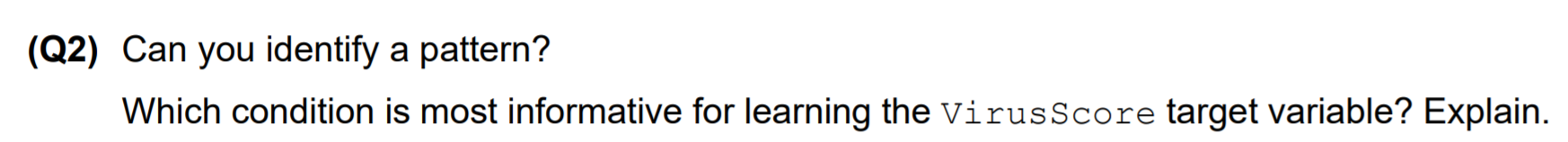
**שאלה 1**

הגרפים המתקבלים הם :

Chart, histogram

Description automatically generated

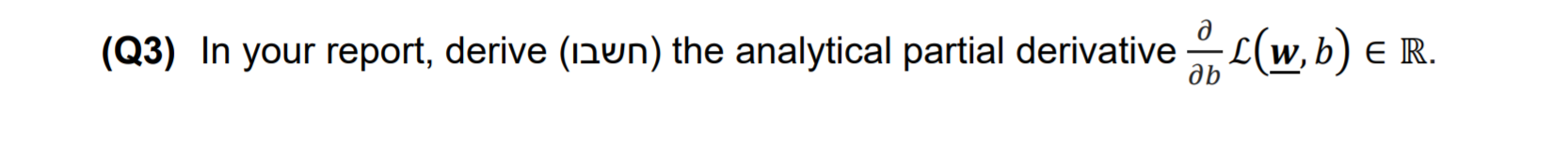
**שאלה 2**



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 :

**שאלה 3**



Text

Description automatically generated

**שאלה 4**

Chart, line chart

Description automatically generated

לפי \* לא גמור\*

**שאלה 5**

\*לסיים\* יש טעות

**שאלה 6**

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

**שאלה 7**

Chart, line chart

Description automatically generatedהשגיאה הקטנה ביותר שהתקבלה היא : 0.00815

היא התקבלה עבור alpha אשר שווה ל: 2.2570197196339215

**שאלה 8**

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Section | Train MSE | Valid MSE |
|  |  | Cross Validated | |
| Dummy | 3 | 0.02001108412543403 | 0.020040073426649308 |
| Ridge | 4 | 0.007982461953909743 | 0.008150394640191199 |

**שאלה 9**

התכונות שנמצאו עם הערכי coefficient הגדולים ביותר בערך מוחלט הם

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

**שאלה 10**

הגרף שהתקבל הוא :

Chart, line chart

Description automatically generated

**שאלה 11**

Chart, line chart

Description automatically generated

הערך שגיאה הטוב ביותר שהתקבל הוא : 0.008125

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)

ישנו שוני בשיפוע הלמידה אשר מתרחש קודם ובאופן מהיר יותר, דבר הנגרם כתוצאה מכך שהפונקציה של Ridge משתמש ב L2 norm דבר הגורם לכך הדעיכה מתרחשת מעקב מדעיכת המשקולות שנגרם כתוצאה מנרמול לפי ולכן דעיכה זו איטית יותר לעומת הדעיכה של Lasso לפי נרמול אשר גורמת לבחירת תכונות יותר ערכי 0 דבר הגורם ללמידה להיות מהירה יותר.

**שאלה 13**

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Section | Train MSE | Valid MSE |
|  |  | Cross Validated | |
| Dummy | 3 | 0.02001108412543403 | 0.020040073426649308 |
| Ridge | 4 | 0.007982461953909743 | 0.008150394640191199 |
| Lasso | 5 | 0.008002384807585961 | 0.008125396810908386 |

**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 generated

**שאלה 16**

**שאלה 17**

**שאלה 18**

**Chart, line chart

Description automatically generated**

ערך השגיאה הטוב ביותר שהתקבל הוא : 0.006103163121906443

עבור alpha : 2.2570197196339215

**שאלה 19**

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Section | Train MSE | Valid MSE |
|  |  | Cross Validated | |
| Dummy | 3 | 0.02001108412543403 | 0.020040073426649308 |
| Ridge | 4 | 0.007982461953909743 | 0.008150394640191199 |
| Lasso | 5 | 0.008002384807585961 | 0.008125396810908386 |
| Ridge polynomial | 6 | 0.005045018030948625 | 0.006103163121906443 |

**שאלה 20**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Section | Train MSE | Valid MSE | Test MSE |
|  |  | Cross Validated | | Retrained |
| Dummy | 3 | 0.02001108412543403 | 0.020040073426649308 | 0.021396513611111112 |
| Ridge | 4 | 0.007982461953909743 | 0.008150394640191199 | 0.0076482366030830575 |
| Lasso | 5 | 0.008002384807585961 | 0.008125396810908386 | 0.0076261886697763645 |
| Ridge Polynomial | 6 | 0.005045018030948625 | 0.006103163121906443 | 0.006632578340822194 |

המודל בעל הביצועים הטובים ביותר על ה- test set הוא המודל Ridge Polynomial , בעל השגיאה הקטנה ביותר על ה test set . אך בהשוואה ל Ridge , Lasso שהשגיאה שלהם על ה test set קטנה מהשגיאה על הסט וולידאציה מראה שאין over-fitting במודלים אלו מפני שאין התאמה טובה יותר לסט אימון מאשר לסט בדיקה. ובנוסף שהגדלת המידע לאימון שיפרה את התוצאות , בניקוד למודל עםת כונות פולינומאליות אשר הגדלת הסט אימון לא שיפרה את התוצאה.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Section | Train MSE | Valid MSE | Test MSE |
|  |  | Cross Validated | | Retrained |
| Dummy | 3 | 0.02058912914699978 | 0.020660172475179036 | 0.021396513611111112 |
| Ridge | 4 | 0.0076807939301818205 | 0.007947173041614703 | 0.0076482366030830575 |
| Lasso | 5 | 0.007692277903751298 | 0.007937073577412513 | 0.0076261886697763645 |
| Ridge Polynomial | 6 | 0.0048262937057628515 | 0.006131792231701941 | 0.006632578340822194 |