

## דו"ח תרגיל 3 – למידת מכונה

### מבנה הרשת

יצרנו רשת נוירונים בעלת שכבה נסתרת אחת עם פונקציית האקטיבציה ReLU. גודל הקלט לרשת הוא 784 (כגודל מספר התכונות בקובץ "train\_x"). משם הקלט עובר לשכבה הנסתרת, שבגודל 100 (גודל ה-Hidden Layer). לאחר מכן, מופעלת פונקציית ReLU על התוצאה שמייצרת את הקלט לשכבה הבאה. השכבה הבאה מעבירה את תוצאת הביניים לגודל 10 שזה כמות הסיווגים האפשריים עבור ערך מהtrain ומהtest ומפעילה softmax לקבלת תוצאה הסתברותית לסיווג הקלט. במהלך התרגיל ניסינו לעבוד עם גדלי Hidden Layer שונים. לבסוף נבחר הגודל 100 שנתן את התוצאות הטובות ביותר.

### תהליך האימון

תהליך האימון כלל מעבר של 20 אפוקים על סט האימון, בו לכל דוגמא חושבו ה-loss והגרדיאנטים, ועודכנו המשקולות עם כלל העדכון של SGD עבור קצב למידה של 0.01.

השתמשנו בvalidation set בכדי לבדוק כמה אפוקים מובילים לתוצאה הטובה ביותר ומהו קצב הלמידה האידאלי.

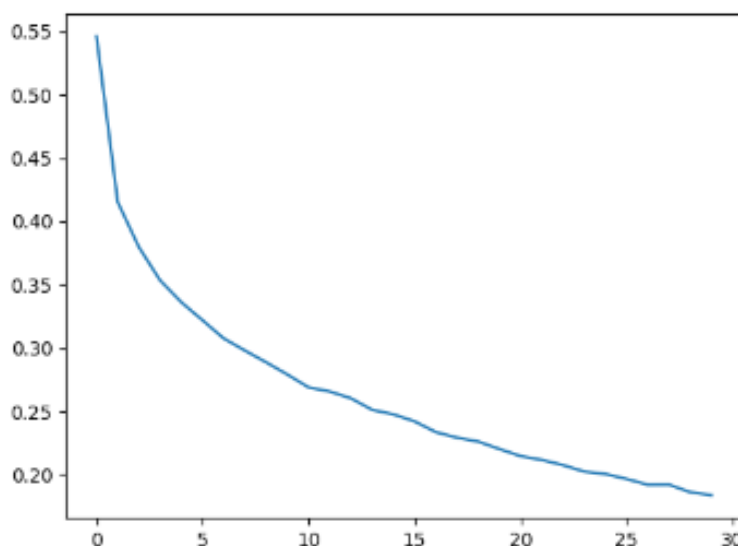
לבסוף, הפרמטרים הטובים ביותר נבחרו ואיתם השתמשנו ל-prediction של ה-test set.

### בחירת הפרמטרים

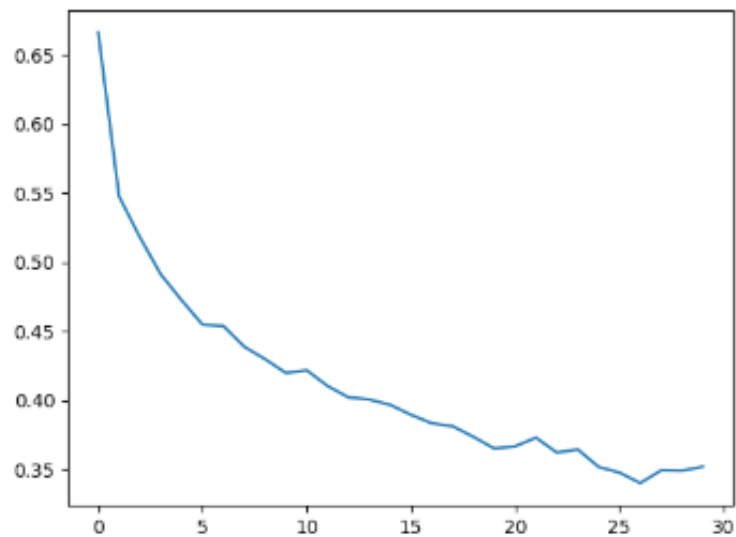
ציר הא מתאר את כמות הepochs וציר הy את אחוז ההצלחה בניבוי של 0.2 מהtrain שנגדמים אקראית ולא אומן כחלק משאר הtrain במודל.

דוגמאות לגרפים שמציגים חלק מהבדיקות שערכנו -

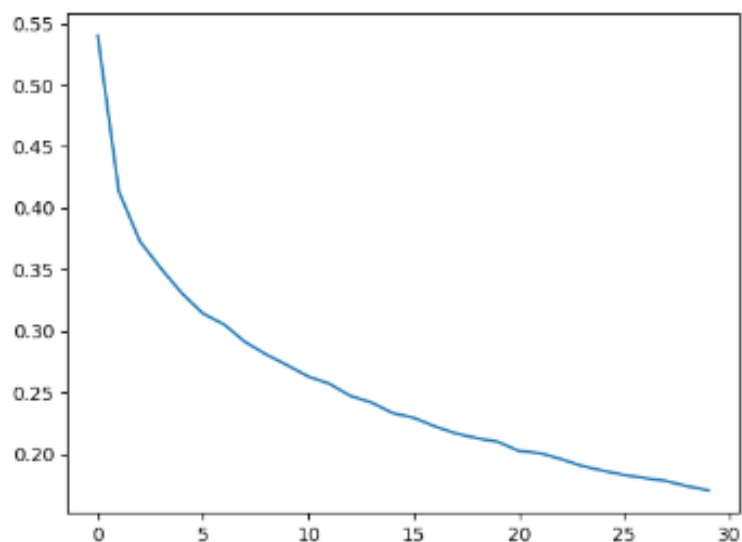
עבור קצב למידה 0.01 ושכבה נסתרת בגודל 100:



עבור קצב למידה 0.03 ושכבה נסתרת בגודל 100:



עבור קצב למידה 0.01 ושכבה נסתרת בגודל 150:



ואכן, עפ"י גרפים אלה ניתן לראות שא' בין 20 ל-30 epochs ההפרש היה יחסית קטן והירידה יחסית התמתנה, ולכן העדפנו לקחת את 20 כמס' epochs. בנוסף, התוצאות הטובות ביותר היו עבור קצב למידה 0.01 וגודל שכבה נסתרת 100.