|  |
| --- |
| **למידה עמוקה**  **תרגיל 2** |
| גל גרנות 315681593 |
| ניר טבת |
|  |
|  |
|  |
|  |

**שאלה 1**

הגדרנו את ה-Bayes Risk ואת הסיכון של בתור:

*נוכיח כי:*

*עבור:*

1. *הע"ע של המטריצה*

***הוכחה:***

מכיוון ש- ו- חסרי קורלציה (( איברי ה-cross יתאפסו תחת התוחלת ולכן נתעלם מהם. נחשב בנפרד את התוחלת עבור הנורמות של שני האיברים:

*הביטוי המתקבל הוא סקלר, או מטריצה מסדר ולכן שווה לעקבה של עצמו. נוכיח זהות קצרה בשימוש בזהות הציקלית של העקבה:*

*נשתמש בזהות עם ו-:*

*נפעיל את התוחלת ונשתמש בלינאריות העקבה והתוחלת, כאשר היא מטריצת הקוואריאנס של הוקטור האקראי שלפי הנתון היא :*

*נחזור לאיבר השני ב-\*, ונשתמש בעובדה ש- סימטרית ולכן שווה לשחלוף שלה:*

*נאחד את התוצאות שקיבלנו:*

*היא מטריצה סימטרית ולכן קיים לה לכסון אורתוגונלי עבור אורתוגונלית ו- אלכסונית. נקבל:*

*נשתמש בעובדה כי ובחילוף לכפל של מטריצת היחידה , נכפול משמאל וימין בהתאמה:*

*המטריצה אלכסונית ולכן ההופכית שלה היא אלכסונית עם כל האלמנטים ההפוכים:*

*כנדרש.*

**שאלה 2**

1. נתון ש-w נדגם בצורה אחידה מתוך Q אשר מכיל q ערכים. וכן נתון שכל רכיב בוקטור w נדגם באופן i.i.d לשאר הרכיבים.

אם w היה וקטור באורך 1, קל היה לראות שמתקבל:

לכן כיוון ש-w באורך k ומכיל k איברים i.i.d ניתן לומר כי:

1. *כדי שהביטוי יתקיים נצטרך שכל המשקולות יתאימו בין הרשתות. נשים לב שיש במטריצה , משקולות, וכן במטריצה יש משקולות. כמו שראינו בסעיף 1, לכל משקולת יש הסתברות של להתאים בין שתי הרשתות. כמו כן אנחנו יודעים שיש רק משקולות שאינן אפס בשכבה 1. סך הכל נקבל:*
2. *T הוא זמן העצירה של הרשת. ניתן להסתכל על T כעל רצף של ניסויי ברנולי עם פרמטר הצלחה שהוגדר בסעיף 2. כלומר T מתפלג גאומטרית.*

*לכן ניתן לומר:*

*נפעיל לוג על שני האגפים:*

1. *משילוב של שני הקירובים שנתונים ברמז והצבה שלהם בתוצאה של סעיף 3 נקבל:*

*נציב את התוצאה מסעיף 1 ונקבל:*

*נגדיר*

*כמו כן במקרה שלנו:*

*כעת נציב הכל במשפט 2 ונקבל:*

*כנדרש.*

1. *ביטוי 4 עבור משקולות שיכולות לקבל q ערכים שונים נקבל:*

*נקבל ביטוי שגדול יותר מהביטוי בסעיף 3 ולכן ההכללה של ביטוי 4 תהיה פחות טובה.*

***שאלה 3***

1. בכל הגרפים הנקודה הקריטית היא הנקודה שבה ה-test loss מפסיק לעלות ומתחיל לרדת שוב.
2. הגרף הכחול והירוק הם שני מודלים שונים, הכחול הוא מודל שמתרגם מגרמנית לאנגלית והירוק הוא מודל שמתרגם מאנגלית לצרפתית. אנו יכולים לראות כי ה-test loss לאחר הגדלת גודל המודל מתחיל לעלות בשלב מסוים ולאחר גודל מסוים מתחיל לרדת, בנוסף ניתן לראות כי ה – train loss מונוטוני יורד עם הגדלת גודל המודל ולכן זהוmodel-wise double decent מכיוון שהאזור הקריטי מופיע ב-test loss.
3. אנו יכולים לראות כי זה model-wise double decentככל שכמות הפרמטרים במודל גדלה כך השגיאה קטנה, עבור מודלים גדולים, בניגוד לשגיאה הקלאסית. הנקודה הקריטית מתקבלת בנק' מקסימום בה הגרף מתחיל לרדת לאחר העלייה )בערך ברוחב 10 פרמטרים(
4. אנו יכולים לראות כי epoch-wise double decentכיוון שעבור המודל הגדול (האדום) אנחנו רואים שעלייה בכמות הEPOCHS גורמת לירידה בשגיאה.

**שאלה 4**

1. ההתפלגות של W סימטרית -> ההתפלגות של U סימטרית:

ניתן לראות כי הביטוי המחושב זהה לחישוב השונות של u אך החלק השלילי מאופס. מכיוון שההתפלגות של u סימטרית, החלק השלילי תרומה שווה לתרומה של החלק החיובי ונקבל חצי מהשונות של u.

*תחת ההנחה שהשונות של היא 1.*

1. *לפי תאוריית הגבול המרכזי:*

*נדיר משתנה Z שמתפלג נורמלי עם תוחלת 0 ושונות 1.*

וההמשך זהה לסעיף 1.

***שאלה 5***

*צד ראשון: אם מתקיים אז היא equivariant:*

*כאשר השוויון האחרון נובע מהנתון ומכך ש פועלת על כל אלמנט בנפרד.*

*מ- equivarianceשל f ל- נקבל:*

*ובסך הכול קיבלנו equivariance של f ל-H.*

*צד שני: אם היא equivariant אז מתקיים*

*מחד-חד ערכיות של ניתן לומר שחייב להתקיים שוויון בארגומנטים:*

*ומכאן ניתן להסיק שלכל איבר בW מתקיים:*

***שאלה 6***

|  |  |  |
| --- | --- | --- |
| Layer | Output\_dim | Number of parametes |
| INPUT | 224X224X3 | 0 |
| CONV3-64 | 224X224X64 | (3X3X3+1)X64=1792 |
| ReLU | 224X224X64 | 0 |
| POOL2 | 112X112X64 | 0 |
| CONV3-128 | 112X112X128 | (3X3X64+1)X128=73856 |
| ReLU | 112X112X128 | 0 |
| POOL2 | 56X56X128 | 0 |
| CONV3-256 | 56X56X256 | (3X3X128+1)X256=295168 |
| ReLU | 56X56X256 | 0 |
| CONV3-256 | 56X56X256 | (3X3X256+1)X256=590080 |
| ReLU | 56X56X256 | 0 |
| POOL2 | 28X28X256 | 0 |
| CONV3-512 | 28X28X512 | (3X3X256+1)X512 = 2359808 |
| ReLU | 28X28X512 | 0 |
| CONV3-512 | 28X28X512 | (3X3X512+1)X512 = 4718592 |
| ReLU | 28X28X512 | 0 |
| POOL2 | 14X14X512 | 0 |
| CONV3-512 | 14X14X512 | (3X3X512+1)X512 = 4718592 |
| ReLU | 14X14X512 | 0 |
| CONV3-512 | 14X14X512 | (3X3X512+1)X512 = 4718592 |
| ReLU | 14X14X512 | 0 |
| POOL2 | 7X7X512 | 0 |
| FC-4096 | 4096X1 | (7X7X512+1)X4096 = 102764544 |
| FC-4096 | 4096X1 | (4096+1)X4096 = 16781312 |
| FC-1000 | 1000X1 | (4096+1)X1000 = 4097000 |
| SOFTMAX | 1000X1 | 0 |

1. *מספר הפרמטרים הכולל: 114,119,336*
2. *החלק היחסי של הפרמטרים של שכבות ה-FC:*