דו"ח לתרגיל 3 – למידת מכונה

תומר גיל  
\*\*\*\*\*\*\*\*\*  
13/05/18

# מבנה הרשת

גודל הקלט לרשת הוא כמובן 784 (28\*28) ומשם הוא עובר לשכבה הליניארית הראשונה, שמעבירה את הקלט לגודל של **200**, ולאחר מכן מופעל **סיגמואיד** על התוצאה לפני שהיא עוברת לשכבה הבאה. השכבה הבאה כמובן מעבירה את תוצאת הביניים לגודל 10 ומפעילה softmax בשביל ההסתברויות.

במהלך האימונים התנסיתי עם גדלי שכבה נסתרת שונים – 200 הוא מבין הגדולים מהם אבל גדולים יותר לא עבדו טוב וקטנים יותר אמנם הקטינו את זמן הריצה אבל הגריעו מהתוצאה (ניסיתי מיני גדלים עד לגדלים בסגנון 16 או 10, אבל גם באיזור ה-100 וה-50).  
כמו כן התנסיתי עם מספר פונקציות אקטיביציה – סיגמואיד שנבחר לבסוף, tanh, ReLU ואפילו leaky-ReLU מהתרגול האחרון. כולם נתנו תוצאות דומות אך סיגמואיד מוביל באחוז או שניים על כולם.

# תהליך האימון

תהליך האימון כלל מעבר של **30 אפוקים** על סט האימון, בו לכל דוגמא חושבו ה-loss והגרדיאנטים, ועודכנו עם כלל העדכון של SGD עבור **קצב למידה של 0.02**. לא השתמשתי בבאטצ'ים כלל.  
עבור כל אפוק חושב הדיוק וה-loss הממוצע עבור ה-validation set ונשמרו הפרמטרים שהובילו לתוצאות המקסימליות מבחינת הדיוק. לבסוף, הפרמטרים הטובים ביותר שומשו ל-prediction של ה-test set.

# תוצאות

המודל שלי מגיע עם הנתונים הנ"ל ברוב ההרצות לאיזור ה-89.5 אחוזים (±0.3 אחוזים) כאשר השיא הנוכחי עומד על 89.8%.

A screenshot of a computer

Description generated with very high confidence