

ניסוי 56-חלק א

חלק א'-רגרסיה לוגיסטית

שאלה 1:

מימד ה-W הינו $D+1$.

שאלה 2:

סדר השימוש הנכון עבור תהליך ה-FORWARD :

- 1. AFFINE
- 2. SIGMOID
- 3. NLL

סדר השימוש הנכון עבור תהליך ה-BACKWARD :

- 1. NLL
- 2. SIGMOID
- 3. AFFINE

המימד הצפוי למוצא כל אחת מהפונקציות הללו הוא:

AFFINE_FORWARD :

`z - output (newFeaturesSize x nExamples)`

NLL_FORWARD :

`loss - vector of loss for each example (1 x nSamples)`

SIGMOID_FORWARD :

`z_out - output of the sigmoid (nFeatures x nExamples)`

AFFINE_BACKWARD :

```
dE_dx - gradient w.r.t. x (nFeatures x nExamples)
dE_dW - gradient w.r.t. weight matrix (nFeatures x newFeaturesSize)
dE_db - gradient w.r.t. bias vector (newFeaturesSize x nExamples)
```

:NLL_BACKWARD

```
dE_dpi - gradient w.r.t. the input vector of the NLL (1 x nSamples)
```

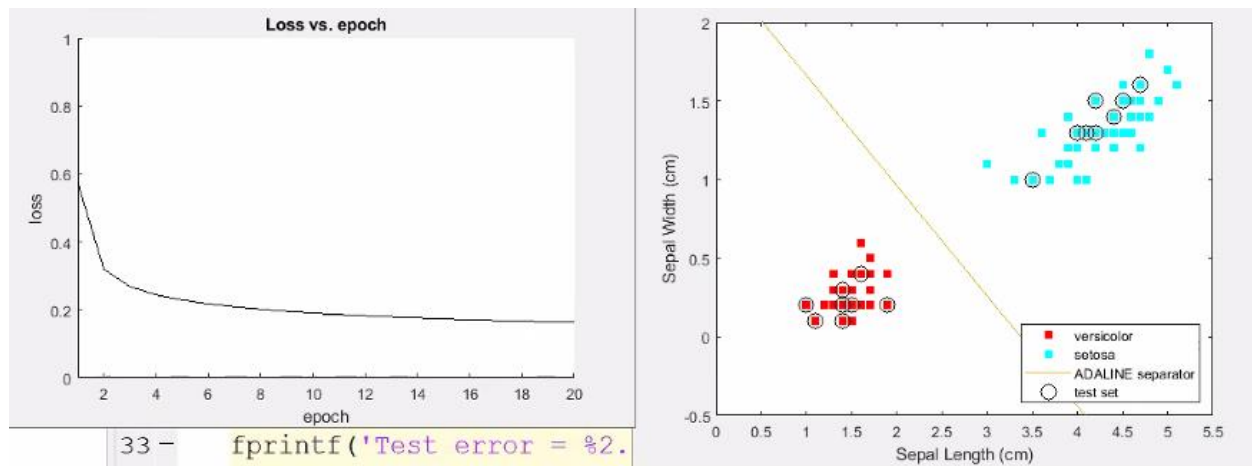
:SIGMOID_BACKWARD

```
dE_dz_in - gradient w.r.t. the inputs of the sigmoid (nFeatures x nExamples)
```

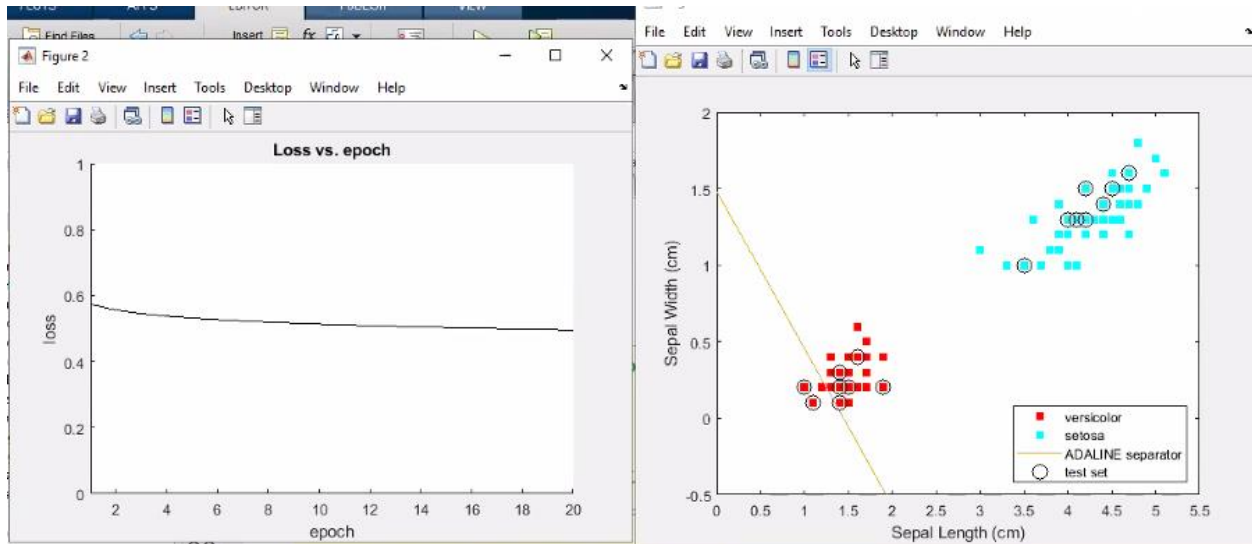
שאלה 3:

אנו מצפים לקבל קו הפרדה לינארי כיוון שוקטור המאפיינים הינו מגודל 2 ומכיל את אורך ורוחב עלי הגביע.

עבור קצב התכנסות של 0.1:



עבור קצב התכנסות של 0.01:



שאלה 4:

עבור גודל צעד נמוך של ה-GRADIENT DESCENT לא הצלחנו להגיע לתוצאה הרצויה בעשרים איטרציות. לעומת זאת, עבור גודל צעד גדול הגענו לתוצאה הרצויה.

חלק ב'-זיהוי ספרות בתמונות

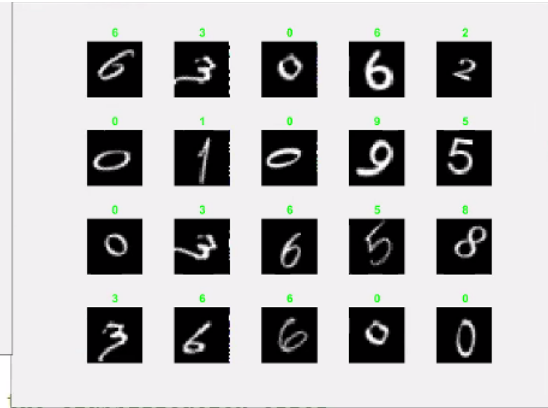
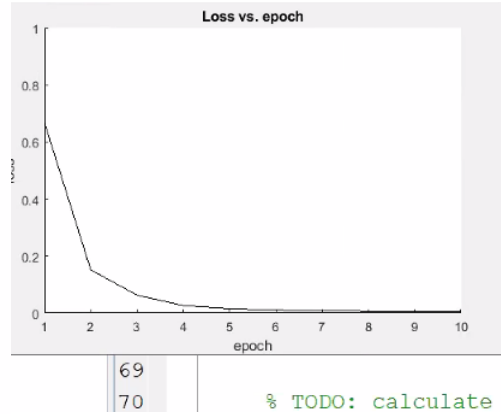
שאלה 1:

הביטויים המתמטיים בעבור שלבי תהליך ה-FORWARD של הרשת הם:

$$X \rightarrow w_1'x + b_1 \rightarrow \text{sigmoid}(w_1'x + b_1) \rightarrow w_2'(\text{sigmoid}(w_1'x + b_1)) + b_2 \rightarrow \text{softmax}(w_2'(\text{sigmoid}(w_1'x + b_1)) + b_2) \rightarrow l(\text{softmax}(w_2'(\text{sigmoid}(w_1'x + b_1)) + b_2), y)$$

הגודל של השכבה הנסתרת צריך להתאים לגודל של שכבת הכניסה והמוצא כך שהמימדים יתאימו ושכבת המוצא תכיל מספר תיגוים נכון.

```
epoch #1
Train error = 22.33%
epoch #2
Train error = 3.06%
epoch #3
Train error = 0.89%
epoch #4
Train error = 0.04%
epoch #5
Train error = 0.00%
epoch #6
Train error = 0.00%
epoch #7
Train error = 0.00%
epoch #8
Train error = 0.00%
epoch #9
Train error = 0.00%
epoch #10
Train error = 0.00%
Final train error: 0.00%
Validation:
Validation error: 0.40%
```



שאלה 2:

גודל השכבה הנסתרת בקוד הינו 700. ככל שהשכבה תגדל נצטרך יותר מקום בזיכרון על מנת לאחסן אותה. בנוסף, עבור שכבה גדולה יותר זמן החישוב יהיה ארוך יותר. יתר על כן, גודל השכבה הנסתרת משפיע על הדיוק.

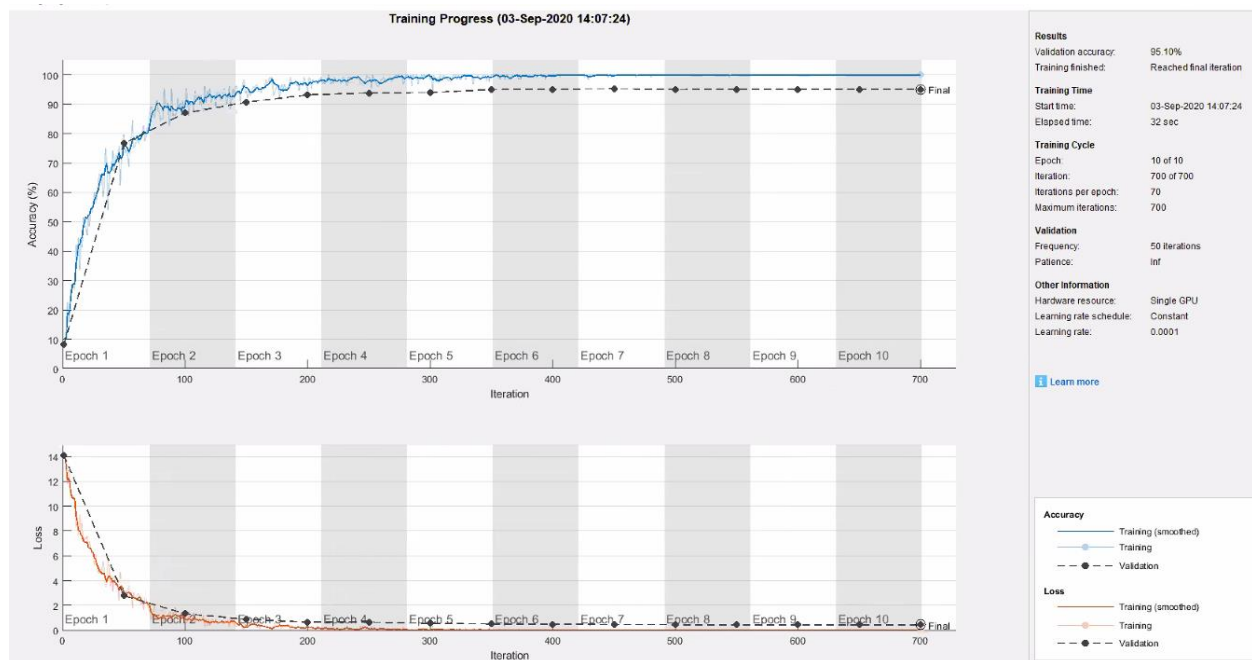
חלק ג'-סיווג באמצעות MATLAB NEURAL NETWORK TOOLBOX

שאלה 1:

המשמעות של כל אחד מהפרמטרים המתקבלים כקלט לפונקציה TRAININGOPTIONS בקובץ MNISTCLASSIFICATIONDL:

- SGDM-לבצע גרדיאנט דיסנט.
- INITIAL RATE-גודל הצעד ההתחלתי.
- MAXEPOCH-מספר האפוקים המקסימאלי שעוצרים את האלגוריתם לאחר שהגענו אליו.
- VALIDATION DATA-מציין את ה-DATA עבורו אנו רוצים לבצע ואלידציה.
- VALIDATION FREQUENCY-לאחר כמה איטרציות אנו רוצים לבצע הדפסה לחלון.
- VERBOSE-מציין האם להציג מידע תוך כדי אימון.
- VERBOSEFREQUENCY-מציין כל כמה איטרציות להדפיס בחלון את המידע של האימו.
- PLOTS-לצייר את התוצאות באופן גרפי.
- TRAINING-PROGRESS-לצייר את התוצאות במהלך אימון.

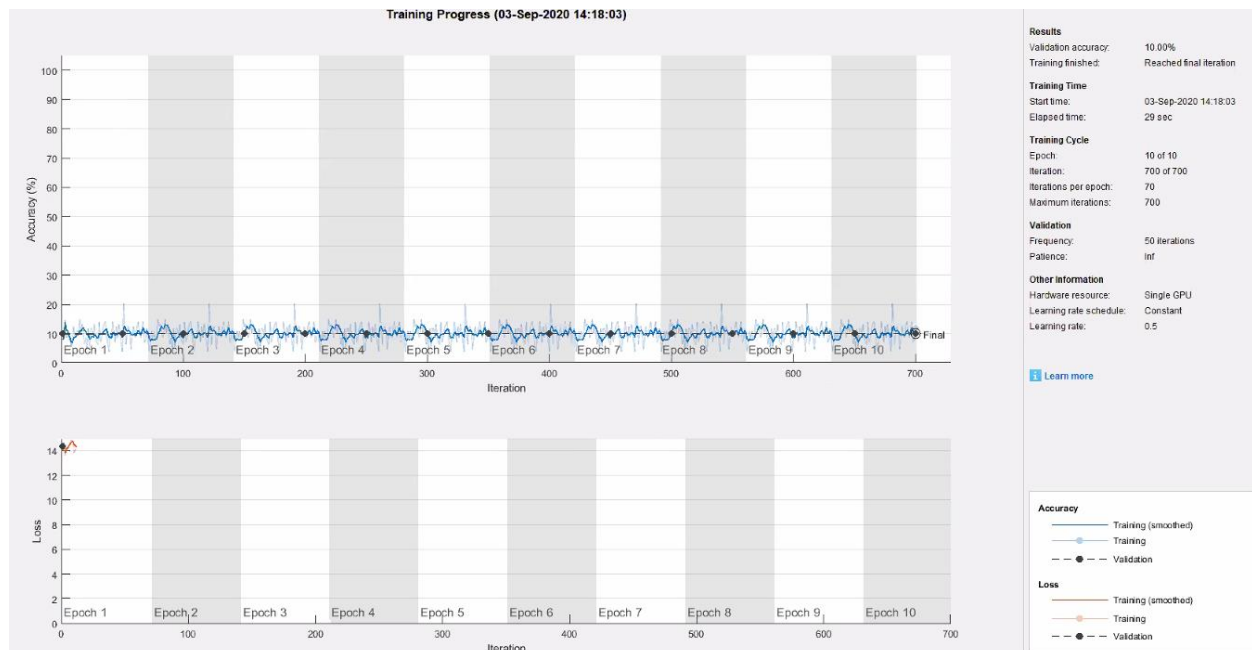
10 EPOCH עם גודל צעד של 0.0001:



Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Validation Accuracy	Mini-batch Loss	Validation Loss	Base Learning Rate
1	1	00:00:03	10.16%	8.20%	14.1967	14.1193	1.0000e-04
1	25	00:00:05	56.25%		6.0273		1.0000e-04
1	50	00:00:06	79.69%	76.70%	2.6256	2.7833	1.0000e-04
2	75	00:00:07	85.16%		2.0080		1.0000e-04
2	100	00:00:08	91.41%	87.00%	0.7501	1.3337	1.0000e-04
2	125	00:00:09	89.06%		0.9070		1.0000e-04
3	150	00:00:10	94.53%	90.70%	0.3206	0.8797	1.0000e-04
3	175	00:00:11	96.09%		0.3362		1.0000e-04
3	200	00:00:12	94.53%	93.20%	0.3942	0.6540	1.0000e-04
4	225	00:00:13	98.44%		0.0636		1.0000e-04
4	250	00:00:14	99.22%	93.80%	0.0184	0.6561	1.0000e-04
4	275	00:00:15	99.22%		0.0397		1.0000e-04
5	300	00:00:16	98.44%	94.00%	0.0442	0.5507	1.0000e-04
5	325	00:00:17	98.44%		0.0246		1.0000e-04
5	350	00:00:18	100.00%	95.00%	0.0009	0.5071	1.0000e-04
6	375	00:00:19	100.00%		0.0021		1.0000e-04
6	400	00:00:20	100.00%	95.10%	0.0003	0.4814	1.0000e-04
7	425	00:00:21	100.00%		0.0002		1.0000e-04
7	450	00:00:22	100.00%	95.30%	0.0023	0.4617	1.0000e-04
7	475	00:00:23	100.00%		0.0004		1.0000e-04
8	500	00:00:24	100.00%	95.00%	0.0003	0.4497	1.0000e-04
8	525	00:00:25	100.00%		0.0007		1.0000e-04
8	550	00:00:26	100.00%	95.10%	3.8878e-05	0.4432	1.0000e-04
9	575	00:00:27	100.00%		1.6276e-05		1.0000e-04
9	600	00:00:28	100.00%	95.10%	5.6313e-05	0.4395	1.0000e-04
9	625	00:00:29	100.00%		7.3306e-05		1.0000e-04
10	650	00:00:30	100.00%	95.10%	8.2642e-05	0.4435	1.0000e-04
10	675	00:00:31	100.00%		2.3524e-05		1.0000e-04
10	700	00:00:32	100.00%	95.10%	4.5187e-05	0.4445	1.0000e-04

Validation error: 4.90%

EPOCH 10 עם גודל צעד של 0.5:



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	7.81%	9.90%	14.4128	14.3641	0.5000
1	25	00:00:01	11.72%		NaN		0.5000
1	50	00:00:02	3.91%	10.00%	NaN	NaN	0.5000
2	75	00:00:03	8.59%		NaN		0.5000
2	100	00:00:04	6.25%	10.00%	NaN	NaN	0.5000
2	125	00:00:05	12.50%		NaN		0.5000
3	150	00:00:06	12.50%	10.00%	NaN	NaN	0.5000
3	175	00:00:07	13.28%		NaN		0.5000
3	200	00:00:08	9.38%	10.00%	NaN	NaN	0.5000
4	225	00:00:09	14.84%		NaN		0.5000
4	250	00:00:10	9.38%	10.00%	NaN	NaN	0.5000
4	275	00:00:11	8.59%		NaN		0.5000
5	300	00:00:12	8.59%	10.00%	NaN	NaN	0.5000
5	325	00:00:13	12.50%		NaN		0.5000
5	350	00:00:15	11.72%	10.00%	NaN	NaN	0.5000
6	375	00:00:16	11.72%		NaN		0.5000
6	400	00:00:17	3.91%	10.00%	NaN	NaN	0.5000
7	425	00:00:18	8.59%		NaN		0.5000
7	450	00:00:19	6.25%	10.00%	NaN	NaN	0.5000
7	475	00:00:20	12.50%		NaN		0.5000
8	500	00:00:21	12.50%	10.00%	NaN	NaN	0.5000
8	525	00:00:22	13.28%		NaN		0.5000
8	550	00:00:23	9.38%	10.00%	NaN	NaN	0.5000
9	575	00:00:24	14.84%		NaN		0.5000
9	600	00:00:25	9.38%	10.00%	NaN	NaN	0.5000
9	625	00:00:26	8.59%		NaN		0.5000
10	650	00:00:27	8.59%	10.00%	NaN	NaN	0.5000
10	675	00:00:28	12.50%		NaN		0.5000
10	700	00:00:29	11.72%	10.00%	NaN	NaN	0.5000

Validation error: 100.00%

~%

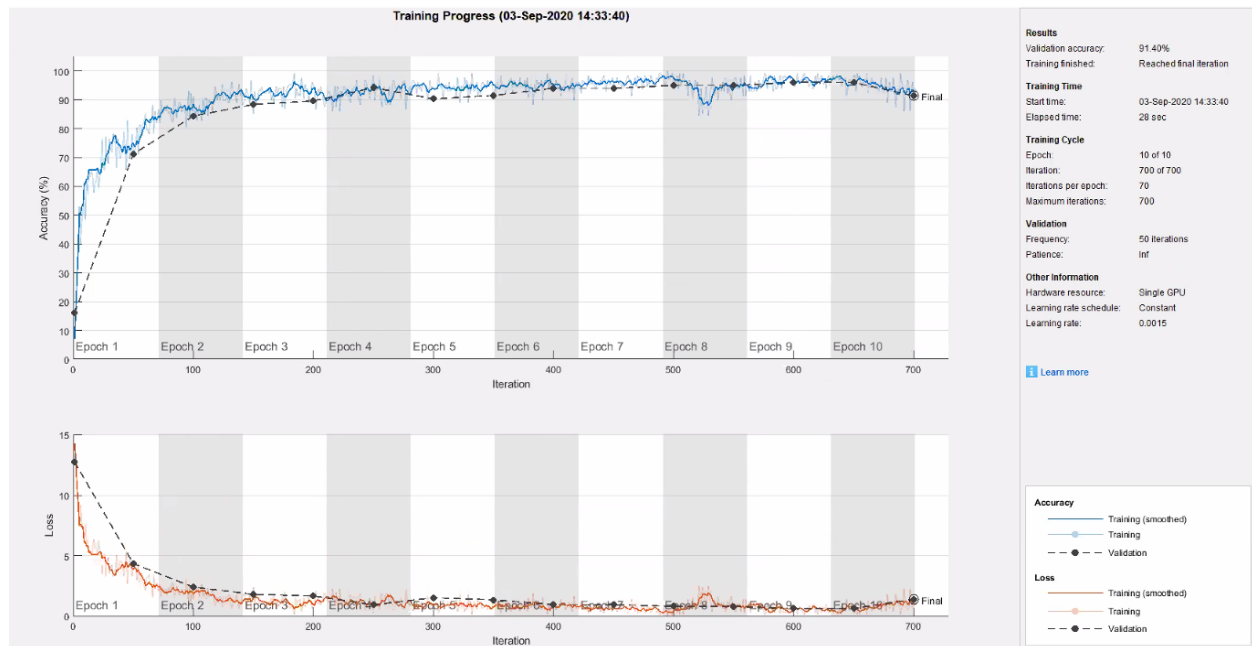
שאלה 2:

ניתן לראות שה-LOSS קטן יותר ב-TRAINING מאשר בואלידציה כיוון שהמערכת למדה מתוך ה-TRAINING והמשקולות הותאמו בהתאם בעוד שבסט ואלידציה לא נעשו שינויים במשקולות (התאמת יתר ב- training DATA ביחס לואלידציה).

שאלה 3:

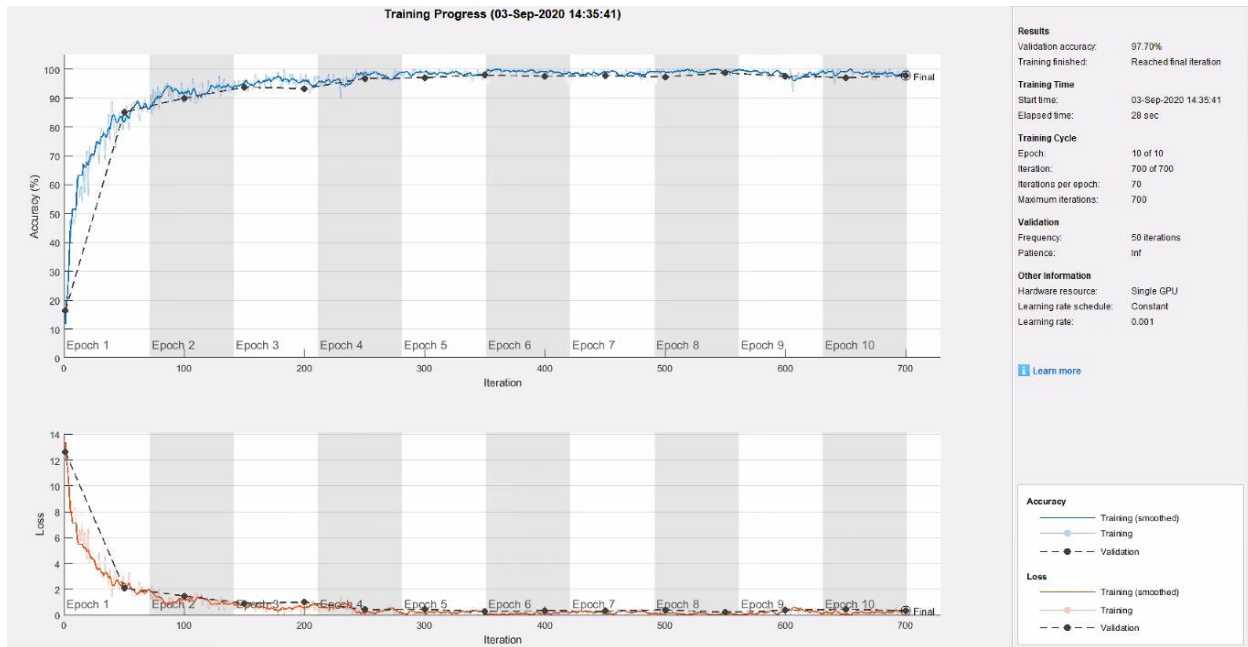
הגרף אינו מתכנס מהסיבה שצעד בגודל 0.5 גדול מדי .

גודל הצעד הגדול ביותר עבורו הרשת מצליחה להתכנס הוא 0.0015:

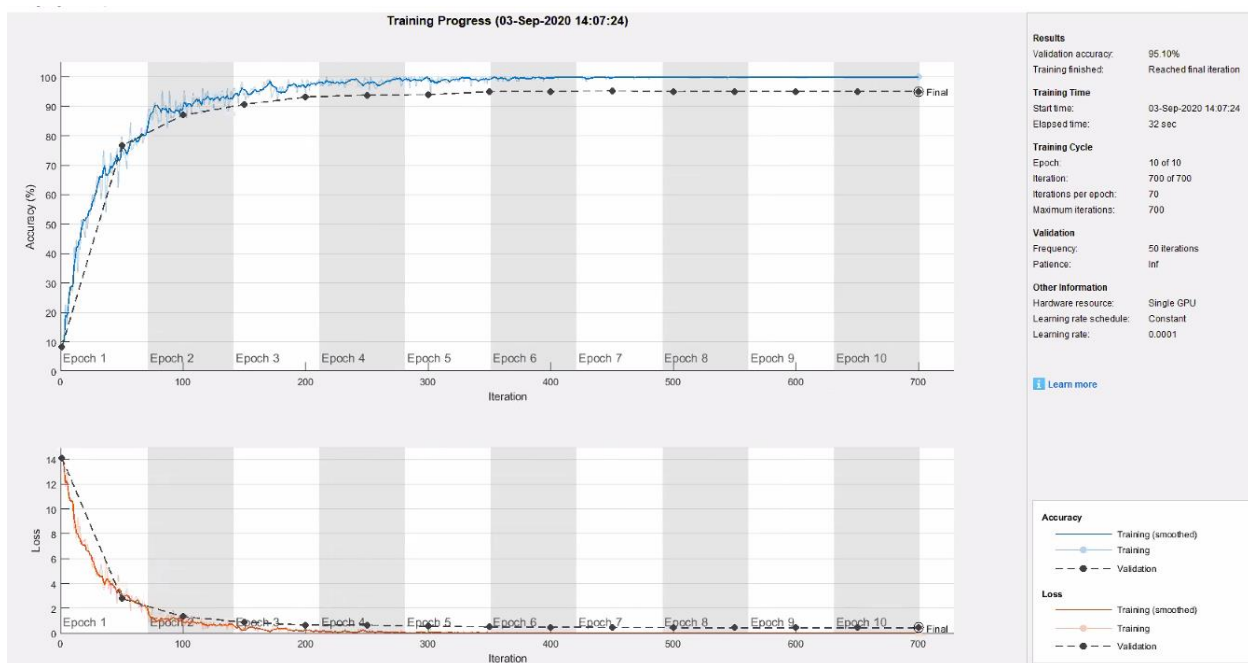


שאלה 4:

עבור גודל צעד של 0.001:



עבור גודל צעד של 0.0001:



ניתן לראות שעבור גדלי צעדים שונים נקבל התכנסות לערכים שונים. אצלנו רואים שגודל הצעד קטן יותר נקבל TEST ERROR גדול יותר, להערכתנו זה קשור ב-OVERFITTING מסעיפים קודמים.

שאלה 5:

סעיף א:

גודל הצעד המתאים ביותר מבין כל שלושת האופציות הינו 0.001 והוא אידיאלי לאורך כל תהליך ההתכנסות.

סעיף ב:

גודל זה לא יתאים לכל רשת נוירונים שנרצה לאמן כיוון שהוא תלוי בגודל המידע שלנו סוג המידע שלנו מספר התיוגים שלנו באופי הרשת מספר שכבות שיש לנו ברשת ובעוד גורמים .