

מבוא ללמידה עמוקה – דוח מסכם חלק א'

פרטי המגישים: אלה בראל 204880421

אור גלזמן 302314174

חלק א' – רגרסיה לוגיסטית

שאלה 1

וקטור הכניסה שלנו הוא בגודל d , ווקטור היציאה בגודל 1. לכן, יהיו d משקלים המחברים בין x ל- γ בתוספת משקל $bias$. סה"כ $d+1$ משקלים.

שאלה 2

עבור תהליך ה- $forward$, הסדר הנכון הוא:

1. פונקציה אפינית
2. סיגמואיד
3. NLL

עבור תהליך ה- $backward$ הסדר הנכון הוא

1. NLL
2. סיגמואיד
3. פונקציה אפינית

סעיף a

בתהליך ה- $forward$ נקבל וקטור אפיני ממימד $d+1$ וקטור סיגמואיד ממימד 1 ווקטור NLL ממימד 1.

בתהליך ה- $backward$ נקבל וקטור NLL ממימד 1, וקטור סיגמואיד ממימד 1. כאשר נעשה גרדיאנט לוקטור האפיני נקבל d נגזרות חלקיות לפי הכניסה ו- $d+1$ נגזרות חלקיות לפי הפרמטרים, לכן הוקטור יהיה ממימד

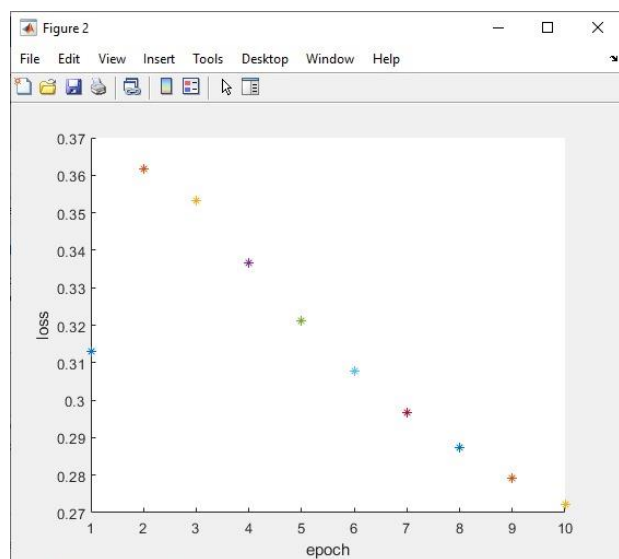
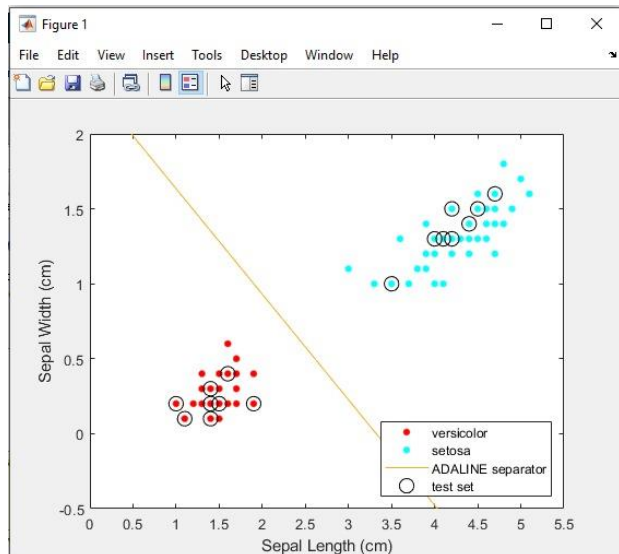
$$2d + 1$$

שאלה 3

נצפה שצורת ההפרדה תהיה קו ישר.

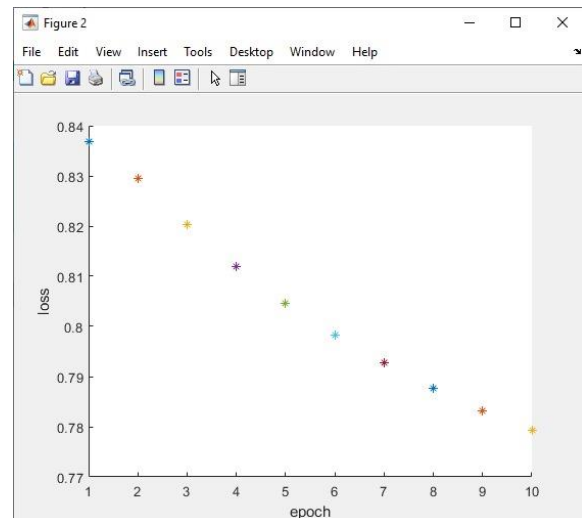
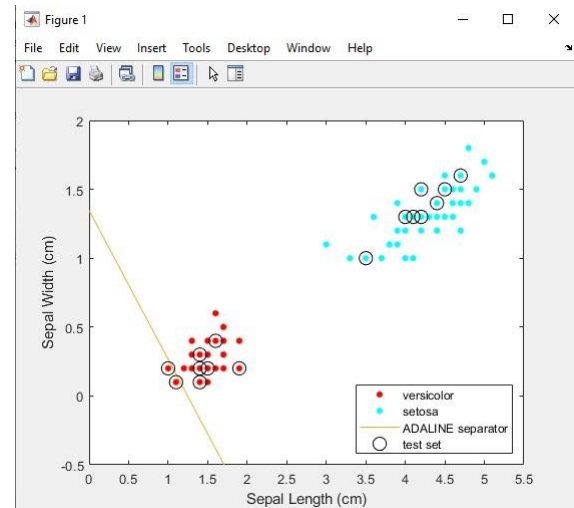
פלט ריצה: iris_logistic_regression.m

Learning Rate = 0.1



קבלנו Train Error Rate & Test Error Rate = 0%.

Learning Rate = 0.01



קבלנו Train Error Rate = 51.25%

Test error rate = 35%

שאלה 4

השוני בין 2 ההרצות נובע מקצב הלמידה. בהרצה הראשונה קצב הלמידה היה גבוה מספיק עבור סט האימון, לכן קיבלנו הפרדה טובה. בהרצה השנייה, קצב הלמידה היה נמוך מדי, לפיכך ההפרדה שנתקבלה לא הייתה איכותית.

חלק ב' – זיהוי ספרות בתמונות (סיווג באמצעות רשת נוירונים)

שאלה 1

אחרי שכבה אפינית ראשונה:

$$z_1 = w_1^T x + b_1$$

לאחר אקטיבציה ראשונה:

$$\sigma_1 = \frac{1}{1 + e^{-(z_1)}}$$

לאחר שכבה אפינית שנייה:

$$z_2 = w_2^T \sigma_1 + b_2$$

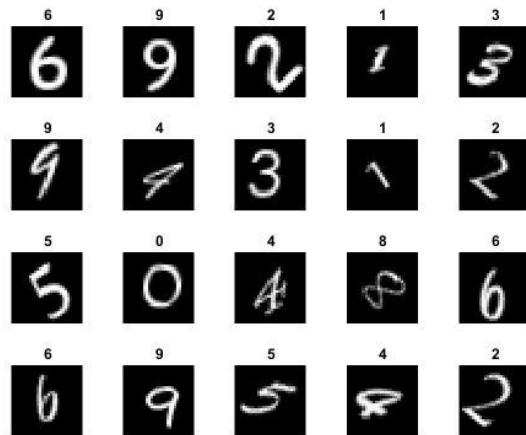
לאחר אקטיבציה שנייה:

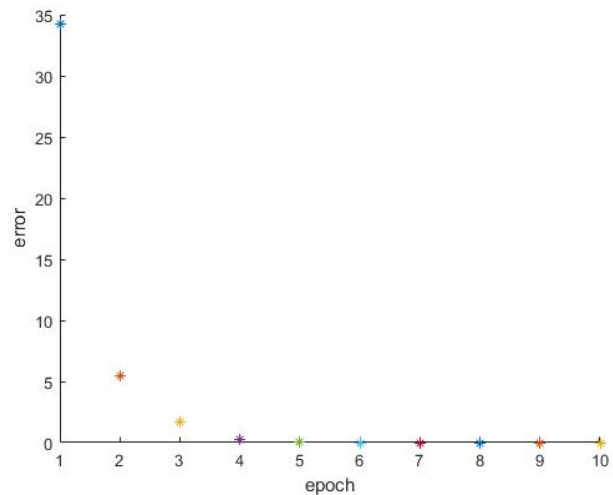
$$y = \text{softmax}(z_2)$$

לאחר NLL:

$$l(\hat{y}, y) = NLL = - \sum_{i=1}^n \hat{y}_i \log(y_i)$$

פלט ריצה: MNIST – פרסטרון דו שכבתי





שאלה 2

גודל השכבה הנסתרת בקוד הוא 700. ככל שגודל השכבה הנסתרת יותר גדול הביצועים יותר טובים והשגיאה יורדת יותר מהר. מצד שני, שכבה גדולה צורכת מקום רב בזיכרון וזמן חישוב גדול יותר.

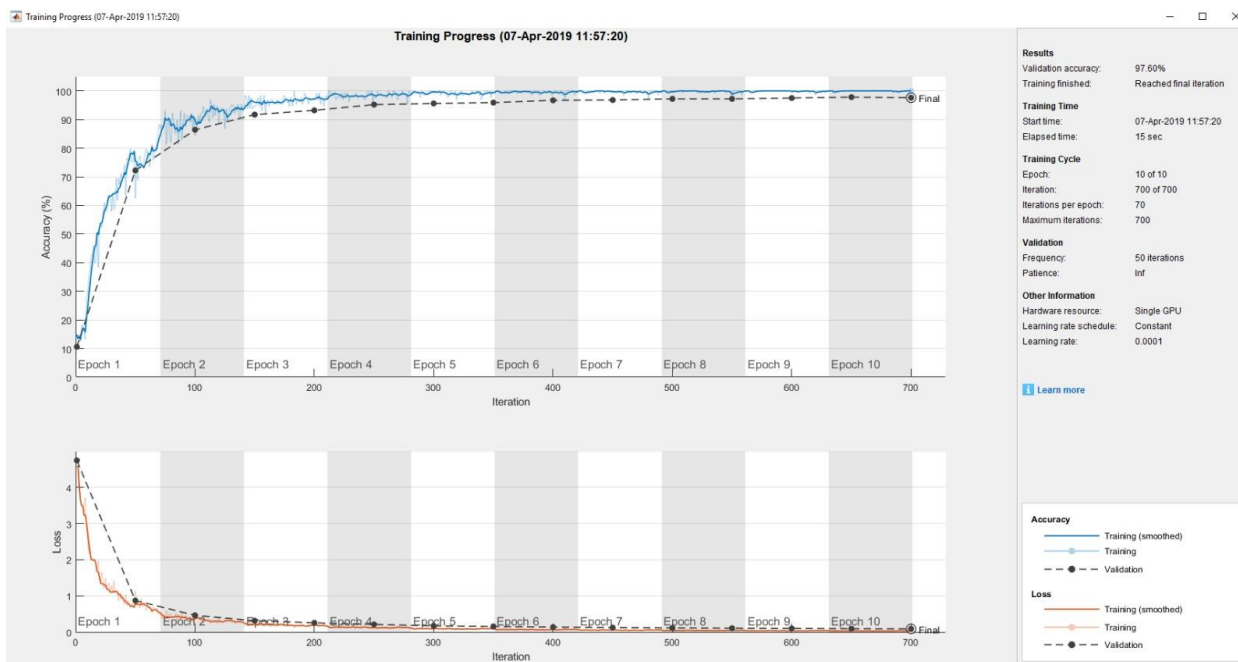
חלק ג' – סיווג ספרות בעזרת Matlab Neural Network Toolbox

שאלה 1

- Sgdm: stochastic gradient descent with momentum: שיטת האימון של הרשת.
- קצב אימון התחלתי
- מספר מקסימלי של epochs
- קבוצת הבקרה
- Validation frequency: כל כמה זמן להשתמש בתמונה מסט הולידציה.
- Verbose: מאחסן פרטים על מהלך האימון
- Verbose frequency: כל כמה זמן נקבל מידע לגבי האימון
- Plots: איזה גרפים להציג (גרף של training progress)

פלט ריצה: mnistClassificationDL.m

```
%% Set model options
options = trainingOptions('sgdm', ...
    'InitialLearnRate', 0.0001, ...
    'MaxEpochs', 4, ...
    'ValidationData', imdsValidation, ...
    'ValidationFrequency', 50, ...
    'Verbose', true, ...
    'VerboseFrequency', 25, ...
    'Plots', 'training-progress');
```



Training on single GPU.
Initializing image normalization.

Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Validation Accuracy	Mini-batch Loss	Validation Loss	Base Learning Rate
1	1	00:00:00	14.84%	10.60%	4.5643	4.7407	1.0000e-04
1	25	00:00:01	57.81%		1.4898		1.0000e-04
1	50	00:00:01	62.50%	72.20%	1.1310	0.8706	1.0000e-04
2	75	00:00:02	92.97%		0.3139		1.0000e-04
2	100	00:00:02	84.38%	86.40%	0.4525	0.4629	1.0000e-04
2	125	00:00:03	86.72%		0.3846		1.0000e-04
3	150	00:00:03	94.53%	91.70%	0.2336	0.3145	1.0000e-04
3	175	00:00:04	98.44%		0.1629		1.0000e-04
3	200	00:00:04	96.88%	93.20%	0.2395	0.2561	1.0000e-04
4	225	00:00:05	99.22%		0.1000		1.0000e-04
4	250	00:00:05	98.44%	95.20%	0.1345	0.2128	1.0000e-04
4	275	00:00:06	98.44%		0.1227		1.0000e-04
5	300	00:00:06	99.22%	95.60%	0.0978	0.1717	1.0000e-04
5	325	00:00:07	98.44%		0.0964		1.0000e-04
5	350	00:00:07	99.22%	95.90%	0.0807	0.1549	1.0000e-04
6	375	00:00:08	100.00%		0.0669		1.0000e-04
6	400	00:00:08	99.22%	96.70%	0.0824	0.1390	1.0000e-04
7	425	00:00:09	99.22%		0.0484		1.0000e-04
7	450	00:00:09	100.00%	96.80%	0.0639	0.1250	1.0000e-04
7	475	00:00:10	99.22%		0.0730		1.0000e-04
8	500	00:00:10	100.00%	97.20%	0.0441	0.1146	1.0000e-04
8	525	00:00:11	100.00%		0.0362		1.0000e-04
8	550	00:00:12	96.88%	97.20%	0.0712	0.1078	1.0000e-04
9	575	00:00:12	100.00%		0.0335		1.0000e-04
9	600	00:00:13	99.22%	97.50%	0.0444	0.1026	1.0000e-04
9	625	00:00:13	100.00%		0.0389		1.0000e-04
10	650	00:00:14	100.00%	97.80%	0.0369	0.0938	1.0000e-04
10	675	00:00:14	100.00%		0.0315		1.0000e-04
10	700	00:00:15	100.00%	97.60%	0.0330	0.0890	1.0000e-04

ה-loss הסופי של האימון: 0.03

ה-loss הסופי של ה-validation: 0.089

ה-accuracy הסופי של האימון: 100%

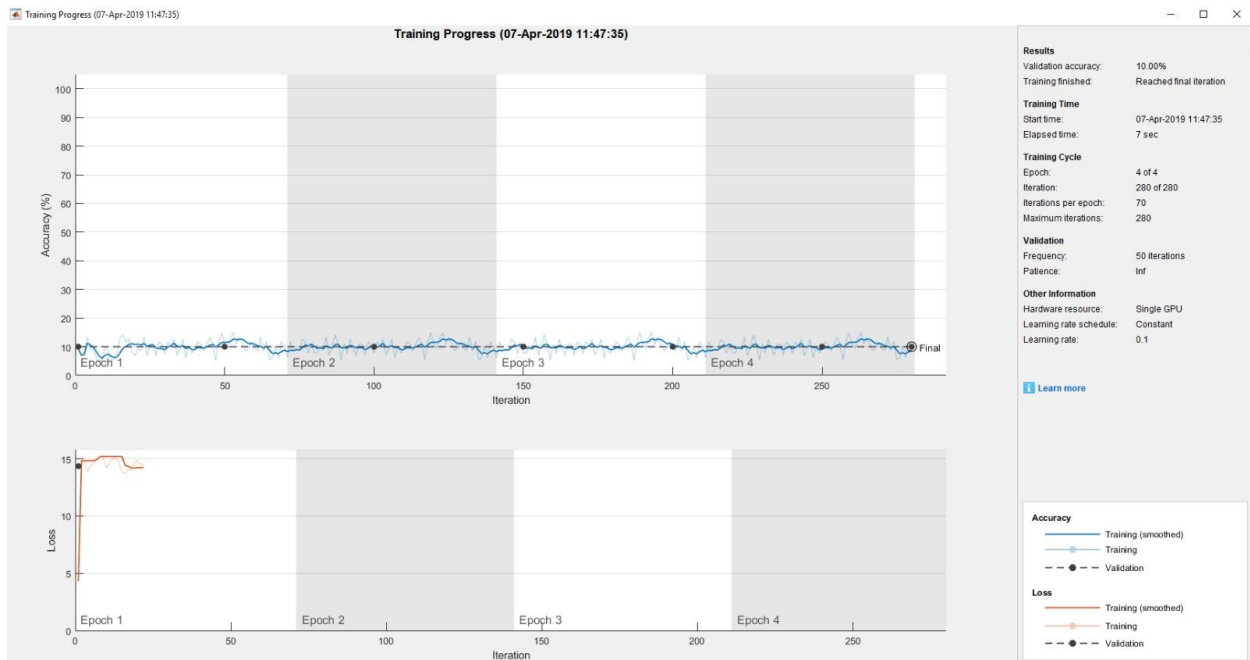
ה-accuracy הסופי של ה-validation: 97.6%

שאלה 2

קיבלנו שה-loss הסופי של האימון יותר נמוך משל ה-validation, זאת מכיוון שכנראה נוצר overfitting לסט האימון. ניתן לפתור את התופעה הזאת ע"י שינוי אקראי של סט האימון או פחות זמן אימון.

פלט ריצה: mnistClassificationDL.m

- יש להוסיף לכל גרף את הפרמטרים איתם ביצעתם את הריצה.
קצב למידה: 0.1



Training on single GPU.

Initializing image normalization.

Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Validation Accuracy	Mini-batch Loss	Validation Loss	Base Learning Rate
1	1	00:00:00	9.38%	10.00%	4.2938	14.3481	0.1000
1	25	00:00:01	10.94%		NaN		0.1000
1	50	00:00:01	10.94%	10.00%	NaN	NaN	0.1000
2	75	00:00:02	8.59%		NaN		0.1000
2	100	00:00:02	7.81%	10.00%	NaN	NaN	0.1000
2	125	00:00:03	13.28%		NaN		0.1000
3	150	00:00:04	7.81%	10.00%	NaN	NaN	0.1000
3	175	00:00:04	9.38%		NaN		0.1000
3	200	00:00:05	9.38%	10.00%	NaN	NaN	0.1000
4	225	00:00:05	7.03%		NaN		0.1000
4	250	00:00:06	9.38%	10.00%	NaN	NaN	0.1000
4	275	00:00:06	5.47%		NaN		0.1000
4	280	00:00:07	10.16%	10.00%	NaN	NaN	0.1000

ה-loss הסופי של האימון: NaN

ה-loss הסופי של ה-validation: NaN

ה-accuracy הסופי של האימון: 10.16%

ה-accuracy הסופי של ה-validation: 10%

שאלה 3

הפעם גרף השגיאה הוא במגמת עלייה (עד כדי יציאה מהמסגרת), זאת מכיוון שגודל הצעד גדול מדי ולכן האלגוריתם של gradient descent לא מצליח להגיע למינימום שגיאה ואף מתרחק ממנה.

גודל הצעד עבורו הרשת באמת מצליחה להתכנס מבחינת loss ו-accuracy הוא 0.001.

פלט ריצה: `mnistClassificationDL.m`

- יש להוסיף לכל גרף את הפרמטרים איתם ביצעתם את הריצה.

שאלה 4

בגודל צעד של 0.001 אנחנו מגיעים לדיוק גבוה יותר מהר. כמו כן, ניתן לראות בטבלה שבגודל הצעד הגדול יותר התופעה של overfitting נעלמת.

שאלה 5

לגודל צעד של 0.00001, נצפה שהאפקט של ה-overfitting יגדל ושקצב ההתקדמות לעבר דיוק גבוה יהיה נמוך יותר.

סעיף a

גודל הצעד המתאים ביותר הוא 0.001: גם מגיע ל-accuracy גבוה וגם אין overfitting. זה הקצב האידיאלי לרשת שלנו: יותר גדול ממנו יביא לחוסר התכנסות ויותר קטן ממנו יביא ל-overfitting.

סעיף b

גודל זה לא יתאים לכל רשת נזירות שנוצרה לאמון, מכיוון שיכול להיות שיהיו רשתות שגודל צעד יותר גדול כן יוביל להתכנסות. רשת עם שכבות שונות תוביל לחישוב גרדיאנט שונה.