**דו"ח תרגיל 3 – אורי דאבוש**

בתרגיל זה מימשתי רשת נוירונים. רשת הנוירונים שלי היא דינמית, היא לא קבועה אלא היא מקבלת את מספר השכבות וגודלן כפרמטרים. הרשת מקבלת כפרמטר רשימה של מימדים – גדלי כל שכבה (במקרה שלנו שכבת הקלט היא בגודל 784 ושכבת הפלט היא בגודל 10), פונקציית אקטיבציה והנגזרת שלה, ו-learning rate.

רשת הנוירונים שלי מאתחלת את ה-w ואת ה-bias במימדים המתאימים בעזרת פונקציה שמגרילה את הערכים בהתפלגות נורמלית עם ממוצע 0 וסטיית תקן 0.1.

פונקציית האימון של הרשת מקבלת פרמטרים נוספים – נתוני האימון, מספר איפוקים וגודל batch. בכל איפוק, הפונקציה מערבבת את הנתונים, מחלקת אותם ל-batch-ים בגודל שהתקבל כפרמטר, ומחשבת את הגרדיאנט לפי כל wi ו-biasi. לאחר מכן היא מחשבת את הממוצע של ערכי הגרדיאנטים ב-batch הנוכחי, ומעדכנת את ה-w-ים ואת ה-bias-ים. היא משתמשת בשתי פונקציות של back-propagation ו-forward-propagation.

הרשת שבניתי ממזערת את פונקציית ה-NLL (negative log likelihood), זה בא לידי ביטוי בגרדיאנטים (הגרדיאנטים הם של פונקציה זו).

כעת אתאר את הפרמטרים של הרשת הספציפית שהשתמשתי בה לתרגיל.

הרשת שהשתמשתי בה בנויה מ-3 שכבות, שכבה ראשונה (קלט) בגודל 784, שכבה שנייה (נסתרת) בגודל 128 ושכבה שלישית (פלט) בגודל 10. פונקציית האקטיבציה שהשתמשתי בה היא sigmoid, ובשכבה האחרונה softmax (כדי לקבל את ה-"הסתברויות" לכל מחלקה).

ה-learning rate שהשתמשתי בו הוא 0.01.

את הנתונים שקיבלתי נרמלתי על ידי חלוקה ב-255 (זו בעצם המרה של הנתונים מהטווח [0,255] ל-[0,1]), והפעלתי את פונקציית האימון עם 50 איפוקים וגודל batch של 8.