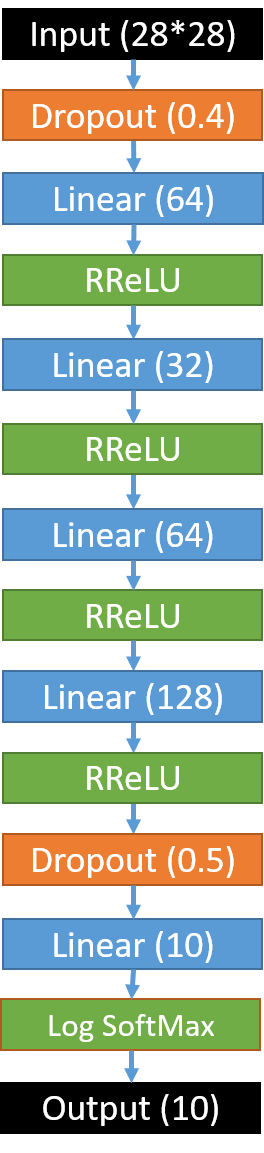
תרגיל בית 1 – נושאים מתקדמים במערכות מידע

*שמעון ארזואן 037036209*

*שגיא חכמון 365656870*

*גיא רוזין 204097984*

<https://github.com/sagi180/Deep-Learning>



**תיאור ארכיטקטורה**

המודל שבחרנו מכיל 4 שכבות, מתואר באיור.

|  |  |
| --- | --- |
| פרמטר | ערך |
| מספר Epocs | 1000 |
| גודל Batch | 256 |
| פונקציה אופטימיזציה | Adam |
| פרמטרי פונקציה האופטימיזציה | Learning Rate = 0.001  beta1 = 0.9  beta2 = 0.999  epsilon = 1e-8 |
| מספר פרמטרי המודל | 64,042 |

**שיטת האימון**

נרמלנו את סט האימון כך שהתוחלת שלו תהיה 0 והשונות 1. בנינו את הרשת לפי המתואר לעיל. לפני כל epoch ביצענו shuffle לסט.

לאחר בניית המודל, ביצענו נרמול לסט המבחן לפי השונות והתוחלת של סט האימון, והרצנו את סט זה לאחר שביטלנו במודל את שכבות ה-dropout.

בתהליך האימון תחילה בחרנו מספר גדול ממש של epochs, ובסיום הריצה הדפסנו גרף של השגיאה כתלות במספר ה- epochובדקנו באיזה epoch קיבלנו דיוק מקסימלי. כך בחרנו את מספר ה- epochsלהרצה.

בסיום קיבלנו כי ה-Accuracy שלנו הוא **98.78%**.

**סיכום הניסיונות שלנו**

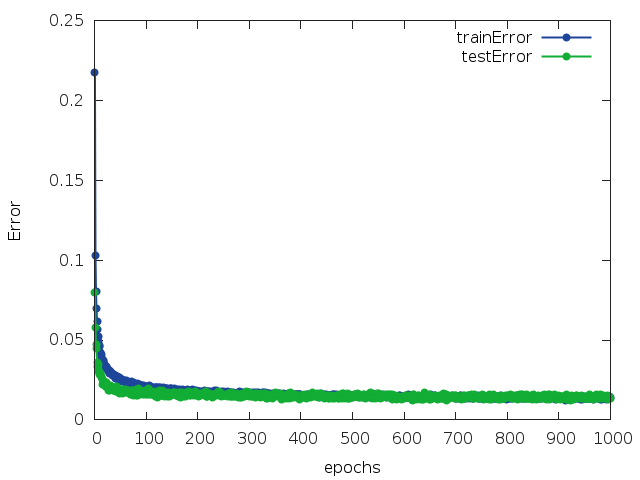
בכדי למצוא את השיטה הטובה ביותר הרצנו ניסיונות רבים על מודלים שונים עם פונקציות אופטימיזציה שונות, ע"י ביצוע grid search. ניסינו את השיטות: AdaDelta, AdaGrand, Adam, AdaMax, ASGD, SGD, ועבור כל שיטה בדקנו כמה פרמטרים אפשריים (כמו learning rate, מומנטום וכו'). ניסינו לבדוק איך משפיעים גודל ה-batch של כל איטרציה ומספר ה-epochs. בנוסף בדקנו עומקים שונים של רשתות לינאריות עם מודלים לא ליניאריים שונים (למשל ReLU, RReLU, PReLU ועוד). ניסינו גם בתחילת כל epoch לבצע shuffing, או לבצע דגימה עם חזרות.

בבדיקות שלנו ראינו כי אנחנו מקבלים Overfitting לסט האימון, לכן לאחר שהגענו לרשת בעלת ביצועים טובים, הוספנו רגולריזציה – הוספנו לרשת בהדרגה שכבות Dropout (גם כאן, בדקנו מיקומים שונים והסתברויות שונות).

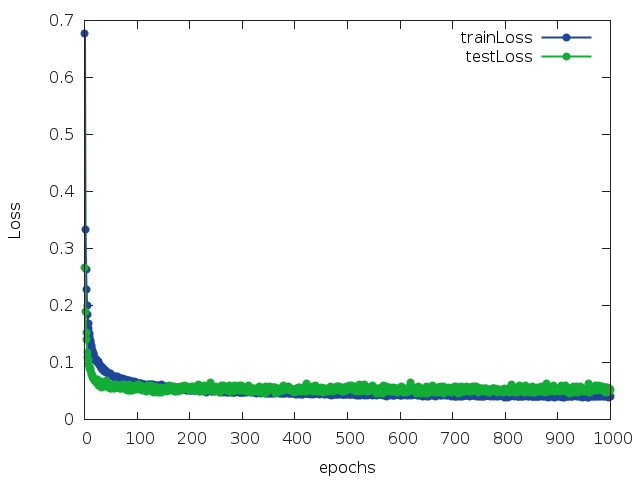
לבסוף בחרנו את השיטה שהניבה את תוצאות הטובות ביותר על סט המבחן.

**גרפים**

שגיאה כתלות ב-epoch:



Loss כתלות ב-epoch:



**מסקנות**

* פונקציות אופטימיזציה שונות (AdaDelta, AdaGrand, Adam, AdaMax, ASGD, SGD) מתכנסות באופן שונה, ושווה לנסות כמה מהן בכדי לראות מי המתאימה ביותר לבעיה.
* מספר ה-epochs משפיע על התוצאות, איטרציות נוספות במקרים מסוימים משפרות את המודל, יש לשים לב כי ישנם מקרים שבהן מספר ה- Epochs יכול להביא ל-overfitting.
* לאחר שמגיעים לרשת בעלת ביצועים מספקים, חשוב לבדוק את גרף ה-loss ולהוסיף רגולריזציה, כאשר dropout עובדת מעולה.