DEEP LEARNING – HW2

Submitted: ilan frank (043493386), Itay Berkovich (039632732)

Model architecture:

**Layer structure:**

* Sequential layer 1:

1. Convolution: kernel = 3 , padding = 1, stride = 1, channels = 16
2. Activation relu
3. Convolution: kernel = 3 , padding = 1, stride = 1, channels = 16
4. Batch normalization
5. Activation relu
6. Max pooling : kernel = 2, stride =2

* Sequential layer 2:

1. Convolution: kernel = 3 , padding = 1, stride = 1, channels = 24
2. Activation: relu
3. Convolution: kernel = 3 , padding = 1, stride = 1, channels = 24
4. Batch normalization
5. Activation: relu
6. Max pooling : kernel = 2, stride =2

* Sequential layer 3:

1. Convolution: kernel = 3 , padding = 1, stride = 1, channels = 32
2. Activation relu
3. Convolution: kernel = 3 , padding = 1, stride = 1, channels = 32
4. Batch normalization
5. Activation: relu
6. Max pooling : kernel = 2, stride =2

* Drop out: p=0.3
* Fully connected (in: 4\*4\*32, out:40)
* Batch normalization
* Activation: relu
* Fully connected (in:40, out:10)
* Activation: log softmax

**Data augmentation:**

* Random crop: padding = 4
* Random horizontal flip

**optimizer:** Adam

**loss function**: NLLLos

**hyper parameters:**

* Bath size = 64
* Learning rate = 0.001
* Dropout rate = 0.1
* Epochs: 200

**Training progress:**

**%**

**Train accuracy:** 88%

**Test accuracy: 85%**

**Train loss**: 0.32

**Test loss:** 0.43

**סיכום:**

מבנה הרשת מנסה לתפוס בצורה המיטבית את אי הלינאריות של הנתונים בהתאם למספר הפרמטרים המקסימלי שהוגדר. ע"מ להשיג מטרה זאת, הוגדרו 3 צמדים של 2 שכבות קונבולוציה ו-2 שכבות FC בסוף הרשת.

ע"מ לשמור על גודל מימדים סביר בקצה שכבות הקונבולציה, כל שכבה הוגדרה כך שלא תוריד את ממדים התמונה (kernel = 3, stride = 1, padding = 1) ורק לאחר כל צמד שכבות הוגדרה שכבת max pooling (kernel = 1, stride = 2) אשר מורידה את ממדי התמונה בחצי. הרצף הנ"ל חזר על עצמו 3 פעמים ( מספר ערוצים שונה) במטרה להגיע בקצה השכבות לממד של 4X4.

מספר הערוצים בכל צמד שכבות קובנובלציה הוגדר במטרה לעלות בצורה הדרגתית את כמות הערוצים אבל לעמוד בתנאי לכמות הפרמטרים המותרת.

ע"מ לעלות את אחוז הדיוק של המבחן, הוספנו אוגמנטציה לנתונים באמצעות random crop ו-random horizontal flip.

ע"מ להשיג לאזן את bias/variance של המודל, בחנו הוספה\הורדה של שכבות נרמול וdrop out. מבנה הרשת שנבחר הינו המבנה שבו קיבלנו את אחוז הדיוק בטסט הגבוה ביותר.