

ארכיטקטורת המודל:

הרשת שבניתי מורכבת מ-3 שכבות קונבולוציה ומשכבה אחת של fully connected. בנוסף היא מבצעת עיבוד מקדים על המידע.

עיבוד מקדים

ראשית, נבצע עיבוד מקדים של המידע:

1. נשתמש בפונקציה RandomHorizontalFlip עם פרמטר $p=0.5$ אשר הופכת את התמונה (ביחס לציר האופקי) בהסתברות p . פעולה זו מבטאת רגולריזציה ומעודדת הכללה. (שלב זה מבוצע רק על סט האימון לצורך עקביות המודל המאומן על סט המבחן).
2. נמיר את התמונה לטיפוס Tensor על מנת שנוכל להשתמש בספריית torch.
3. ננרמל את התמונות ע"י שימוש בפונקציה Normalize.

שכבות הקונבולוציה

עבור כל אחת מן השכבות, בעת הפעלת כל אחד מן הפילטרים, נבצע את הפעולות הבאות:

1. חישוב פיקסלי התוצאות ע"י ביצוע קונבולוציה בין התמונה לבין הפילטר.
2. נרמול פיקסלי התוצאות ע"י שימוש בפונקציה batchNorm2d.
3. הפעלת פונקציית אקטיבציה ReLu על פיקסלי התוצאות.
4. ביצוע maxpooling עם פרמטר 2.

שכבת הקונבולוציה הראשונה

הפרמטרים שנבחרו עבור שכבה זו הינם $\text{filters}=16, \text{kernel}=5, \text{padding}=2, \text{stride}=1$.

שכבה זו מקבלת כקלט תמונה בגודל $32 \times 32 \times 3$ (התמונה המקורית). עבור כל פילטר, ע"פ הנוסחה שראינו בתרגול, פלט ביצוע הקונבולוציה הינו תמונה בגודל 32×32 . לאחר מכן מתבצע maxpooling (עם פרמטר 2) ולכן הפלט הסופי הינו תמונה בגודל 16×16 . ישנם 16 פילטרים, ולכן הפלט הסופי של שכבה זו הינו תמונה בגודל $16 \times 16 \times 16$.

מספר פרמטרי הפילטרים הנלמדים הינו $1216 = 16 * (5 * 5 * 3 + 1)$. כמו כן, אנו משתמשים בפונקציה batchNorm2d בעלת $16 * 2 = 32$ פרמטרים נלמדים. לכן מספר הפרמטרים הנלמדים הכולל של שכבה זו הינו **1,248**.

שכבת הקונבולוציה השנייה

הפרמטרים שנבחרו עבור שכבה זו הינם $\text{filters}=32, \text{kernel}=5, \text{padding}=2, \text{stride}=1$.

שכבה זו מקבלת כקלט תמונה בגודל $16 \times 16 \times 16$ (פלט השכבה הקודמת). עבור כל פילטר, ע"פ הנוסחה שראינו בתרגול, פלט ביצוע הקונבולוציה הינו תמונה בגודל 16×16 . לאחר מכן מתבצע maxpooling (עם פרמטר 2) ולכן הפלט הסופי הינו תמונה בגודל 8×8 . ישנם 32 פילטרים, ולכן הפלט הסופי של שכבה זו הינו תמונה בגודל $8 \times 8 \times 32$.

מספר פרמטרי הפילטרים הנלמדים הינו $12832 = 32 * (5 * 5 * 16 + 1)$. כמו כן, בפונקציה batchNorm2d בעלת $32 * 2 = 64$ פרמטרים נלמדים. לכן מספר הפרמטרים הנלמדים הכולל של שכבה זו הינו **12,896**.

שכבת הקונבולוציה השלישית

הפרמטרים שנבחרו עבור שכבה זו הינם **filters=64, kernel=3, padding=1, stride=1**.

שכבה זו מקבלת כקלט תמונה בגודל $8 \times 8 \times 32$ (הפלט מהשכבה הקודמת). עבור כל פילטר, ע"פ הנוסחה שראינו בתרגול, פלט ביצוע הקונבולוציה הינו תמונה בגודל 8×8 . לאחר מכן מתבצע maxpooling (עם פרמטר 2) ולכן הפלט הסופי הינו תמונה בגודל 4×4 . ישנם 64 פילטרים, ולכן הפלט הסופי של שכבה זו הינו תמונה בגודל $4 \times 4 \times 64$.

מספר הפרמטרי הפילטרים הנלמדים הינו $64 * (3 * 3 * 32 + 1) = 18496$. כמו כן, אנו משתמשים בשכבה זו בפונקציה batchNorm2d בעלת $64 * 2 = 128$ פרמטרים נלמדים. לכן מספר הפרמטרים הנלמדים הכולל של שכבה זו הינו **18,624**.

שכבת fully connected

לצורך ביצוע שכבה זו, נשטח את התמונה שקיבלנו משכבות הקונבולוציה לוקטור באורך $4 \times 4 \times 64$ (פלט שכבת הקונבולוציה האחרונה). בנוסף, על מנת לבצע רגולריזציה, נשתמש בפונקציה Dropout עם פרמטר $p=0.5$. פונקציה זו מאפסת את ערך האקטיבציה של כל נירון בהסתברות p . ע"י כך, נמזער over-fitting ונאפשר הכללה טובה יותר.

שכבה זו מכילה 10 נירונים (כמספר התוויות). לכן לאחר חישוב ערך הצירוף הלינארי בין משקולות כל נירון בשכבה זו לבין ערכי האקטיבציה של הנירונים בשכבה הקודמת (התמונה המשוטחת), נקבל וקטור המכיל 10 כניסות.

מספר הפרמטרים הנלמדים הינו $10 * (4 * 4 * 64 + 1) = 10,250$.

לבסוף, על מנת לספק פרדיקציה של תווית התמונה נפעיל את פונקציית logsoftmax ונבחר באינדקס בעל הערך המקסימלי.

מספר הפרמטרים הנלמדים הכולל של המודל הינו **43,018** (קטן מ-50,000 כנדרש בשאלה).

```
total number of model parameters: 43018
```

אימון המודל

לצורך אימון המודל וביצוע backpropagation, נבחר בפונקציה Cross Entropy Loss (multiclass), כלומר זוהי הפונקציה אותה אנו מנסים למזער בשלב האימון. לצורך ביצוע מטרה זו, נבצע שיטת גרדיאנט עם קצב למידה של 0.001.

מספר סבבי האימון שנקבע הינו לכל היותר 60, כאשר נעצור בסבב בו המודל הצליח להשיג אחוז שגיאה על סט המבחן הקטן מ-20% (אחוז השגיאה הנדרש בשאלה). בכל סבב אימון נעבור על כל סט האימון כאשר גודל ה-batches עליהם נעבור הינו 128.

בחירת hyperparameters של המודל:

בחירת קצב הלמידה: קצבי הלמידה שנבדקו הינם 0.1, 0.01, 0.001. עבור הערכים 0.1, 0.01, לאחר מספר סיבובים אחדים, ערך פונקציית ה-loss על מדגם האימון הפסיק לקטון. בעקבות כך הסקתי כי קצבי למידה אלו גדולים מידי ואינם מאפשרים למזער את פונקציית ה-loss בצורה אפקטיבית. לכן הערך הבא שנבדק הינו 0.001. ערך זה, אומנם בקצב איטי, אכן הצליח למזער את פונקציית ה-loss ככל שמספר סבבי האימון גדל. לכן זהו הערך שנבחר.

בחירת גודל ה-batch: גודל ה-batch ההתחלתי שנבחר הינו 128. מספר זה התאים מבחינת מפרט המחשב וכן קיבלתי עבורו תוצאות המקיימות את הנדרש בשאלה. לכן זהו הערך שנבחר.

בחירת מספר סבבי האימון: לאחר קביעת פרמטר קצב הלמידה, הבנתי כי הירידה בערך פונקציית ה-loss תהיה קטנה בין סיבוב לסיבוב (כפי שצוין) ולכן הסקתי כי יש צורך במספר סבבים יחסית גדול. אם זאת, כל סבב אימון דורש זמן לא מבוטל ולכן לא נוכל לבחור מספר סבבים גדול מידי. לבסוף הערך שנקבע הינו לכל היותר 60. ערך זה מבטא איזון בין השגת אחוז השגיאה הנדרש בשאלה לבין זמן אימון לא ארוך מידי של המודל.

בחירת מספר שכבות הקונבולוציה: בשכבות הקונבולוציה יש מספר רב של פרמטרים (מספר הפילטרים בכל שכבה, גדלי kernel, padding, stride, פרמטר ה-maxpooling ועוד). לכן עבדתי לפי הכללי האצבע הבאים:

- בכל שכבת קונבולוציה לאחר מעבר על תמונת הקלט עם פילטר, גודל תמונת הפלט יהיה שווה לגודל תמונת הקלט (מבחינת מימדי אורך ורוחב ולפני ביצוע maxpooling).
- בכל שכבת קונבולוציה נבצע maxpooling עם פרמטר 2 על מנת להקטין את אורך ורוחב תמונת הפלט פי 2.
- מספר הפילטרים בין שכבה לשכבה גדל פי 2.

מספר השכבות הראשוני שניסיתי לפי כללים אלו הינו 3 אשר הצליח להשיג את אחוז השגיאה הנדרש של המודל על סט המבחן. על מנת לנסות לשפרו הוספתי שכבת נוספת (סה"כ 4 שכבות). הוספה זו אינה הובילה לירידת אחוז השגיאה ואף גרמה להגדלתו. לכן מספר שכבות הקונבולוציה שנבחר הינו 3.

בחירת מספר שכבות fully connected: מספר השכבות ההתחלתי שנבחר הינו 1 מכיוון שרציתי לתת דגש על שכבות הקונבולוציה. לאחר קביעת מספר שכבות זה, המודל השיג את אחוז השגיאה הנדרש, אך ניסיתי לשפרו. לצורך כך בדקתי את טיב המודל עבור 2 ו-3 שכבות fully connected עם מגוון רחב של גדלי שכבות. הגדלת מספר השכבות לא הוביל לירידת אחוז השגיאה על סט המבחן ואף הגדיל אותו. לכן בחרתי להישאר עם שכבת fully connected יחידה.

תוצאות המודל המאומן

לבסוף, לאחר 52 סבבי אימון המודל הצליח להשיג אחוז שגיאה העונה על הנדרש. אחוז השגיאה של המודל המאומן על סט המבחן הינו **19.6%** (אחוז דיוק של **80.04%** העולה על אחוז הדיוק הנדרש בשאלה).

```
trained model test error: 0.196
```

כמו כן, מצורפים הגרפים הבאים כנדרש בשאלה:

- אחוז השגיאה של המודל על סט האימון ועל סט המבחן כפונקציה של מספר סבבי האימון.



- ערך פונקציית Cross Entropy Loss (multiclass) של המודל על סט האימון ועל סט המבחן כפונקציה של מספר סבבי האימון.

