**ניסוי 56-חלק א**

**חלק א'-רגרסיה לוגיסטית**

**שאלה 1:**

מימד ה-W הינו 1+D .

**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:

FORWARD\_AFFINE:



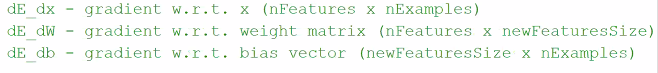
FORWARD\_NLL:



FORWARD\_SIGMOID:



BACKWARD\_AFFINE:



BACKWARD\_NLL:



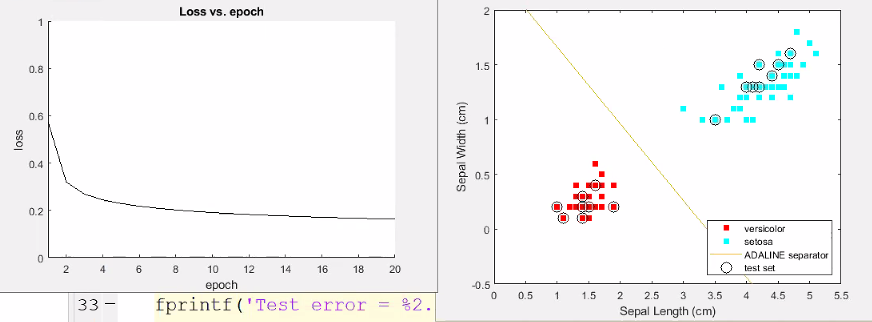
BACKWARD\_SIGMOID:



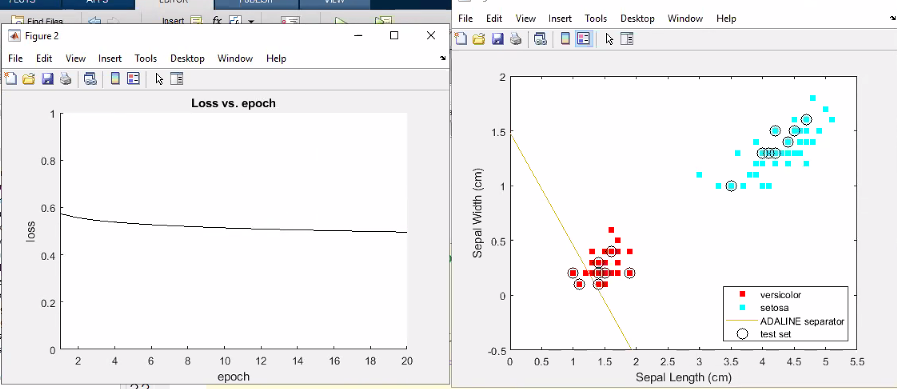
**שאלה 3:**

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:

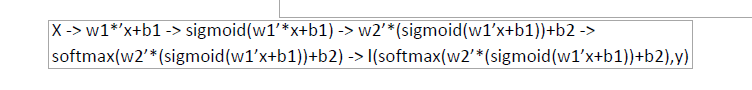


**שאלה 4:**

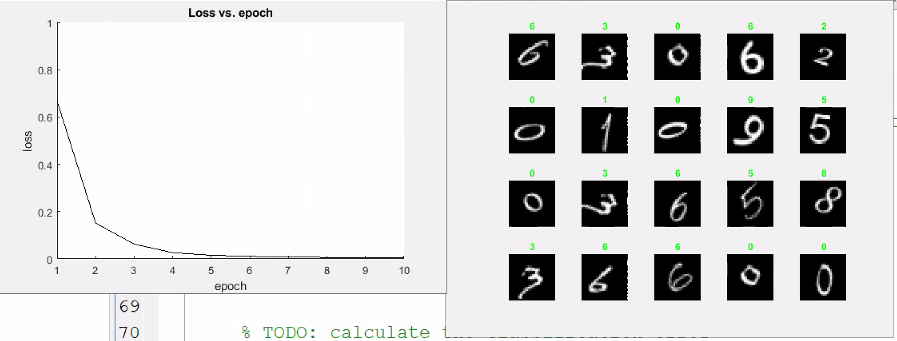
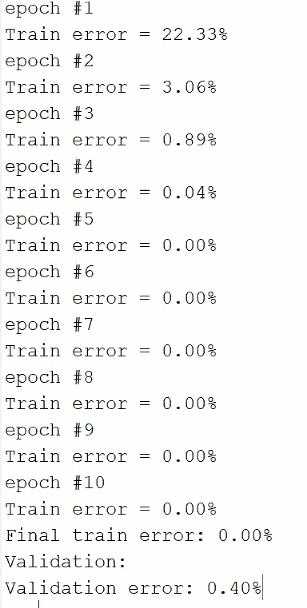
עבור גודל צעד נמוך של ה-GRADIENT DECENT לא הצלחנו להגיע לתוצאה הרצויה בעשרים איטרציות.לעומת זאת, עבור גודל צעד גדול הגענו לתוצאה הרצוייה.

**חלק ב'-זיהוי ספרות בתמונות**

**שאלה 1:**

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

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



**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.

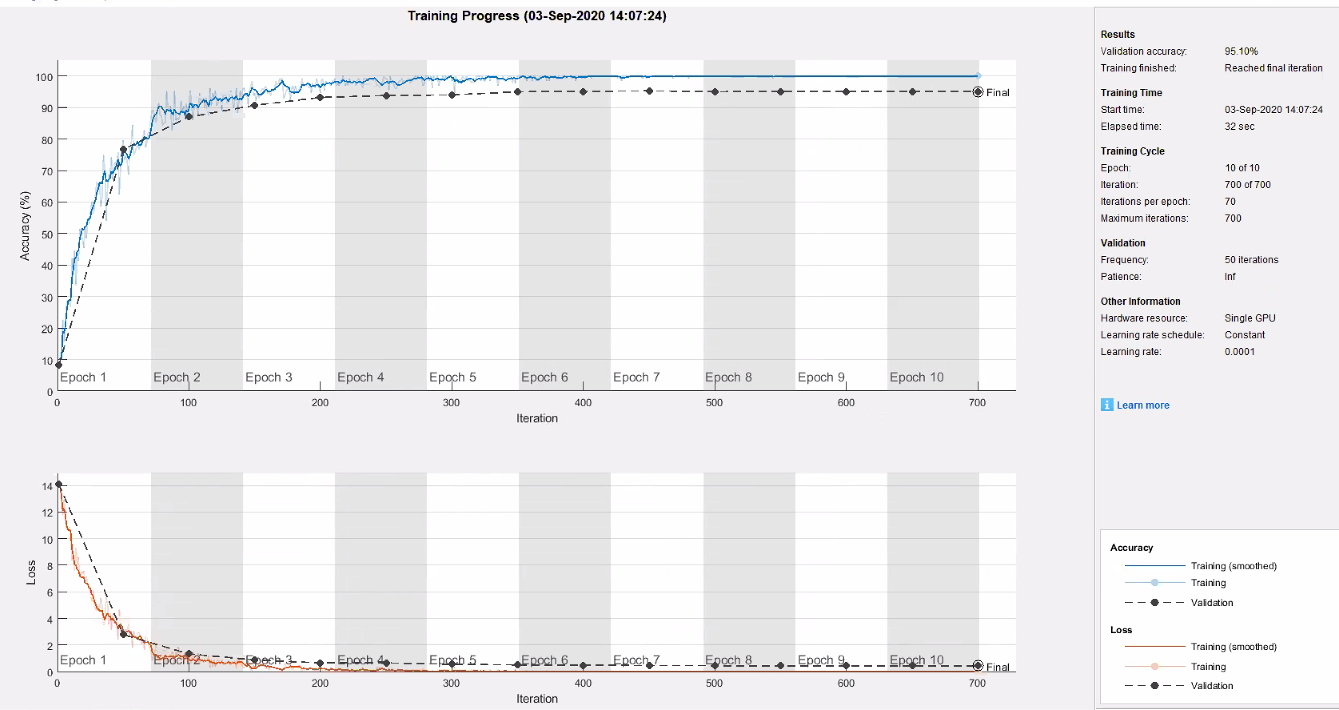
**חלק ג'-סיווג באמצעות MATLAB NEURAL NETWORK TOOLBOX**

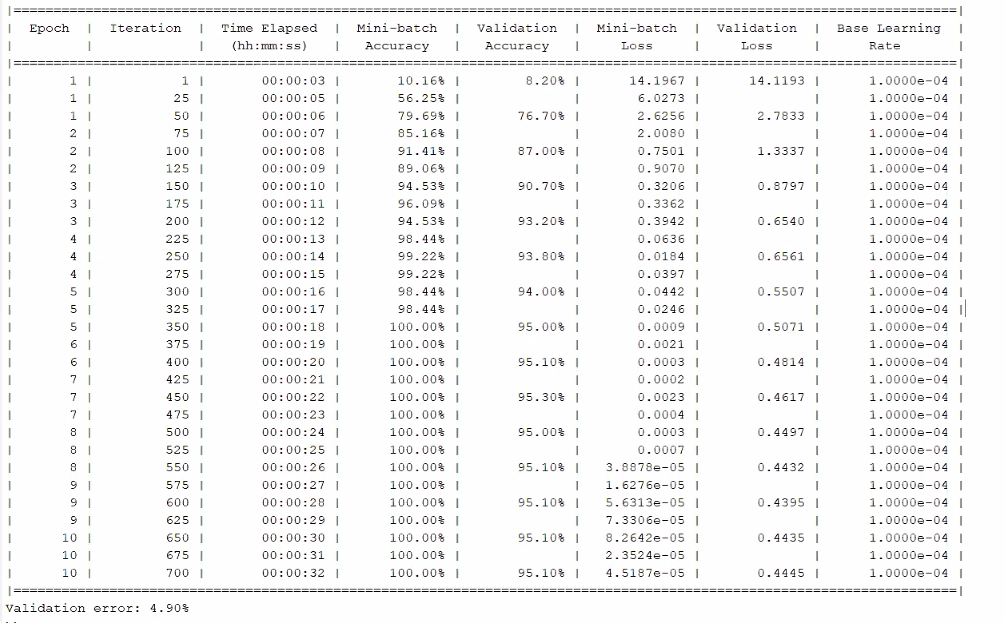
**שאלה 1:**

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

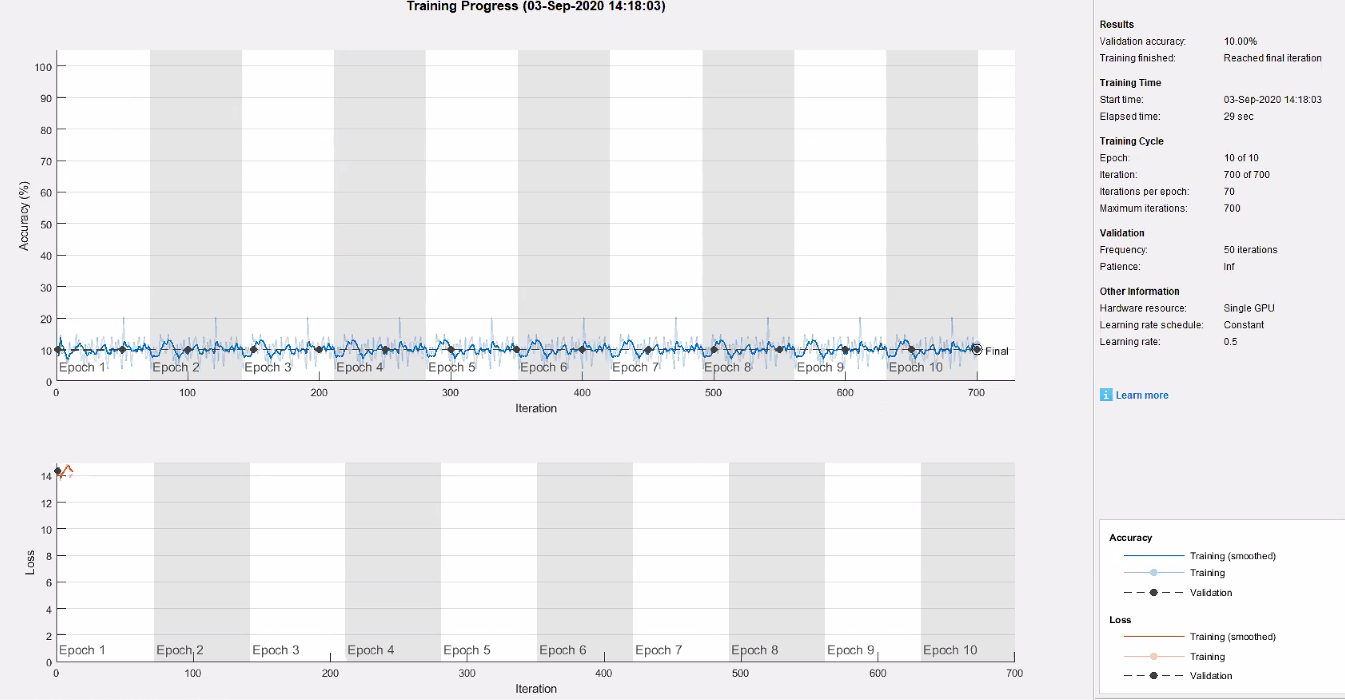
* SGDM-לבצע גרדיאנט דיסנט.
* INITIAL RATE-גודל הצעד ההתחלתי.
* MAXEPOCH-מספר האפוקים המקסימאלי שעוצרים את האלגוריתם לאחר שהגענו אליו.
* VALIDATION DATA-מציין את ה-DATA עבורו אנו רוצים לבצע ואלידציה.
* VALIDATION FREQUENCY-לאחר כמה איטרציות אנו רוצים לבצע הדפסה לחלון.
* VERBOSE-מציין האם להציג מידע תוך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

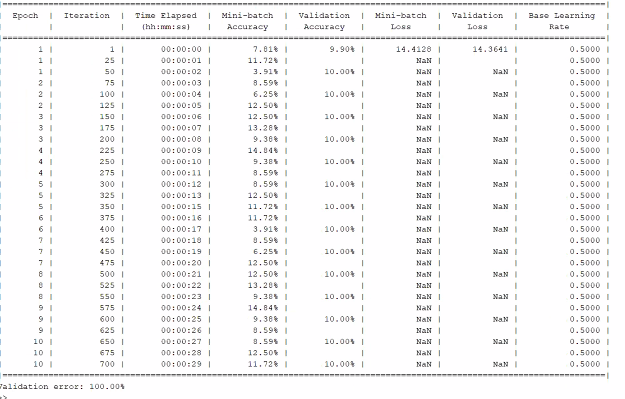
**10 EPOCH עם גודל צעד של 0.0001:**





**10 EPOCH עם גודל צעד של 0.5:**





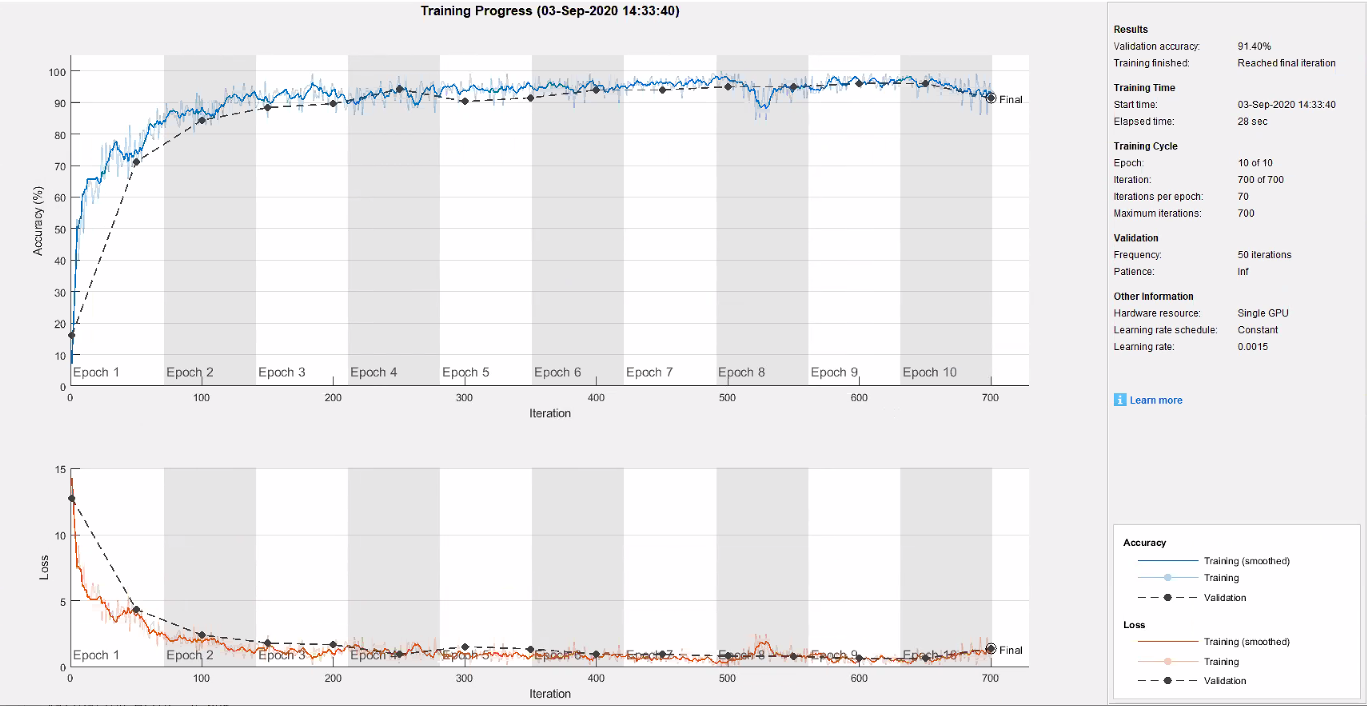
**שאלה 2:**

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

**שאלה 3:**

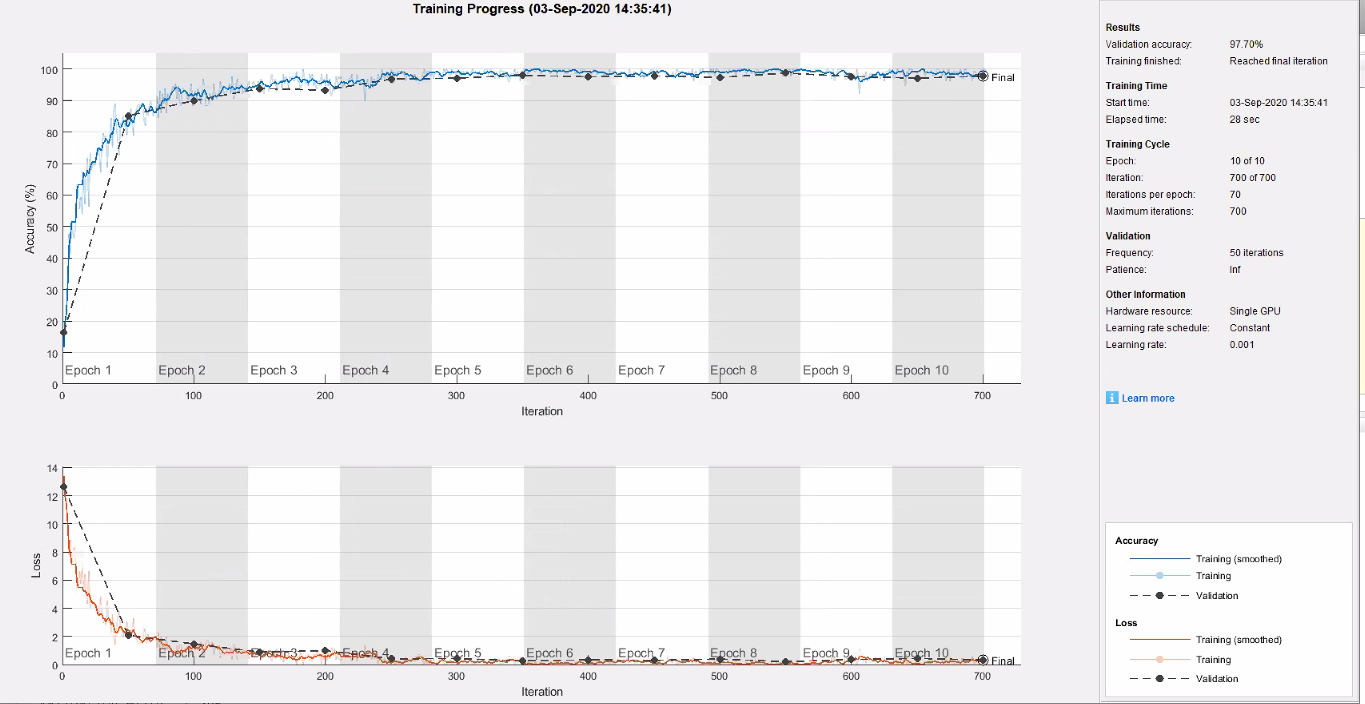
הגרף אינו מתכנס מהסיבה שצעד בגודל 0.5 גדול מידי .

גודל הצעד הגדול ביותר עבורו הרשת מצליחה להתכנס הוא 0.0015:

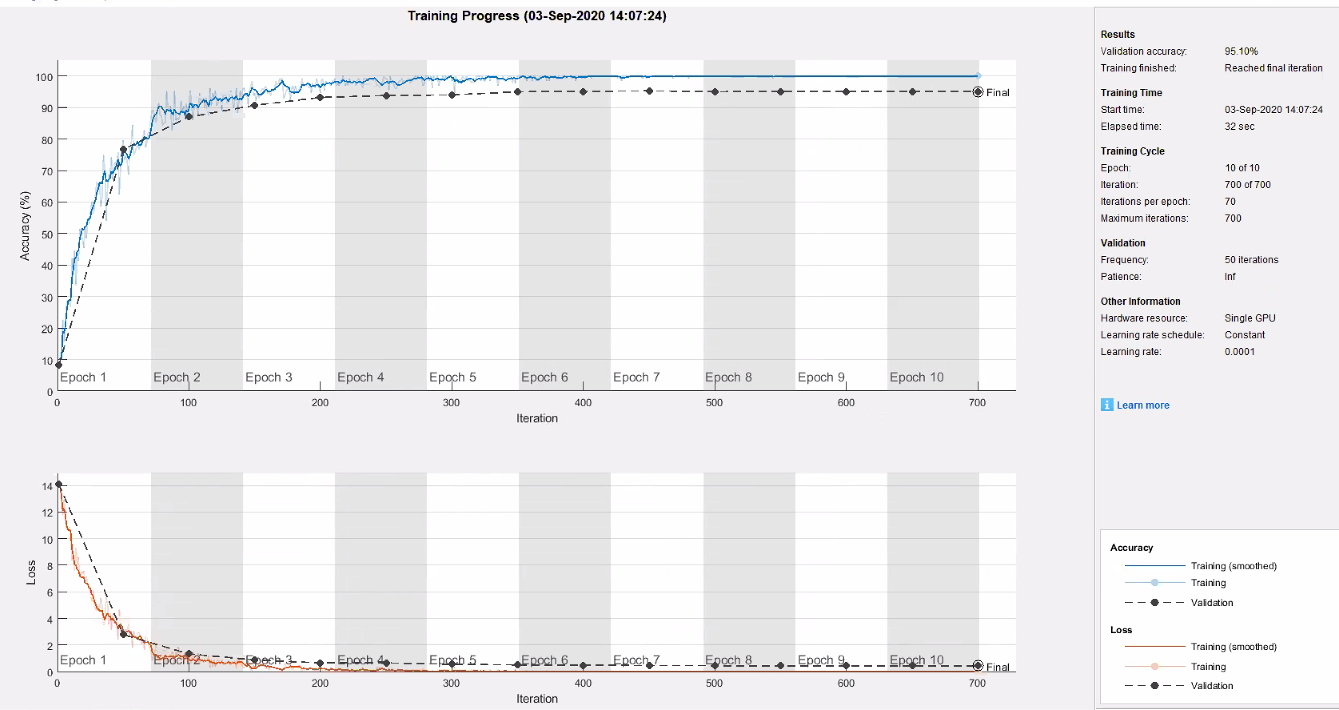


**שאלה 4:**

**עבור גודל צעד של 0.001:**



**עבור גודל צעד של 0.0001:**



ניתן לראות שעבור גדלי צעדים שונים נקבל התכנסות לערכים שונים.אצלנו רואים שככל 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 .