

דו"ח בניית הרשת

פרטי הרשת:

1. נרמלתי את הערכים של הפיקסלים ע"י חלוקה שלהם ב- 256.
 2. **אתחול W ו-b:** השתמשתי באיתחול של ה-W ו-b ע"י `np.random.uniform` עם $\mu = 0.2, \sigma = 0.05$. כך הערכים נמצאים בסביבת האפס שזה מביא תוצאות טובות לאחר מכן בפונקציית האקטיבציה. בנוסף הערך $\mu = 0.2$ הוא כדי לתת קצת הטיה לכיוון החלק החיובי ראיתי שזה נותן לי תוצאות יותר טובות. לאחר מכן הכפלתי את ה-W ו-b ב- `np.sqrt(1/size_previous_layer)`. זה לפי האתחול הנקרא Xavier, שהבנתי שיכול להועיל באתחול של W.
 3. **כמות השכבות:** השתמשתי בשכבה פנימית אחת, וגודלה 85. עבור שכבה גדולה יותר לא התקבלו תוצאות יותר טובות.
 4. **פונקציית האקטיבציה:** השתמשתי בפונקציית Relu (בפיאצה נכתב שניתן להשתמש בפונקציות אקטיבציה שונות מ-Sigmoid). לשם התייעוד אכתוב:
$$Relu(x) = \max(0, x)$$
והנגזרת היא:
$$\frac{dRelu}{dx} = \begin{cases} 0, & x \leq 0 \\ 1, & x > 0 \end{cases}$$
בנוסף לפני הפעולה של פונקציית האקטיבציה, נרמלתי את הערכים ע"י $score - z$, כך יוצא שהערכים יהיו סביב האפס ואז פונקציית האקטיבציה מביאה תוצאות טובות.
 5. **קצב הלמידה:** ניסיתי כל מיני ערכים עבור קצב הלמידה ושמתי לב כי עבור $\eta = 0.1$ מתקבלת התוצאה הכי טובה. (עבור ערכים קטנים או גדולים התקבלו תוצאות פחות טובות).
 6. **מספר ה-Epochs:** בדקתי כל מיני ערכים וראיתי שעבור 15 Epochs מתקבלות תוצאות ממש טובות (למעלה מ-97%). בנוסף שמתי לב שעבור 20 או 30 Epochs ההצלחה לא גדלה. (אולי בכמות Epochs ממש גדולה ההצלחה תגדל אבל זה יהיה הרבה מעבר לזמן הריצה שהוגדר לנו בתרגיל שהתוכנית יכולה לרוץ).
- עבור בדיקת ההצלחה של התוכנית, לקחתי את ה- *train set* וחילקתי אותו באופן רנדומלי ל- *train set* ו- *validation set*. לקחתי 50000 דוגמאות ל- *train* ו-5000 דוגמאות ל- *validation*.

עד כאן.