**עבודה מספר 3 -זיהוי זווית באמצעות *Backpropagation***

מגישים :נתנאל כהן , נטע דג'ן - ,אפרים פבריקר , אלון דור (קבוצה מספר 6).

נראה את הרצת הרשתות בצורה סדרתית ורנדומלית (רשת בעלת שכבה חסויה אחת ורשת בעלת שתי שכבות חסויות) בשלושה מקרים שונים ונראה איזה צורת לימוד עבור כל רשת עדיפה ,איזו רשת עובדת טוב יותר ומהי הסיבה.

נראה הרצה בצורה הסדרתית עבור 7 קבוצות למידה (בשתי הרשתות):

|  |  |
| --- | --- |
| **צעדי אימון** | **אחוזי הצלחה** |
| **1** | **14** |
| **20** | **42** |
| **30** | **61** |
| **40** | **71** |
| **42** | **80** |

**שכבה חסויה אחת שתי שכבות חסויות**

|  |  |
| --- | --- |
| **צעדי אימון** | **אחוזי הצלחה** |
| **1** | **14** |
| **10** | **52** |
| **19** | **80** |
|  |  |
|  |  |

נראה הרצה בצורה הסדרתית עבור 12 קבוצות למידה (בשתי הרשתות):

|  |  |
| --- | --- |
| **צעדי אימון** | **אחוזי הצלחה** |
| **1** | **25** |
| **20** | **61** |
| **40** | **63** |
| **50** | **72** |
| **51** | **80** |

|  |  |
| --- | --- |
| **צעדי אימון** | **אחוזי הצלחה** |
| **1** | **33** |
| **20** | **47** |
| **30** | **50** |
| **40** | **61** |
| **52** | **80** |

**שכבה חסויה אחת שתי שכבות חסויות**

נראה הרצה בצורה הסדרתית עבור 18 קבוצות למידה (בשתי הרשתות):

**שכבה חסויה אחת שתי שכבות חסויות**

|  |  |
| --- | --- |
| **צעדי אימון** | **אחוזי הצלחה** |
| **1** | **35** |
| **20** | **33** |
| **30** | **37** |
| **50** | **64** |
| **64** | **81** |

|  |  |
| --- | --- |
| **צעדי אימון** | **אחוזי הצלחה** |
| **1** | **31** |
| **10** | **35** |
| **20** | **50** |
| **30** | **59** |
| **49** | **81** |

נראה הרצה בצורה הרנדומלית עבור 7 קבוצות למידה (בשתי הרשתות):

|  |  |
| --- | --- |
| **צעדי אימון** | **אחוזי הצלחה** |
| **1** | **23** |
| **10** | **38** |
| **20** | **57** |
| **30** | **76** |
| **34** | **80** |

**שכבה חסויה אחת שתי שכבות חסויות**

נראה הרצה בצורה הרנדומלית עבור 12 קבוצות למידה (בשתי הרשתות):

|  |  |
| --- | --- |
| **צעדי אימון** | **אחוזי הצלחה** |
| **1** | **30** |
| **10** | **58** |
| **20** | **72** |
| **30** | **77** |
| **33** | **83** |

**שכבה חסויה אחת שתי שכבות חסויות**

|  |  |
| --- | --- |
| **צעדי אימון** | **אחוזי הצלחה** |
| **1** | **30** |
| **10** | **38** |
| **20** | **55** |
| **30** | **77** |
| **38** | **80** |

נראה הרצה בצורה הרנדומלית עבור 18 קבוצות למידה (בשתי הרשתות):

**שכבה חסויה אחת שתי שכבות חסויות**

|  |  |
| --- | --- |
| **צעדי אימון** | **אחוזי הצלחה** |
| **1** | **23** |
| **10** | **66** |
| **19** | **80** |
|  |  |
|  |  |

|  |  |
| --- | --- |
| **צעדי אימון** | **אחוזי הצלחה** |
| **1** | **25** |
| **20** | **48** |
| **40** | **62** |
| **50** | **77** |
| **52** | **85** |

|  |  |
| --- | --- |
| **צעדי אימון** | **אחוזי הצלחה** |
| **1** | **31** |
| **20** | **46** |
| **40** | **70** |
| **50** | **68** |
| **52** | **85** |

**הסברים לכל מקרה:**

על פי הנתונים מהטבלאות אפשר לשים לב שבכל מקרה יש עדיפות לרשת עם **השכבה החסויה אחת**, אך על פי מה שלמדנו הרשת עם שני השכבות החסויות צריכה לעבוד בצורה יותר אפקטיבית, דבר זה קורה בעקבות כמות הקטנה של קבוצות הלמידה שאנחנו מתעסקים איתם.

הרשת עם השני שכבות כנראה תעבוד בצורה יותר אפקטיבית אם נכניס עוד נתונים.

במקרה פשוט זה הרשת עם השכבה האחת עובדת בצורה יותר אפקטיבית בעקבות הכמות המעטה של הקבוצות.

בנוסף אפשר לשים לב לכך שהוספת שכבה אחת בודדה אינה תורמת בצורה משמעותית לזמני הריצה נוכל לראות בכמה מקרים שזמני הלמידה הם כמעט אותו דבר בין שני הרשתות (**הרצה סדרתית ל- 12 קבוצות, הרצה רנדומלית ל -18 קבוצות)**, אולי כאשר נוסיף כמה שכבות חסויות וניקח קבוצת למידה יותר גדולה נוכל לראות שינוי משמעותי בזמני הלמידה של הרשתות ונוכל להשוות הבדלים כי זמני הריצה יהיו יותר ארוכים ולא יסתיימו לאחר כמות מעטה של סיבובים.

**עדיפות לצורת תהליך הלמידה:**

אפשר לראות שתהליך הלמידה הרנדומלי עובד בצורה יותר טובה מאשר הצורה הסדרתית בעקבות שצעדי האימון שהוא לוקח להגיע לאחוזי הצלחה שהצבנו לו שהינם 80% יותר מהר מאשר תהליך הלמידה הסדרתי.

על פי דעתנו אנחנו רוצים שתהליך הלמידה של הרשת לא יהיה תלוי בסדר מסוים של המטריצות בקבוצות כי אז עקומת הלמידה תהיה תלויה בסדר שאתה מציב לו את הקלט, ולכן כשאר אנחנו "מבלגנים" את הסדר של המטריצות בקבוצות הלמידה אפשר להבין שהרשת תלמד לקבל קבוצות קלטים בכל סדר ולא תהיה תלויה בסדר מסוים של הקלט וככה תהליך הלמידה יהיה יותר אפקטיבי ואנחנו רואים את זה בתוצאות של כל המקרים.

**טסטים:**

**טסטים ל-7 קבוצות למידה :**

**שכבה חסויה אחת שתי שכבות חסויות**

**תהליך למידה סדרתי**

|  |  |
| --- | --- |
| **טסט סדרתי** | **טסט רנדומלי** |
| 33% | 33% |

|  |  |
| --- | --- |
| **טסט סדרתי** | **טסט רנדומלי** |
| 33% | 33% |

**תהליך למידה רנדומלי**

|  |  |
| --- | --- |
| **טסט סדרתי** | **טסט רנדומלי** |
| 33% | 33% |

|  |  |
| --- | --- |
| **טסט סדרתי** | **טסט רנדומלי** |
| 33% | 33% |

**טסטים ל-12 קבוצות למידה :**

**שכבה חסויה אחת שתי שכבות חסויות**

**תהליך למידה סדרתי**

|  |  |
| --- | --- |
| **טסט סדרתי** | **טסט רנדומלי** |
| 33% | 33% |

|  |  |
| --- | --- |
| **טסט סדרתי** | **טסט רנדומלי** |
| 33% | 33% |

**תהליך למידה רנדומלי**

|  |  |
| --- | --- |
| **טסט סדרתי** | **טסט רנדומלי** |
| 33% | 33% |

|  |  |
| --- | --- |
| **טסט סדרתי** | **טסט רנדומלי** |
| 66% | 66% |

**טסטים ל-18 קבוצות למידה :**

**שכבה חסויה אחת שתי שכבות חסויות**

**תהליך למידה סדרתי**

|  |  |
| --- | --- |
| **טסט סדרתי** | **טסט רנדומלי** |
| 66% | 66% |

|  |  |
| --- | --- |
| **טסט סדרתי** | **טסט רנדומלי** |
| 33% | 33% |

**תהליך למידה רנדומלי**

|  |  |
| --- | --- |
| **טסט סדרתי** | **טסט רנדומלי** |
| 66% | 66% |

|  |  |
| --- | --- |
| **טסט סדרתי** | **טסט רנדומלי** |
| 33% | 33% |

**הסברים על טסטים**

אחוזי ההצלחה שמופיעים בטבלאות אינם קבועים, אחוזי ההצלחה של הטסט משתנים בין 0%-66% בקבוצות 12 ו- 18, בעיקר אחוזי ההצלחה מושפעים ע"י המשקלים הסינפטים כאשר אנחנו נריץ את התוכנית שוב ושוב נקבל תוצאות שונות לכן אין אפשרות להגיד שהטסט תמיד יצליח, זה תמיד יהיה תלוי באיתחול המשקלים הסינפטים וגודל קבוצת הלמידה, כאשר אנחנו ניתן לקבוצה יותר קווים שבורים, תהליך הלמידה יהיה יותר מדויק ונוכל להצליח בבדיקות בצורה יותר מדויקת.

לעיתים יותר קרובות הרשת עם השני שכבות החסויות מקבלת אחוזי הצלחה גבוהים יותר מאשר הרשת עם השכבה האחת ולפי מה שהסברנו מקודם לרשת עם שני שכבות החסויות זמן הלמידה הוא ארוך יותר אך עפ"י התוצאות של הטסטים הוא יותר מדויק מאשר שכבה חסויה אחת.

**מקרים נוספים שאפשר להתייחס אליהם:**

נוכל לראות מקרים מיוחדים שבהם הקוד לא יודע לעבוד איתו:

לדוגמא : ניקח את המקרה שאנחנו מאתחלים את כל המשקלים הסינפטים במספרים בין (0,1) , כאשר אנחנו עובדים עם מספרים חיובים צריך לשמור על העולם שאנחנו עובדים בו כדי שנוכל לקבל תהליך למידה נכון, במקרה זה אם אנחנו נאתחל את כל הקלטים שלנו מהמטריצות למספרים שהם 1 ו מינוס 1 הקוד עצמו לא יוכל לעבוד כי אנחנו מכניסים המון מספרים שלילים לחישוב ה – Sum

בפונקצית Cauclateoutput() ובעקבות זאת נקבל מספר שלילי קטן שקרוב ל-0 שגורם לה להיכשל כי היא מחזירה את אותו איבר מתהליך הלמידה ולא מצליחה ללמוד אז נצטרך לשמור על היחס של העולם שאנחנו עובדים איתם.

נפתור את בעיה זו כשנאתחל את המשקלים בין (1,-1) ואז נוכל לעבוד גם עם מספרים שלילים, כאשר נעשה פה הכפלה בין המספרים רובם יצאו חיובים ולא תמיד נקבל מספר שלילי ולכן הלמידה תהיה אפקטיבית.

**לסיכום:**

אנו עוסקים בעבודה בזיהוי קו שבור הנתונות בקבוצות הלמידה אך העבודה עוסקת בזיהוי צורות מסוימות ולכן לא נתנו לזה יחוס משמעותי בעבודה עצמה מבחינת הסברים כי צורת הקלט אינה חשובה אלה שהצורה חוזרת על עצמה במטריצות השונות כדי שיהיה תהליך למידה אפקטיבי ונכון.

עפ"י כל נתונים שאספנו ראינו שלכל רשת יש את היתרונות שלה ויש עדיפות ללמוד בשיטה רנדומלית, אך תמיד צריך לשים לב שההבדל בין רשתות עם יותר שכבות יבוא לידי ביטוי כאשר נתעסק עם קבוצות למידה יותר גדולות ונוכל לחקור בצורה טובה יותר את תהליכי הלמידה ותוצאות הטסטים.