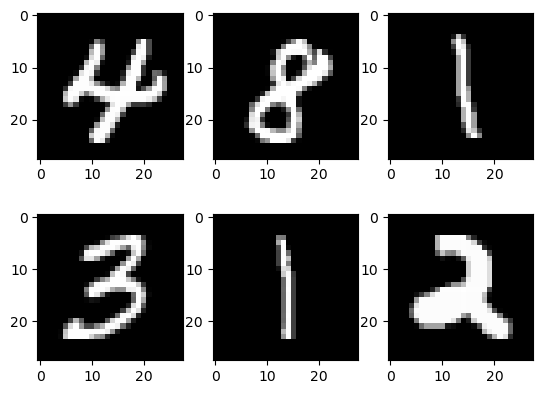
**מסמך מלווה ללמידת מכונה על MNIST:**

מגיש - איתי גלילי

**הנחות היסוד שקבעתי לתהליך בניית המודל:**

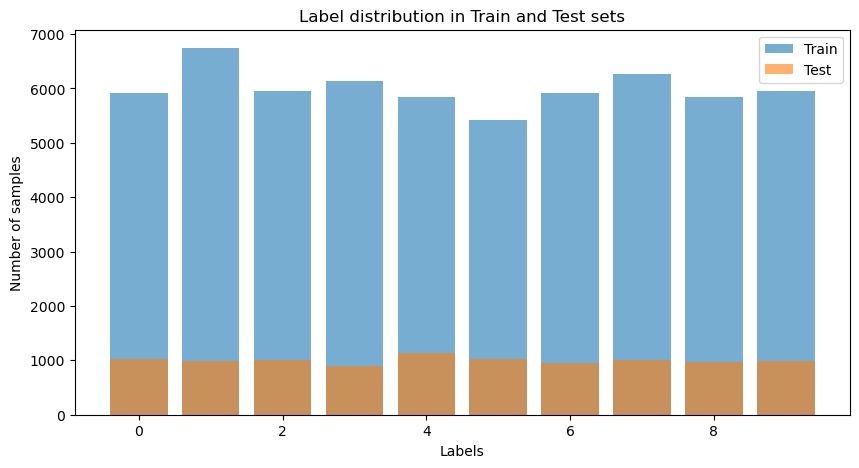
1. מודל ANN, כמתבקש בדרישות הפרוייטק.
2. שכבת קלט – 28\*28 נוירונים



1. שכבת פלט – 10 נוירונים
2. על כן, תהליך המעבר בין השכבות צריך להיות במגמת צמצום נירונים בין שכבה לשכבה.
3. **קריטריון להפסד (Loss Function)**: הקוד עושה שימוש ב- CrossEntropyLoss שהוא קריטריון נפוץ עבור בעיות סיווג.
4. Hyperparameters:
   1. Learning rate נמדדו האופציות 0.01 או 0.001.
   2. Batch size
   3. Num of epochs
   4. פונקציית אקטיבציה – relu

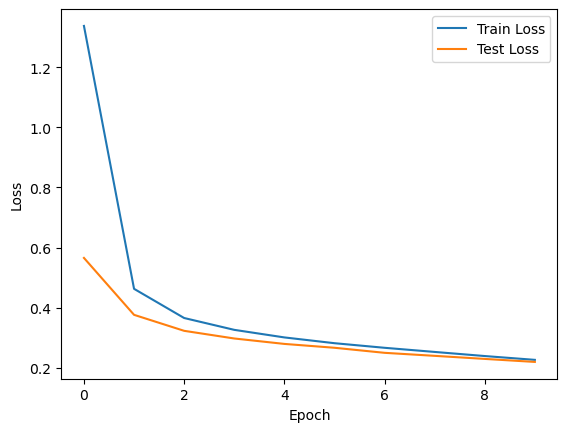
**תהליך הלמידה:**

1. בדיקת תקינות MNIST, טרם הרצת המודלים, נבדוק שיש לנו מספיק דוגמאות לאמן את המודל, נבדוק גם את החלוקה אין ה train ל test ובין ה labels הגיונית:



1. רשת מס' 1:

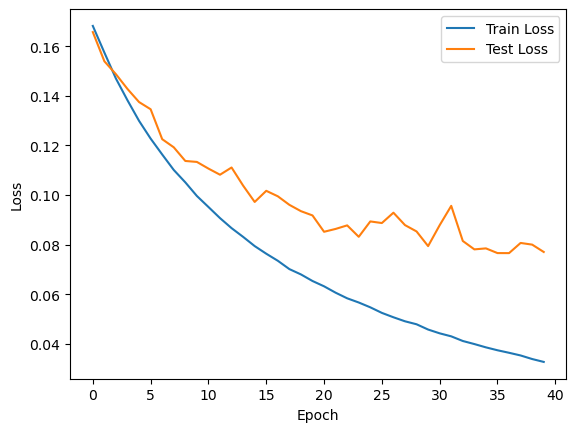
|  |  |  |
| --- | --- | --- |
| השכבות: | קצב למידה: | number of epochs |
| 784 | 0.01 | 10 |
| 128 | dropout |
| 64 | NO |
| 10 |



לקחים –

1. להגדיל את כמות המחזורים ליצירת למידה איכותית יותר.
2. רשת מס' 2:

|  |  |  |
| --- | --- | --- |
| השכבות: | קצב למידה: | number of epochs |
| 784 | 0.01 | 40 |
| 128 | dropout |
| 64 | NO |
| 10 |



לקחים:

1. אפשר לראות שמאוד מהר הtest מתקשה לרדת מ0.08.

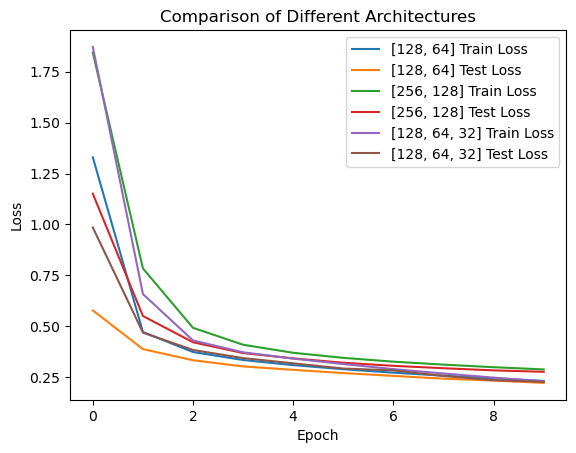
לאחר מכאן, ניסיתי לבצע מספר וריאציות של שינויים בספר וגודל השכבות ושינוי בגודל קצב הלמידה:

1. רשת מס' 3

|  |  |  |
| --- | --- | --- |
| השכבות: | קצב למידה: | number of epochs |
| 784 | 0.005 | 10 |
| 256 |  |  |
| 128 |  |  |
| 10 |  |  |

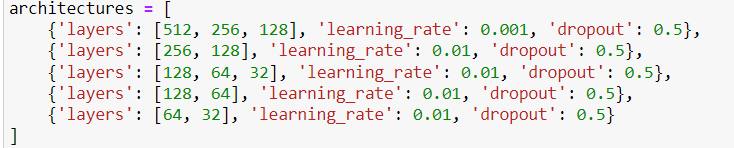
1. רשת מס' 4

|  |  |  |
| --- | --- | --- |
| השכבות: | קצב למידה: | number of epochs |
| 784 | 0.01 | 40 |
| 128 |  |  |
| 64 |  |  |
| 32 |  |  |
| 10 |  |  |



לקחים:

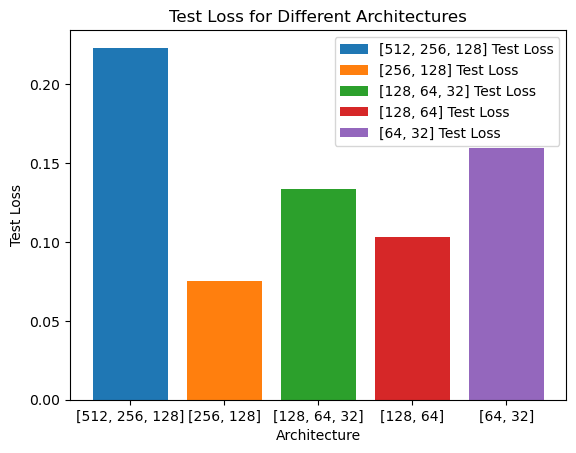
1. ניתן לראות שהבדל בין סוגי השכבות נותן דיוקים שונים (למידה באיכות שונה).
2. הרשתות הבאות (שמהוות את ההרצה הרצינית הראשונה למודל):



עם כמות epochs – 100 לכל רשת, כמו כן, חילקתי את הדאטה עם ולידאציה ובחנתי לאחר מכאן את הtest:

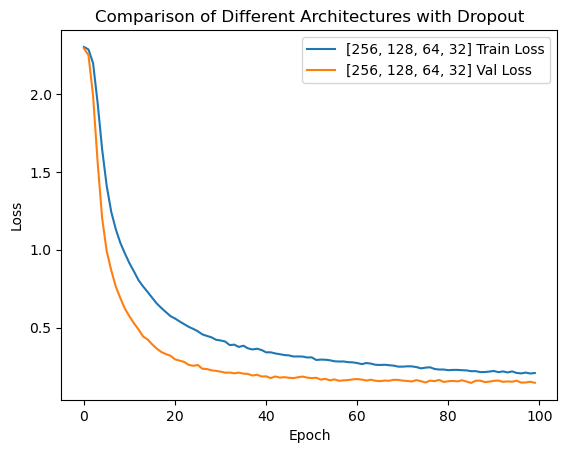
תמונה שמכילה טקסט, צילום מסך, קו, תרשים

התיאור נוצר באופן אוטומטי



הרשת בעלת השכבות 784,256,128,10 עם dropout וקצב למידה 0.01 עם תוצאות של 0.075!

1. רשת עם יותר שכבות (784,256,128,64,32,10):



וloss גבוה של 0.158 (פחות טוב מהמודלים הקודמים).

**מסקנות ולמידה מהתהליך:**

1. לחלק את הtrain עם ולידאציה. לא היה נכון להשתמש בtest בתוך התהליך.
2. הגדרת early stopping מאפשר להימנע מבזבוז זמן של ריצה שמתחילה להיכנס ל overfitting.
3. להבא, לנסות פונקציות אקטיבציה נוספות.
4. להגדיר את פונקציות הלמידה עם פרמטרים משתנים, ולהימנע מכתיבת קוד מיותרת שיוצרת סתם בלאגן (פעם אחת לאתחל אבסטרקטית את המודל, ואז לרוץ עם הגדרות ה hyper parametersבלבד).
5. להימנע מלהריץ מספר מודלים יחד, זה מעמיס על היכולת לנתח בגרפים.