

רשת הנורונים שלנו בנויה מ-3 שכבות.

בשכבה הראשונה (הקלט) אנחנו מקבלים ערכי פיקסלים בעבור תמונה בגודל  $28 \times 28$ , כלומר מערך בגודל 784, כך שכל ערך של פיקסל מהווה נירון ברשת.

השכבה השנייה היא שכבה חבויה שמכילה כ- 24 נירונים והשכבה השלישית (שכבת הפלט) מכילה 10 נירונים כך שכל נירון מייצג תווית (פרדיקציה) אחרת.

בתהליך בחירת ההיפר-פרמטרים השתמשנו בשיטת ה- Cross Validation (עם  $k=5$ ) על ה data set שקיבלנו (שמכיל בסה"כ 55,000 דוגמאות).

כלומר, חילקנו את ה data set שקיבלנו ל-5 קבוצות בגודל זהה. ביצענו 5 איטרציות, כך שבכל איטרציה קבוצה אחרת שימשה כ- validation set ושאר הקבוצות היוו training set, לבסוף חישבנו ממוצע של התוצאות השונות שהתקבלו (עבור כל validation set).

התהליך הזה בוצע מספר פעמים ובכל פעם שינינו את במעט את ההיפר-פרמטרים שלנו וכך בעצם הם נקבעו.

להלן טבלת מעקב, המציגה את הממוצע המתקבל מהרצת cross validation כתלות בערכי היפר-פרמטרים שונים:

epochs	learning rate	Hidden Layers	size	5-k-Cross Validation result (average)
1	0.1	1	24	10.02727273
2	0.1	1	24	83.69818182
3	0.1	1	24	84.64909091
4	0.1	1	24	85.32545455
5	0.1	1	24	85.41272727
6	0.1	1	24	85.41454545
7	0.1	1	24	85.83090909
8	0.1	1	24	85.66545455
1	0.2	1	24	10.02727273
2	0.2	1	24	83.27454545
3	0.2	1	24	83.86727273
4	0.2	1	24	84.36909091
5	0.2	1	24	84.23636364
6	0.2	1	24	84.92
7	0.2	1	24	84.83818182
8	0.2	1	24	84.67272727
1	0.1	1	16	10.04727273
2	0.1	1	16	83.35454545
3	0.1	1	16	84.32727273
4	0.1	1	16	84.79636364
5	0.1	1	16	84.99818182
6	0.1	1	16	84.79090909
7	0.1	1	16	85.46727273
8	0.1	1	16	85.21454545
1	0.111	1	100	9.936363636
2	0.111	1	100	84.37636364
3	0.111	1	100	85.31818182
4	0.111	1	100	86.24181818
5	0.111	1	100	86.36545455
6	0.111	1	100	86.74727273
7	0.111	1	100	86.71636364
8	0.111	1	100	86.82181818