

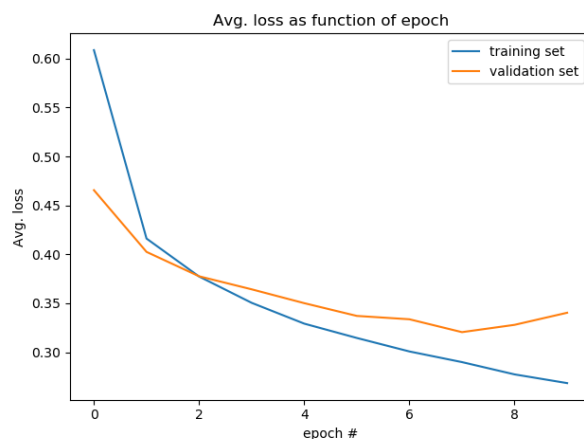
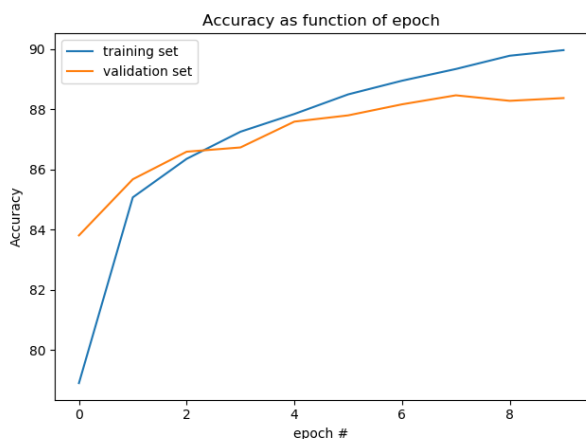
## תרגיל 4

רועי פוקס 205415342

בכל המודלים השתמשתי ב-learning rate של 0.001, ובערך הדיפולטיבי של Adam:  $\gamma_1 = 0.9, \gamma_2 = 0.999$ . בנוסף, השתמשתי בגודל batch של 64 ובנרמול הנתונים שמתקבלים בין 0 ל-1. כל מודל רץ במשך 10 חזרות, כפי שהתבקשנו בתרגיל. המודל עם הלוס הקטן ביותר יצא מודל C, ובו השתמשתי גם להגשה.

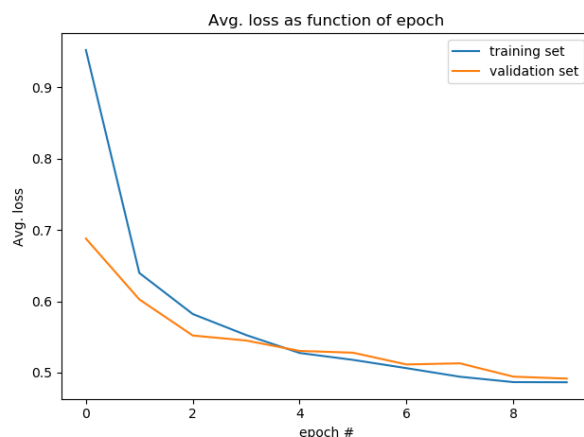
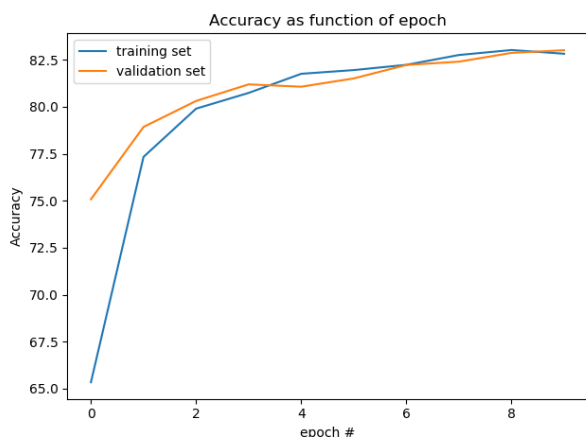
### 1 Model A

מודל זה השתמש ב-2 שכבות נסתרות עם פונקציית אקטיבציה  $ReLU$ . גדלי השכבות:  $[100, 50]$  עבור סט המבחן, הגעתי לדיוק של: 88%.



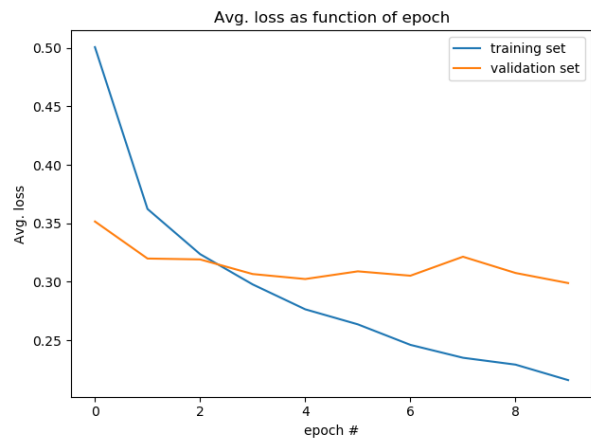
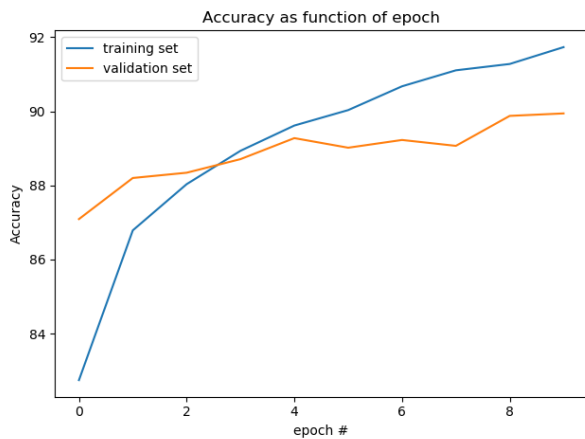
### 2 Model B

מודל זה, הוא כמו מודל A, עם הוספה של  $dropout$  לאחר כל שכבה נסתרת. ההסתברות להתעלמות - 50%. עבור סט המבחן, הגעתי לדיוק של: 86%. בעת חישוב הולדישן, השתמשתי במוד אימון, על מנת להשוות אותם באותו כח חישוב (ולא חלק ע"י השבתה של חלק מהניורונים וחלק לא).



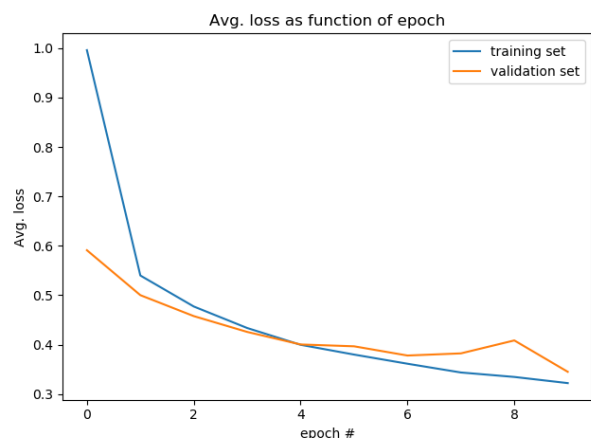
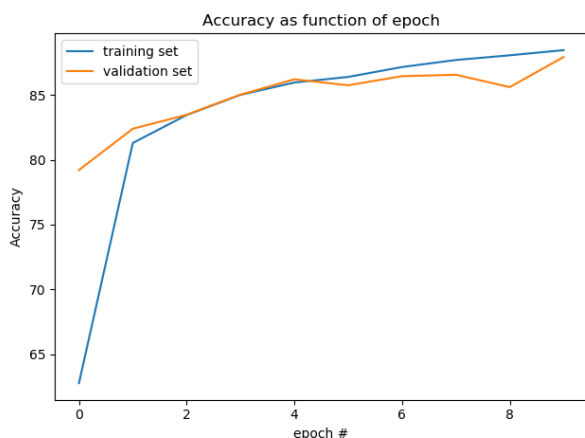
### Model C 3

מודל זה, הוא כמו מודל A, עם הוספה של נרמול ה-Batch לפני כל פונקציית אקטיבציה. עבור סט המבחן, הגעתי לדיוק של: 88%. זה גם הסט שבחרתי להשתמש בו.



### Model D 4

מודל זה כולל 5 שכבות נסתרות עם פונקציית האקטיבציה  $ReLU$ . גדלי השכבות:  $[128, 64, 10, 10, 10]$ . עבור סט המבחן, הגעתי לדיוק של: 87%.



### Model E 5

מודל זה כולל 5 שכבות נסתרות עם פונקציית האקטיבציה  $Sigmoid$ . גדלי השכבות:  $[128, 64, 10, 10, 10]$ . עבור מודל זה, השתמשתי ב-learning rate שונה בגובה 0.01. עבור סט המבחן, הגעתי לדיוק של: 82%.

