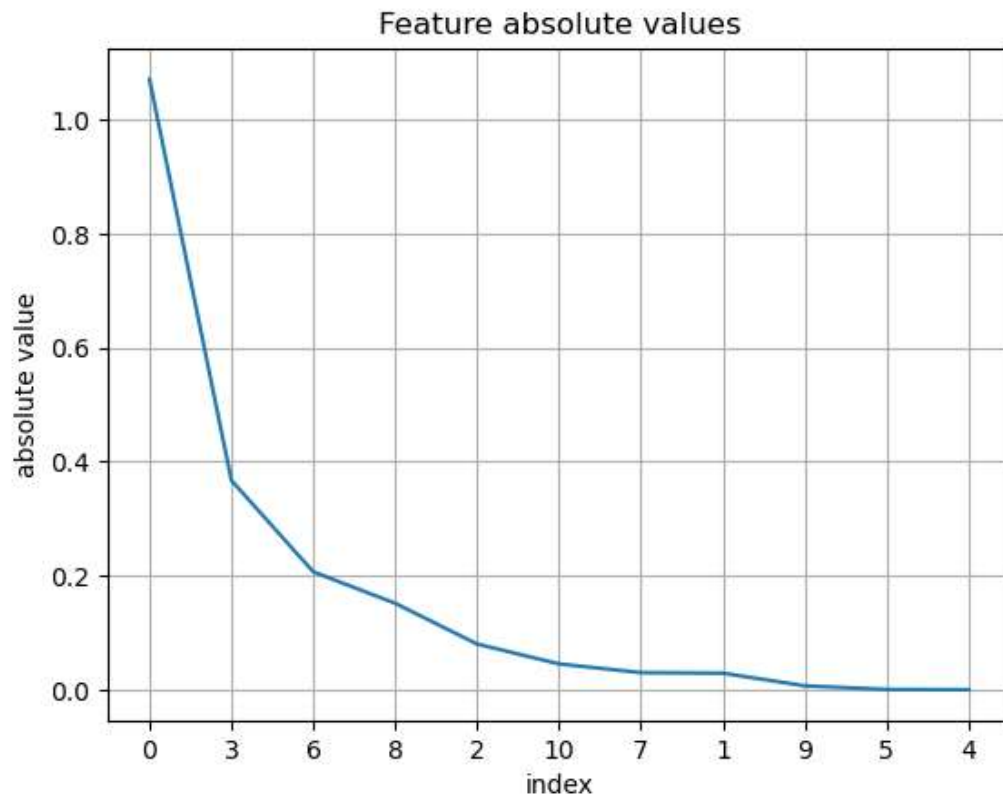
(Q8)

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

(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). לכן עוצמת המקדמים משקף את חוזק הקשר בין features למשתנה היעד.

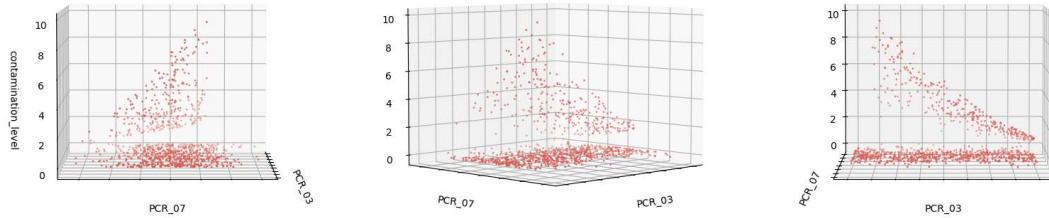
(Q12)

אי-נרמול הנתונים פוגע בביצועי המודל על נתוני האימון מאחר וכעת נוסף איבר רגולריזציה אשר "מעניש" features בעלי טווח ערכים גדול יותר באופן יחסי באמצעות הקטנת המקדם, ללא קשר למידת ההשפעה של כל features.

(Q13)

היינו מצפים למודל פחות דליל, בעל יותר מקדמים שאינם אפסיים. כתוצאה מכך שבניגוד ל-Lasso, Ridge משתמש ברגולריזציה L2. זו לרוב לא נותנת ערכים מסדר גודל 0 לאף אחד מה-features, אך עדיין קיימת העדפה ל-features מסוימים על ידי הקטנת ערכי המקדמים הפחות משמעותיים בהתאם למידת חשיבותם.

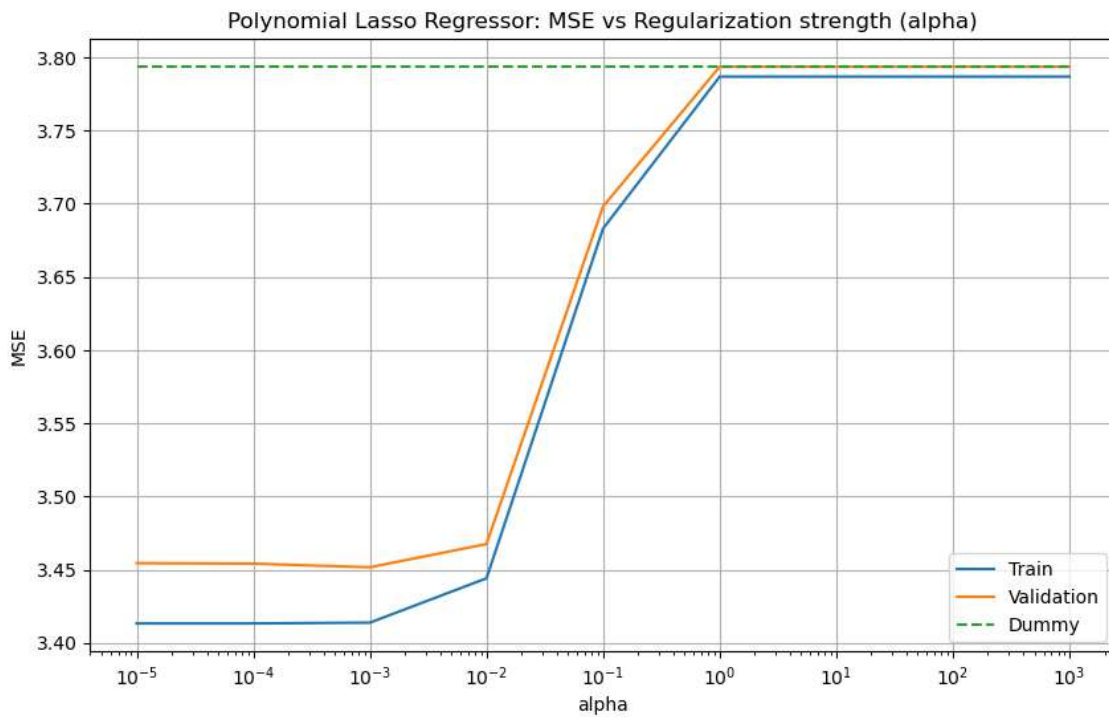
contamination_level as a function of the PCR_03 and PCR_07 features



(Q14)

מיפוי פולינומי משנה את מרחב הfeatures באופן שסביר שיצור אי-אזון 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.

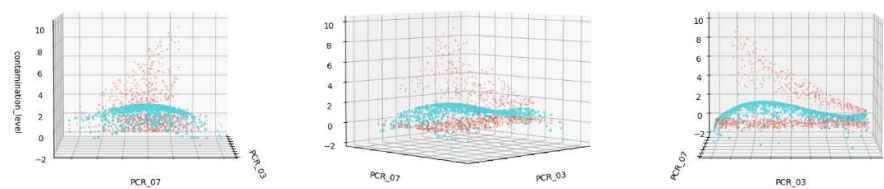
(Q15)



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

(Q16)

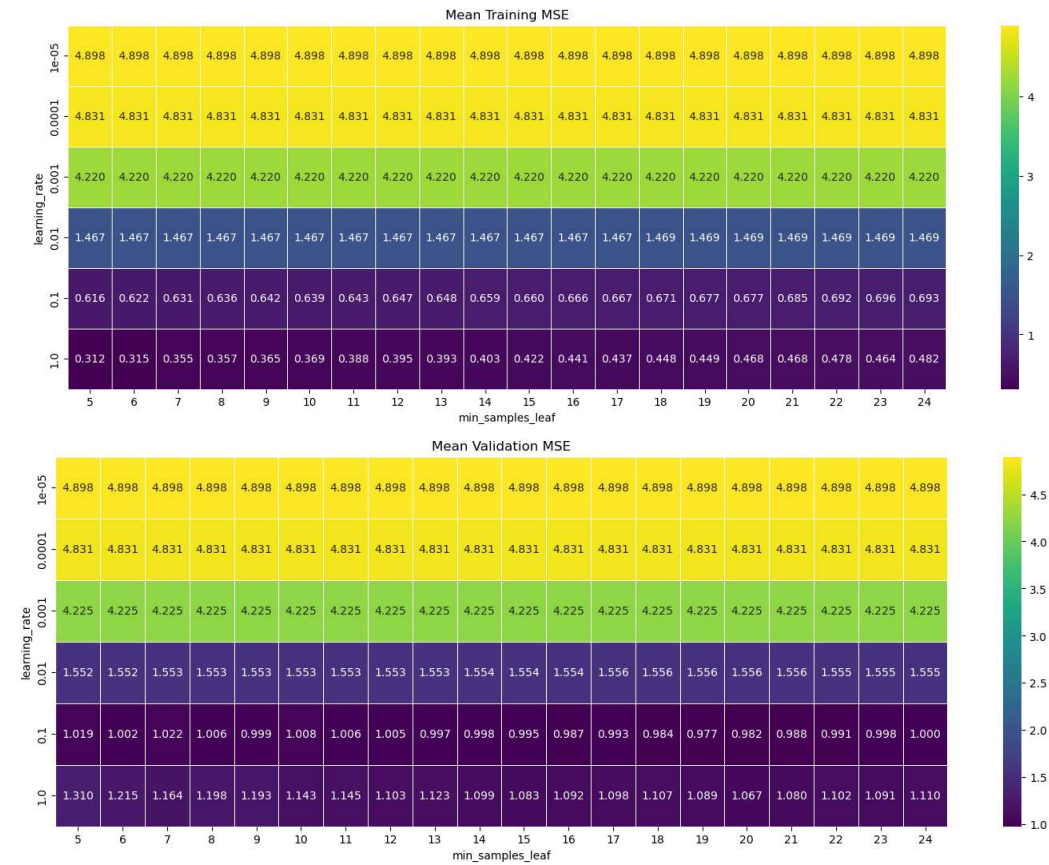
contamination_level as a function of the PCR_03 and PCR_07 features (red) with predictions (blue)



(Q17)

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

(Q18)



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)

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

(Q20)

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

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.