**Section 1: Linear regression implementation**

(Q1)

הפונקציה היא פונקציית מקרים.

נתייחס לנגזרת שלה לפי b לפי המקרים:

* עבור המקרה , הטעות היא ריבועית ביחס לb והפונקציה היא גזירה ביחס לb. אין נקודות אי גזירות בתחום זה ולכן הגדרת הנגזרת והsubderivative מתלכדות.
* עבור המקרה , הטעות היא מוגדרת באמצעות הזזה והכפלה בסקלר של פונקציית הערך המוחלט- . פונקציה זו היא גזירה ביחס לb בכל מקום מלבד באופן פוטנציאלי בנקודה שבה . עם זאת, מאחר שאנו בתחום שבו , הפונקציה גזירה באזורים אלו גם כן, ולכן הגדרת הנגזרת והsubderivative מתלכדות.
* עבור המקרים בהם , כלומר כאשר
* הפונקציה עדיין גזירה ביחס לb בנקודות אלו כיוון שהנגזרות מימין מתלכדות עם הנגזרות משמאל.

סה״כ מאחר והפונקציה גזירה ביחס לb על כל התחום נוכל לחשב את הנגזרות החלקיות בנפרד:

נחשב כל חלק בנפרד:

1. :
2. *Otherwise :*

*נקבל :*

(Q2)

נניח שיש לנו m זוגות של דוגמאות ותיוגים, כאשר כל דוגמה ממימד d כלומר :

נסמן את האלמנטים של דוגמה מסוימת, באופן הבא:

אזי נגדיר את המטריצה ונגדיר את וקטור העמודה .

עבור וקטור נגדיר פונקציית אינדיקטור באופן הבא

: .

עבור וקטור נגדיר פונקציית אינדיקטור באופן הבא

: , כאשר הוא האלמנט במיקום הi בוקטור z .

נשים לב כי

בנוסף, עבור וקטור נגדיר פונקציית אינדיקטור באופן הבא

: , כאשר הוא האלמנט במיקום הi בוקטור z .

אזי נוכל לבטא את באופן הבא:

נבדוק שהנוסחה חוקית:

* = היא מכפלה חוקית ונותנת לנו וקטור עמודה בעל m מימדים, לכן גם הסכום () כיוון שכל הוקטורים שייכים ל.
* הקלט ל הוא ואכן והפלט הוא גם כן וקטור
* מכפלת הדמרד (איבר – איבר) בין ובין

ולידית כי שניהם מאותו מימד והפלט הוא גם כן ב

* ולכן חוקי להכפיל בין המטריצה לווקטור

ונקבל פלט של וקטור בגודל המצופה לוקטור גרדיאנט כיוון שאנו גוזרים לפי .

* סה״כ הביטוי מחזיר לנו כפי שאנחנו רוצים.
* הביטוי הוא ולידי כיוון שראינו כי ולכן חישוב ההפרש בין 2 בוקטורים הוא לידי ומתקבל וקטור ששייך ל .
* לפי איך שהגדרנו את sign לוקטורים ממימד m, אזי הביטוי הוא ולידי ומחזיר וקטור ששייך ל .
* ולידי כיוון שאנו מבצעים מכפלה איבר-איבר בין שני וקטורים ממימד m.

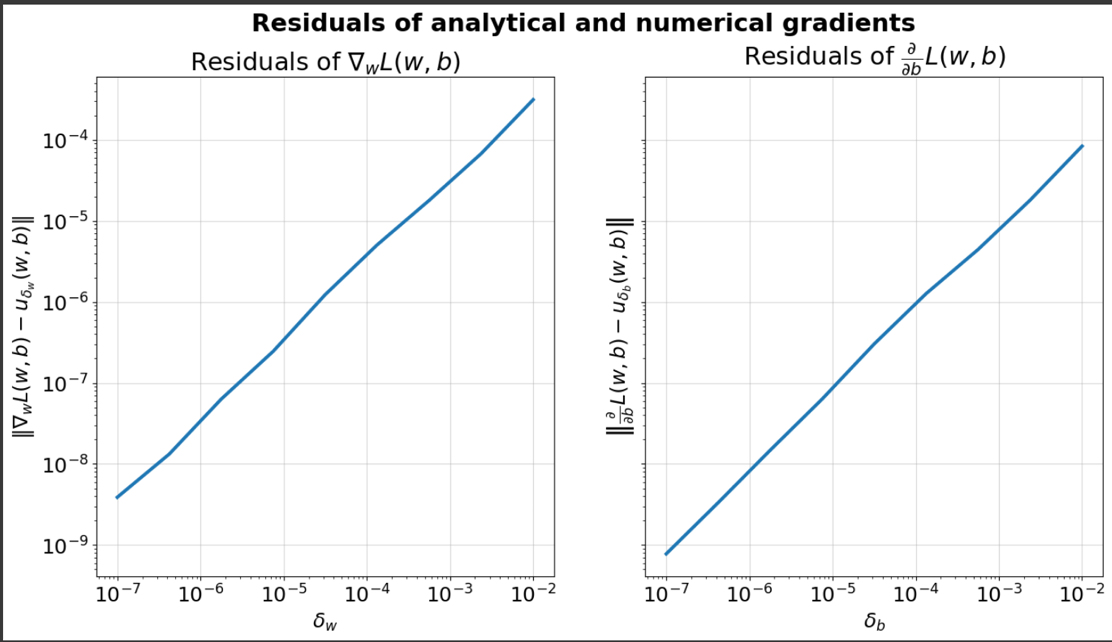
כמו קודם, גם המכפלה חוקית כיוון שכופלים בוקטור מימד m. נקבל פלט של וקטור בגודל המצופה לוקטור גרדיאנט כיוון שאנו גוזרים לפי .

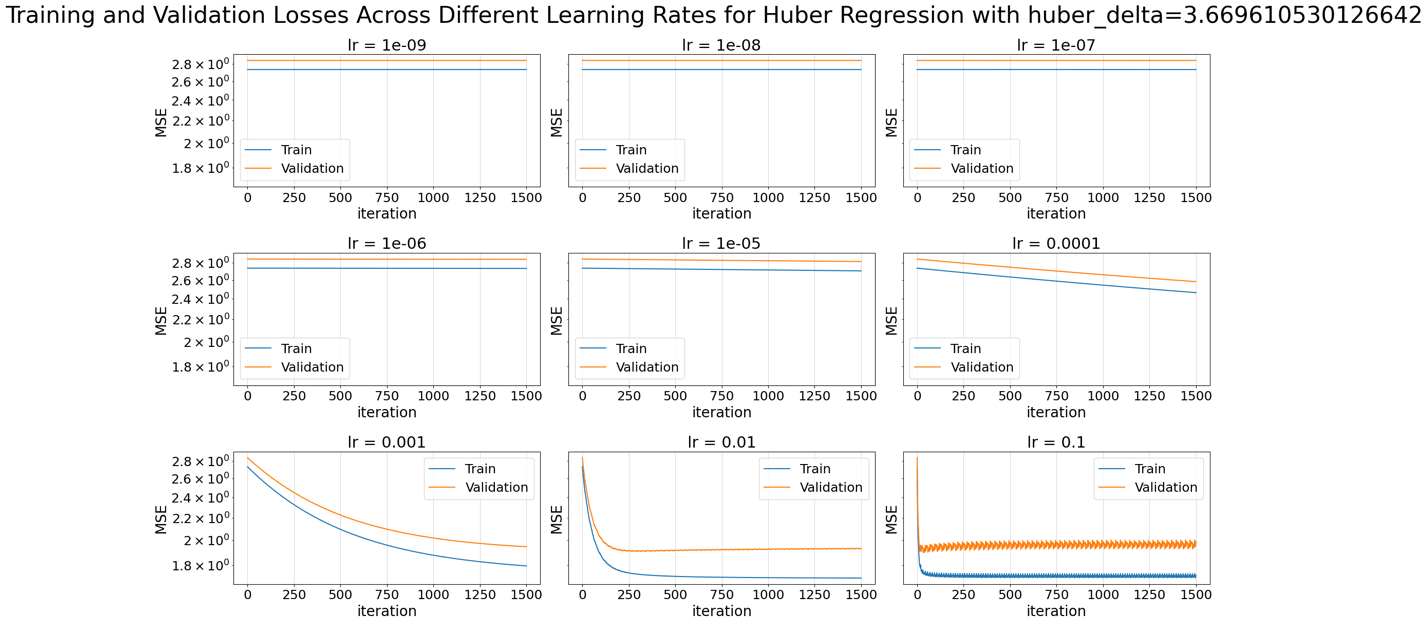
* סה״כ נקבל מהסכום של 2 הביטויים, 2 וקטורים ממימד d, לכן פעולת החיבור הזו היא אכן ולידית ואנחנו מקבלים וקטור גרדיאנט ממימד d כמצופה.

כעת נפתח נוסחה עבור

נבדוק שהנוסחה חוקית:

* = היא מכפלה חוקית ונותנת לנו וקטור עמודה בעל m מימדים, לכן גם הסכום () כיוון שכל הוקטורים שייכים ל.
* הקלט ל הוא ואכן והפלט הוא גם כן וקטור
* ולידי מאחר ומדובר במכפלה פנימית בין 2 וקטורים מימד m ובסופו של דבר הפלט מהמכפלה הזו הוא מספר כפי שהיינו מצפים.
* גם הביטוי הוא מספר , כמצופה.
* הביטוי הוא ולידי כיוון שראינו כי ולכן חישוב ההפרש בין 2 בוקטורים הוא לידי ומתקבל וקטור ששייך ל .
* לפי איך שהגדרנו את sign לוקטורים ממימד m, אזי הביטוי הוא ולידי ומחזיר וקטור ששייך ל .
* ולידי מאחר ומדובר במכפלה פנימית בין 2 וקטורים מימד m ובסופו של דבר הפלט מהמכפלה הזו הוא מספר כפי שהיינו מצפים.
* *סה״כ הביטוי מחזיר מספר ממשי.*
* *לכן הסכום בין שני הביטויים של הנגזרות לפי b חוקי ומחזיר לנו מספר ממשי , כמצופה.*

*(Q4 )*

*(Q5)*

*השתמשנו בhuber\_delta= ע״פ האלוריתם שייצרנו בסעיף 3.*

*ניתן להצדיק שרואים בגרפים כך:*

*\*עבור קצבי למידה נמוכים (1e-09-1e-5), ניתן לראות שבקושי יש ירידה בloss הן בtrain והן בvalidation מאחר וככל הנראה גודל הצעד בgardient קטן מדי ולכן ההתכנסות איטית מאוד ולא קורית ב1500 צעדים.*

*\* כאשר קצב הלמידה עולה לlr=0.0001 אנו רואים כי כן יש ירידה קטנה בloss, הן בvalidation והן בtrain, אך עדיין הירידה הזו קטנה ביחס לloss שמגיעים אליו בקצבי למידה גבוהים יותר. עבור קצב למידה הזה, אנו אכן מתקרבים לכיוון המינימום אך כמות הצעדים לא מספיקה עדין להתכנסות.*

*\* עבור קצבי הלמידה 0.001,0.01 אנו רואים כבר שחלה ירידה משמעותית בloss במהלך האימון והם מגיעים לערכי loss יחסית דומים בסוף האימון עבור שני הקצבים. עם זאת, הירידה עבור הקצב 0.001 היא מתונה יותר, בהשוואה ל0.01 , מאחר שגודל הצעד קטן יותר ולכן אנחנו מתקדמים לעבר המינימום בצורה איטית יותר.*

*\* עבור קצב הלמידה 0.1 , ההתכנסות היא מהירה מאד אבל יש הרבה יותר תנודות בערך הloss בתהליך האימון וגם בקבוצת ולידיה. התנהגות זו מצביעה על כך שקצב למידה זה גדול מדי ולכן האלגוריתם מפספס את המינימום בכל איטרציה.*

*\* עוד נשים לב, שערך הloss של קבוצת האימון ושל קבוצת הוולידציה דומה במהלך כל תהליך האימון של המסווג, מה שמצביע על כך ש*