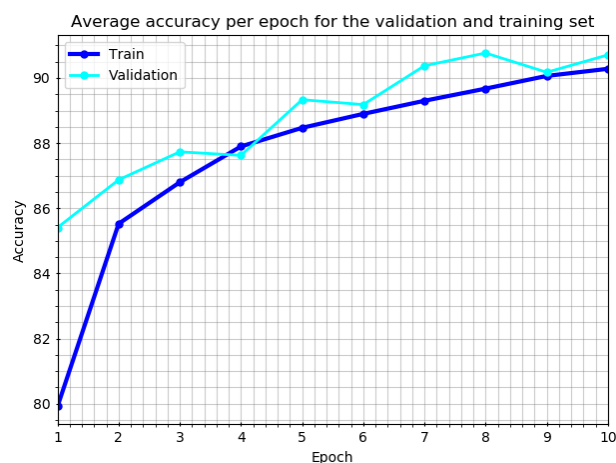
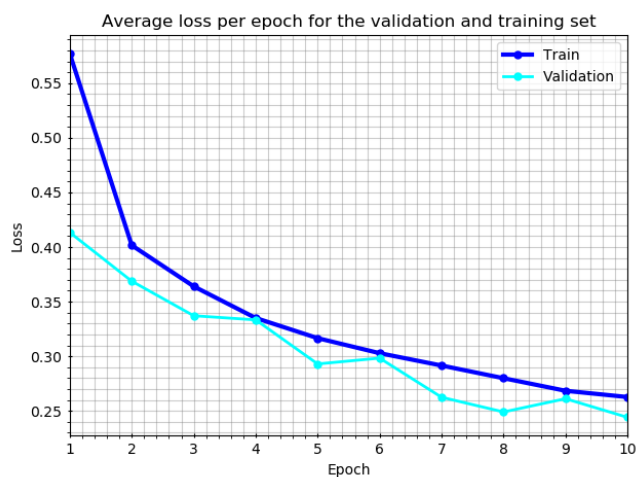


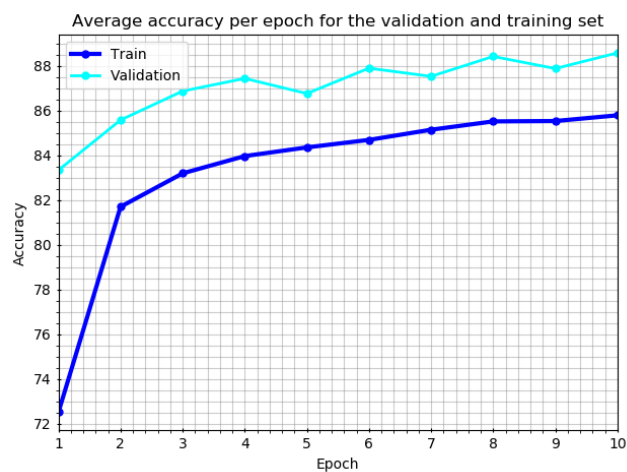
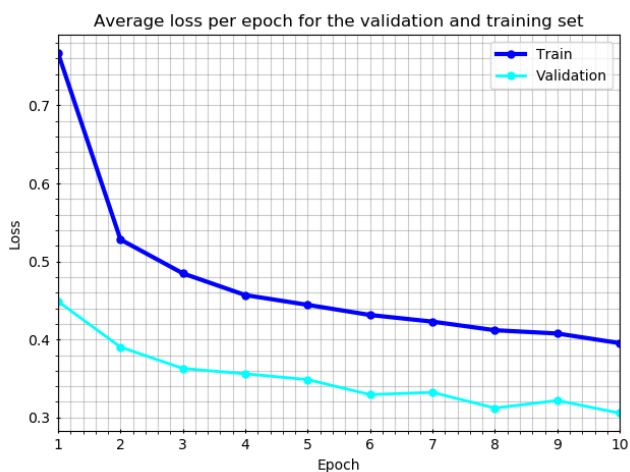
## למידת מכונה תרגיל 4 אלעד ישראל 31344888

ממוצע loss וממוצע דיוק ל training set ול- validation set כפונקציה של מס' epoch לכל מודל:

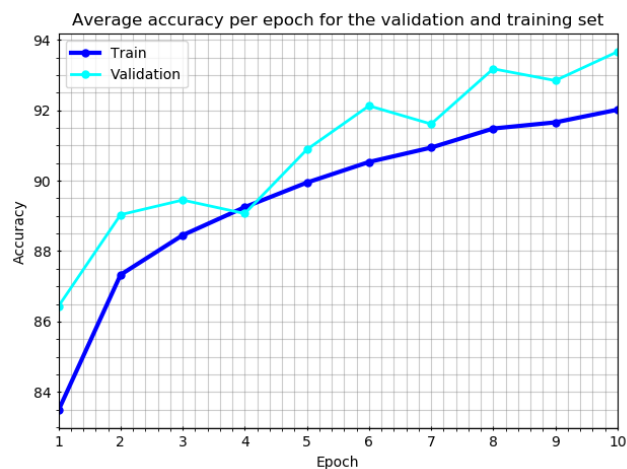
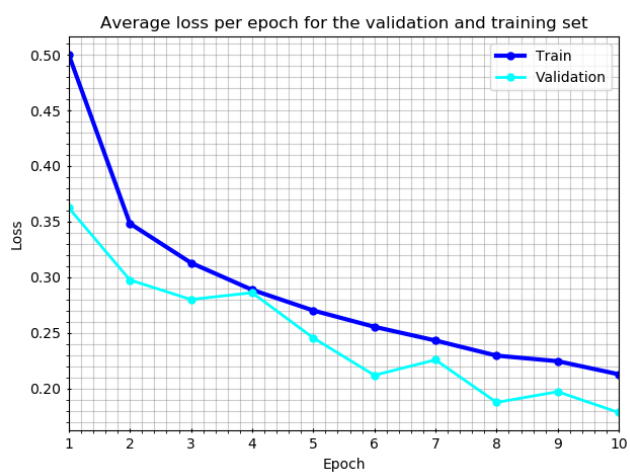
### מודל A:



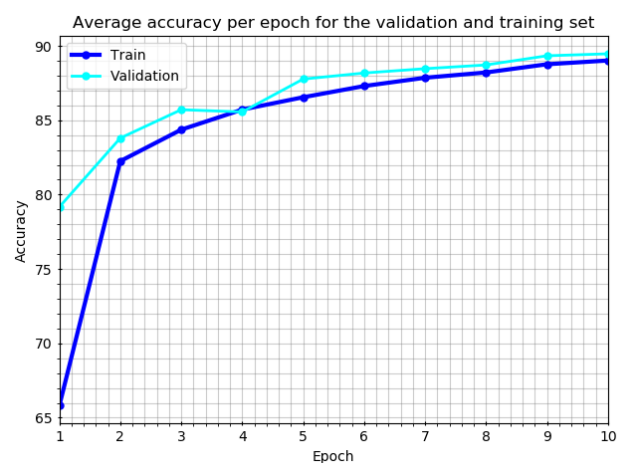
### מודל B:



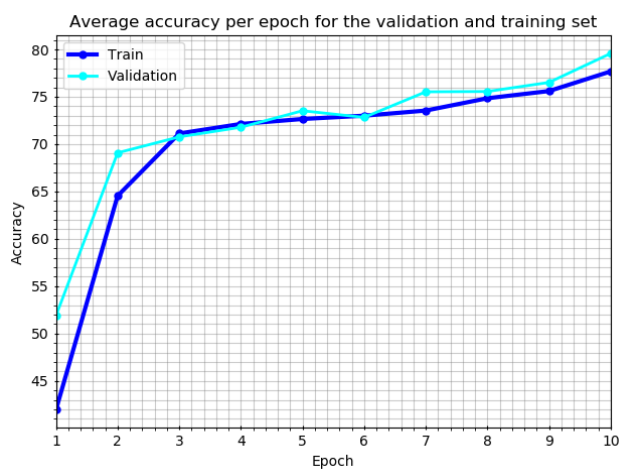
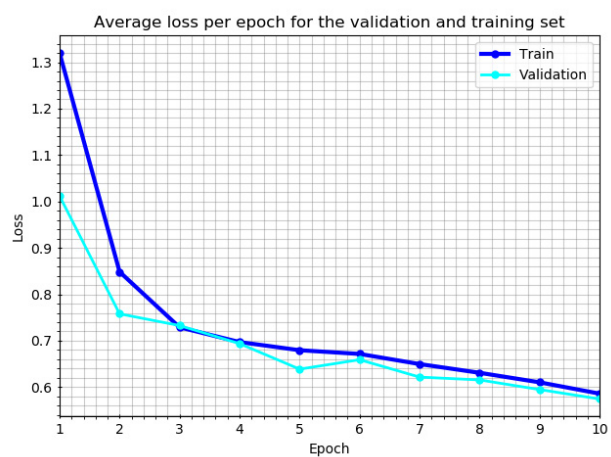
## מודל C:



## מודל D:



## מודל E:



## המודל שבחרתי: C

Validation set accuracy	Test set accuracy (original FashionMNIST test set)	מודל
90.70%	87.35%	A
88.57%	86.40%	B
<b>93.65%</b>	<b>88.32%</b>	<b>C</b>
89.46%	86.87%	D
81.04%	80.09%	E

### Hyper-parameters:

**eta:** ניסיתי ערכי eta רבים ושונים מכיוון שלבחירת eta הייתה השפעה גדולה מאוד על אחוזי ההצלחה.

עבור ערכי eta גבוהים קיבלתי אחוזי הצלחה נמוכים יחסית. ככל שהקטנתי את eta התחלתי לקבל אחוזים גבוהים יותר, עד לגבול מסויים.

לבסוף בחרתי ב-0.001 מכיוון שזו הייתה eta שמיקסמה את אחוזי ההצלחה לרוב המודלים.

במודל E בלבד השארתי את eta גבוהה יותר - 0.01 מכיוון שרק עבורה מודל זה eta נמוכה יותר הורידה את אחוזי ההצלחה.

**epochs:** בכל המודלים מספר epochs הוא 10 מכיוון שזאת כמות האיטרציות שנדרשה בתרגיל

**אקטיבציה:** לפי דרישות התרגיל לכל מודל.

**נרמול:** נרמלתי את המידע הנתון לטווח [0,1] מכיוון שזה נתן התפזרות טובה ולכן תוצאות טובות.

**optimizer:** Adam נתן אחוזי הצלחה גבוהים מאוד! ניסיון לשנות לSGD לדוגמא הוריד את אחוזי ההצלחה של כל המודלים, לחלקם אף בעשרות אחוזים!

**dropout:** במודל B ערך dropout הטוב ביותר שעזר למנוע overfitting (גרם לכך שהרשת לא קיבלה כל פעם את אותו המידע ולכן מונעת היקשרות של נירונים משכבות שונות ועוזרת לתת דגש לנירונים שאנחנו רוצים) היה dropout=0.4.

**batch size:** ניסיתי גדלים שונים, אך 64 הביא לתוצאות הכי טובות. ערך 1 לדוגמא הוריד את אחוזי ההצלחה לאחוזים בודדים.