

ארכיטקטורת הרשת:

ראשית, הרשת בנויה מ-3 שכבות: שכבת הקלט, שכבה נסתרת (כנדרש בשאלה) ושכבת הפרדיקציה.

השכבה הנסתרת מכילה 100 נוירונים ושכבת הפרדיקציה מכילה 10 נוירונים (כל אחד מייצג תווית אפשרית).

המטרה היא שערך כל נוירון בשכבת הפרדיקציה יכיל את ההסתברות שהתווית הנכונה היא התווית אשר הוא מייצג. לאחר השגת מטרה זו, נבחר בתווית בעל ההסתברות הגבוהה ביותר.

עבור השכבה הנסתרת, עבור כל נוירון, נפעיל על ערך הצירוף הלינארי בין משקולותיו לבין התצפיות את פונקציית sigmoid. פונקציה זו ממפה כל מספר ממשי לקטע $[0, 1]$. ערך זה הינו ערך האקטיבציה של הנוירון (מציין את חשיבות הנוירון בהמשך החישוב לצורך קביעת הפרדיקציה).

עבור שכבת הפרדיקציה, עבור כל נוירון, לאחר חישוב ערך הצירוף הלינארי בין משקולותיו לבין ערכי האקטיבציה של הנוירונים בשכבה הנסתרת, נקבל וקטור המכיל 10 כניסות.

על מנת להשיג את מטרתנו, נפעיל על וקטור זה את פונקציית softmax. כפי שראינו בהרצאה ובתרגול, פונקציה זו ממפה וקטור של ערכים ממשים לוקטור של ערכים הסתברותיים. ע"י כך נוכל לחזות את התווית.

לצורך אימון המודל וביצוע backpropagation, נבחר בפונקציית Cross Entropy Loss (multiclass). זהו הפונקציה אותה אנו מנסים למזער בשלב אימון המודל. לצורך מזעור הפונקציה, נבצע שיטת גרדיאנט עם קצב למידה (גודל צעד) של 0.1. על מנת לבצע את השיטה נחשב את הנגזרות החלקיות של הביטויים הרלוונטיים ביחס לפונקציית loss.

מספר הסבבים שנקבע לצורך אימון המודל הינו כ-50. (כל סבב הינו מעבר על כל מדגם האימון).

בחירת hyperparameters של המודל:

בחירת קצב הלמידה: קצב הלמידה ההתחלתי שנבחר הינו 0.1. ערך זה מבטא איזון בין זמן אימון ארוך (נגרם כתוצאה מבחירת קצב למידה קטן מידי), לבין פגיעה במזעור פונקציית ה-loss (נגרם כתוצאה מבחירת קצב למידה גדול מידי). עבור ערך זה קיבלתי תוצאות מצוינות העולות על הנדרש בשאלה ולכן נבחר להשאירו.

בחירת גודל ה-batch: גודל ה-batch ההתחלתי שנבחר הינו 128. מספר זה התאים מבחינת מפרט המחשב וכן קיבלתי עבורו תוצאות מצוינות העולות על הנדרש בשאלה. לכן נבחר להשאירו.

בחירת מספר סבבי האימון: מספר סבבי האימון שנבחר הינו כ-50. גודל מדגם האימון הינו 60,000 רשומות, לכן כל סבב אימון לוקח זמן לא מבוטל. כפי שניתן לראות בגרף המצורף, מספר זה מבטא יחס טוב בין הזמן שהשקענו באימון המודל לבין דיוק המודל על מדגם המבחן.

בחירת גודל השכבה הנסתרת: ביצעתי מספר הרצות עם גדלי שכבה נסתרת שונים (50, 100, 200). עבור גודל 50, התוצאות שקיבלתי אינן היו לשביעות רצוני (למרות שעברנו את אחוז הדיוק הנדרש בשאלה), ואילו עבור גדלים 100, 200 קיבלתי תוצאות מעולות (אחוז דיוק העולה על 90%). לבסוף, בחרתי בגודל 100 על מנת לקבל מודל פשוט יותר ומהיר יותר.

לסיום, אחוז הדיוק של המודל המאומן הינו 91.7%, אחוז הדיוק הינו מעל ל-75% ולכן עומד בתנאי המשימה.

The accuracy of the trained model on the test dataset: 0.917

בנוסף מצורף גרף של דיוק המודל על סט האימון ועל סט המבחן כתלות במספר סבבי האימון, כנדרש בשאלה.

