**Section 1**

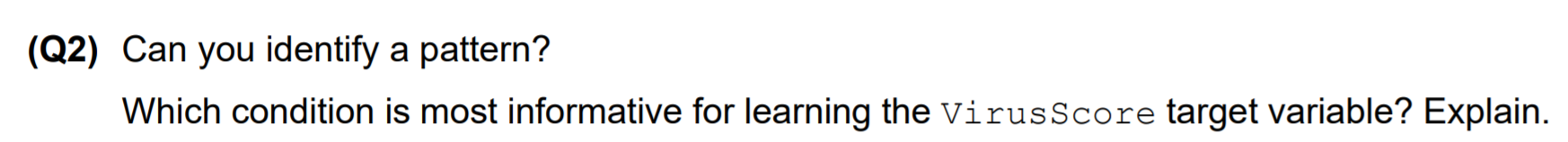
**שאלה 1**

הגרפים המתקבלים הם :

Chart, histogram

Description automatically generated

**שאלה 2**

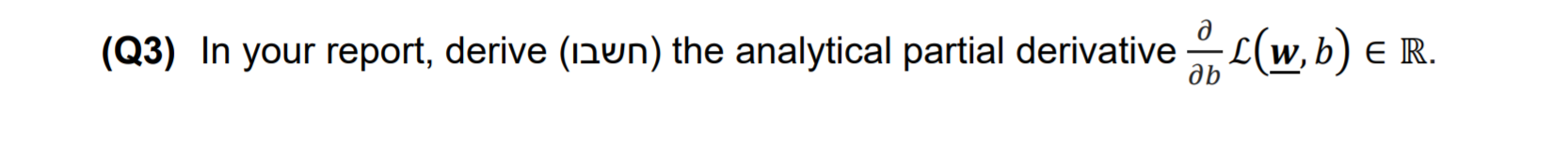


ניתן לראות שהקשר הברור ביותר בין VirusScore לאחד מן הגרפים הוא הגרף המתאר את התכונה

“A or not A” שמציג באופן ברור שההתפלגות של אנשים בעלי A נותת לכיוון ערכים גבוהים יותר של VirusScore מאשר אנשים ללא A .   
ולכן קישרנו בין אנשים עם סוג דם A+-,AB ל-VirusScore כאשר התכונה הבינארית שהוספנו בהתאם לתנאי זה היא :

**Section 2**

**שאלה 3**



Text

Description automatically generated

להלן חישוב הנגזרת,

**שאלה 4**

Chart, line chart

Description automatically generated

ניתן ללראות ב 2 הגרפים קוים לינארים ובשינהם ה delta משפיעה רק על השגיאה הנומרית ולכן בכל שמגדילים את ה delta , השגיאה נומרית גדלה וההפרש בין השגיאה האנליטית לשגיאה הנומרית גדל ביחס ישר.

**שאלה 5**

Diagram

Description automatically generated

ניתן לראות שהדיוק המירבי שהתקבל נמצא ב אך במקרה הזה ניתן לראות שישנו רעש ולא ניתן להבטיח את התשובה – למעשה ה GD מבצע קפיצות מסביב לנקודת המינימום ולכן אין התכנסות למינימום במקרה זה.

ולכן הדיוק המריבי שנתעדף הוא עבור שבמקרה זה הדיוק עלול להיות נמוך יותר אך נבטיח התכנסות.

lr size = 0.01, Best train loss = 0.007996124967117628, Best validation loss = 0.009689819450089773

lr size = 0.1, Best train loss = 0.007654105913138692, Best validation loss = 0.009199048010549963

**Section 3**

**שאלה 6**

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Section | Train MSE | Valid MSE |
|  |  | Cross Validated | |
| Dummy | 3 | 0.02001108412543403 | 0.020040073426649308 |

**Section 4**

**שאלה 7**

**Chart, line chart

Description automatically generated**

השגיאה הקטנה ביותר שהתקבלה היא : 0.008068282291301469

היא התקבלה עבור alpha אשר שווה ל: 2.848035868435799

**שאלה 8**

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Section | Train MSE | Valid MSE |
|  |  | Cross Validated | |
| Dummy | 3 | 0.02001108412543403 | 0.020040073426649308 |
| Ridge | 4 | 0.007899551580020458 | 0.008068282291301469 |

**שאלה 9**

התכונות שנמצאו עם הערכי coefficient הגדולים ביותר בערך מוחלט הם

1. Shortness of breath
2. Fever
3. PCR\_08
4. Household\_income
5. Blood\_viruse (our added feature according to blood type)

**שאלה 10**

הגרף שהתקבל הוא :

Chart, line chart

Description automatically generated

**Section 5**

**שאלה 11**

Chart, line chart

Description automatically generated

הערך שגיאה הטוב ביותר שהתקבל הוא : 0.008049

עבור alpha : 0.000259

**שאלה 12**

ישנו שוני בשיפוע ומיקומו, דבר הנגרם כתוצאה מכך שהפונקציה של Ridge משתמשת ב L2 norm אשר שומר על ערכים נמוכים ברכיבים של ווקטור w עם שונות נמוכה בין הערכים.

בניגוד לLasso המשתמש ב L1 norm אשר מבצע feature selection על ידי הגדרת ערך 0 לחלק מן הרכיבים בוקטור w – ובסופו של דבר לא להתייחס כלל לחלק מן התכונות ולכן הוא רגיש יותר לערכי alpha (כי בlasso למעשה הרכיבים השונים מ 0 בווקטור w מקפיצים מאוד את השגיאה בשונה מ Ridge הנותן לערכים המרכבים את וקטור w ערכים נמוכים הקרובים ל 0)

נשים לב שיש שוני קטן בערכי MSE המיטייבים בין המודלים LASSO , RIDGE, הדבר היחיד שמשתנה הוא מידת ההשפעה של alpha , אשר ניתן לראות לפי הגרף של LASSO שבשאלה 11 שההשפעה של alpha מתחילה באזור ערכים של לעומת ההשפעה של alpha על RIDGE לפי הגרף בשאלה 7 המשפיעה באזור alpha =

**שאלה 13**

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Section | Train MSE | Valid MSE |
|  |  | Cross Validated | |
| Dummy | 3 | 0.02001108412543403 | 0.020040073426649308 |
| Ridge | 4 | 0.007899551580020458 | 0.008068282291301469 |
| Lasso | 5 | 0.007919 | 0.008049 |

**שאלה 14**

חמש התכונות המשפיעות ביותר הן :

1. Shortness of breah
2. PCR\_08
3. Fever
4. Blood\_viruse
5. Household\_income

**שאלה 15**

Chart, line chart

Description automatically generatedניתן לראות שהערכים דומים וההתנהגות לינארית בשניהם \*להוסיף את RIDGE לגרף לשם השוואה\*

**שאלה 16**

יתרחש שינוי בשגיאת האימון ושגיאת הוולידאציה עקב הוספת תכונות נוספות, אשר מאפשרות לנו למידה למידה עם ביצועים טובים יותר.

נשים לב כי התכונות המשפיעות ביותר שקיבלנו הינן:

1)PCR\_08 blood\_viruse

2)cough shortness\_of\_breath

3)fever blood\_viruse

4)shortness\_of\_breath^2

5)shortness\_of\_breath

נשים לב שהתכונות המשפיעות ביותר הינן התכונות החדשות שנוספו לנו על ידי זה שהרחבנו את את המודל הלינארי למודל פולינומי ובכך הורדנו את שגיאת האימון ושגיאת הולידצטיה.

**שאלה 17**

נראה זאת מתמטית,

נגדיר את המשתנה כמשתנה בינארי כך ש אם אנחנו עונים על התנאי של שאלה אחת- כלומר מכיל A ו במידה ולא,

ולכן נוכל לומר כי:

כעת נשים לב לדבר הבא,

הינו פולינום ממעלה שנייה וזאת משום שכופלים כל תכונה בתכונה "מכיל את A" ולכן זה פולינום ממעלה שנייה.

הינו פולינום ממעלה ראשונה

הינו פולינום ממעלה ראשונה

*הינו קבוע*

*ולכן סהכ מקבלים כי ואת מה שנדרש.*

*המודל שיביא לתוצאות הטובות ביותר יהיה ב וזאת משום שכל פתרון שנמצא ב גם מוכל ב ,ולכן יכול להניב את אותם תוצאות של ואף יותר טובות על ה train ועל ה validation , נשים לב שמבדיקה שעשינו, קיבלנו את התוצאה הבאה –*

Text

Description automatically generated

**שאלה 18**

**Chart, line chart

Description automatically generated**

ערך השגיאה הטוב ביותר שהתקבל הוא : 0.0058869

עבור alpha : 2.2519719633

**שאלה 19**

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Section | Train MSE | Valid MSE |
|  |  | Cross Validated | |
| Dummy | 3 | 0.02001108412543403 | 0.020040073426649308 |
| Ridge | 4 | 0.007899551580020458 | 0.008068282291301469 |
| Lasso | 5 | 0.007919 | 0.008049 |
| Ridge polynomial | 6 | 0.004905504397315521 | 0.005886920556338286 |

**שאלה 20**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Section | Train MSE | Valid MSE | Test MSE |
|  |  | Cross Validated | | Retrained |
| Dummy | 3 | 0.02001108412543403 | 0.020040073426649308 | 0.021396513611111112 |
| Ridge | 4 | 0.007899551580020458 | 0.008068282291301469 | 0.007651409804232747 |
| Lasso | 5 | 0.007919 | 0.008049 | 0.00762251432378922 |
| Ridge Polynomial | 6 | 0.004905504397315521 | 0.005886920556338286 | 0.0066804870306428656 |

המודל בעל הביצועים הטובים ביותר הינו ה RIDGE POLYNOMAIL

נשים לב כי במודל זה , גם ה train mse וגם valid mse יחסית נמוכים – זה כנראה בעקבות זה שהרחבנו את מרחב האיפודזות בהרבה (במקום 21 תכונות, יש לנו 252 תכונות) מה מאשפר להגיע לביצועים טובים יותר.

נשים לב כי ב RIDGE ו LASSO אנחנו מקבלים בערך את אותה תוצאה על ה test

ועל dummy מקבלים את התוצאה הכי פחות טובה