**חלק מעשי**

**חלק תיאורטי**

1. פונקציות לינאריות ואפיניות:

יהיו ו- פונקציות לינארית כאשר , ו-, . פונקציית ההרכבה נתונה לפי

*נסמן כאשר ונקבל ש- היא פונקציה לינארית , כנדרש.*

באותו אופן אם ו- , כאשר , הרכבה תיתן

נסמן ו- כאשר ו- ונקבל ש- היא פונקציה אפינית כנדרש

1. *אינפי של gradient descent:*
2. תנאי העצירה של הצעד יכול להיות למשל:

* היפר-פרמטר טריוויאלי שמגדיר מספר איטרציות (מספר epoch-ים):

For :

* לולאה שעוצרת כאשר ה-גרדיאנט לא משתנה, או קטן מאיזשהו היפר-פרמטר מסוים :

While :

* במקרים אידיאלים/ריאליזבילי הלולאה תעצור כשהגענו לנקודה סטציונרית
* ניתן לשלב את הטענות, למשל להגדיר מספר epoch-ים שבו נעצור בכדי להימנע מלולאות אינסופיות או ארוכות במיוחד (מה שיכול לקרות למשל כאשר הצעד איננו אידיאלי)

1. נשתמש בפיתוח טיילור לפונקציית ה-loss כדי לקבל תנאי לקבלת נקודה סטציונרית

בנקודה סטציונרית מתקיים ומכאן נשארנו עם

אם היא נקודת מקסימום, כל הזזה תביא אותנו לנקודה עם ערך נמוך יותר, דהיינו . לעומת זאת, אם היא נקודת מינימום, משיקולים דומים תקיים . כלומר בסה"כ התנאי הנדרש לסיווג הוא

1. *פונקציית הפסד למרחב מעגלי:*

לכל מרחק בין שני ערכים יש שני פירושים – הראשון הוא אורך הקשת הקצרה והשני הוא אורך הקשת הארוכה. לצורך חישוב מרחק מינימלי נרצה תמיד לבחור במרחק שמיוצג על ידי אורך הקשת הקצרה, ולשם כך נבחר במינימום בין המרחק עצמו (בזווית) לבין פחות המרחק. פורמלית נרשום:

בהינתן מרחב דגימה נגדיר פונקציית מרחק

1. נגזרות חלקיות*:*
2. *נסמן לשם נוחות ו-*

*ולכן בגזירה לפי :*

1. *נגזור לפי כלל שרשרת (כאשר סימנו )*

*נחזור שוב על הפעולה עבור המוכפל הימני*

*כלומר*

*נחזור על הפעולה באינדוקציה:*

*או בקיצור:*

1. *נגזור לפי כלל שרשרת, כאשר הגזירה יופיעו סכום הנגזרות החלקיות לפי כל פרמטר.*

*נשתמש בהגדרת הדיפרנציאל:*

*כלומר באופן מפורש:*

*נעשה את אותו תהליך עבור :*

*כלומר כעת הביטוי שבידינו הוא*

*נמשיך זאת באופן אינדוקטיבי, כאשר הנגזרת של היא (לכל )*

*ולבסוף נישאר עם*

1. *נסמן , וכעת השאלה היא*

*את הנגזרת של נחשב מפורשות*

*נסמן באופן דומה .*

*את הנגזרת של נחשב מפורשות*

*לבסוף נאחד את כל הביטויים:*

1. *האנטרופיה היחסית היא גודל אי שלילי*

*בהינתן התפלגויות מעל אותו מרחב הסתברות , מגדירים*

*נשתמש בקירוב הלינארי של , ובפרט בכך שלכל מתקיים :*

*כאשר המעבר האחרון נובע מכך שסכימת ערך פונקציית הסתברות על כל המרחב נותנת תמיד .*

*כמו כן, קל לראות שכאשר מתקיים , שהרי*

1. *האנטרופיה היחסית היא פונקציה קמורה*

*בהינתן נקודות שמייצגות כל אחת זוג של פונק' הסתברות נראה שמתקיים*

*לשם ההוכחה נציג את הלמה הבאה:*

*כאשר סימנו , ונדרש*

*הוכחת הלמה:*

*נסמן ונקבל*

*כעת נסמן*

*נשכתב את הביטוי של האנטרופיה היחסית במונחי :*

1. *התיאוריה של בCybenko ו-Hornik עבור פונקציית*

*התיאוריה של בCybenko ו-Hornik עוסקת בפונקציות חסומות, ואילו איננה חסומה. עם זאת, הפרש של פונקציות כן ייתן לנו פונקציה רציפה וחסומה כפי שנדרש מהתיאוריה. כלומר אם נסתכל על זוגות איברים מתוך , כל צירוף של זוג איברים כנל שהוא מהצורה* מהווה פונקציה חסומה ורציפה, ועל כן בסך הכל הדרישה של התיאוריה מתקיימת. ביתר פירוט, קיימת בחירה של ו- כך שהפרש ה-reluים היא פונקציה חסומה (למשל אם לכל זוג ו, במצב זה נקבל פונקציה מונוטונית עולה חלש כך ש ו-)

1. *מבנה הרשת העמוקה עם התחשבות הסימן הוא של זרועות במקום , כאשר הזרוע החדשה תפעל בדומה לזרוע העליונה , רק עבור ביטויים עם מקדם שלילי:*

* *בזרוע התחתונה נעביר קדימה את הקלט המוזז*
* *בזרוע האמצעית אנו נחשב את הפונקציה האפינית שרלוונטית לכל נוירון*
* *בזרוע העליונה אנחנו סוכמים את כל הפלטים של שכבות הביניים עבורם והסכום יהיה רק על גורמים חיוביים*
* *בזרוע החדשה אנו סוכמים את כל הפלטים של שכבות הביניים עבורם והסכום יהיה רק של גורמים שליליים*

*בסך הכל הוספנו עוד נוירונים, לכן בסך הכל נשארנו עם נוירונים*