

### תרגיל 3-למידת מכונה

דניאל מושיוף - Daniel Moshayof

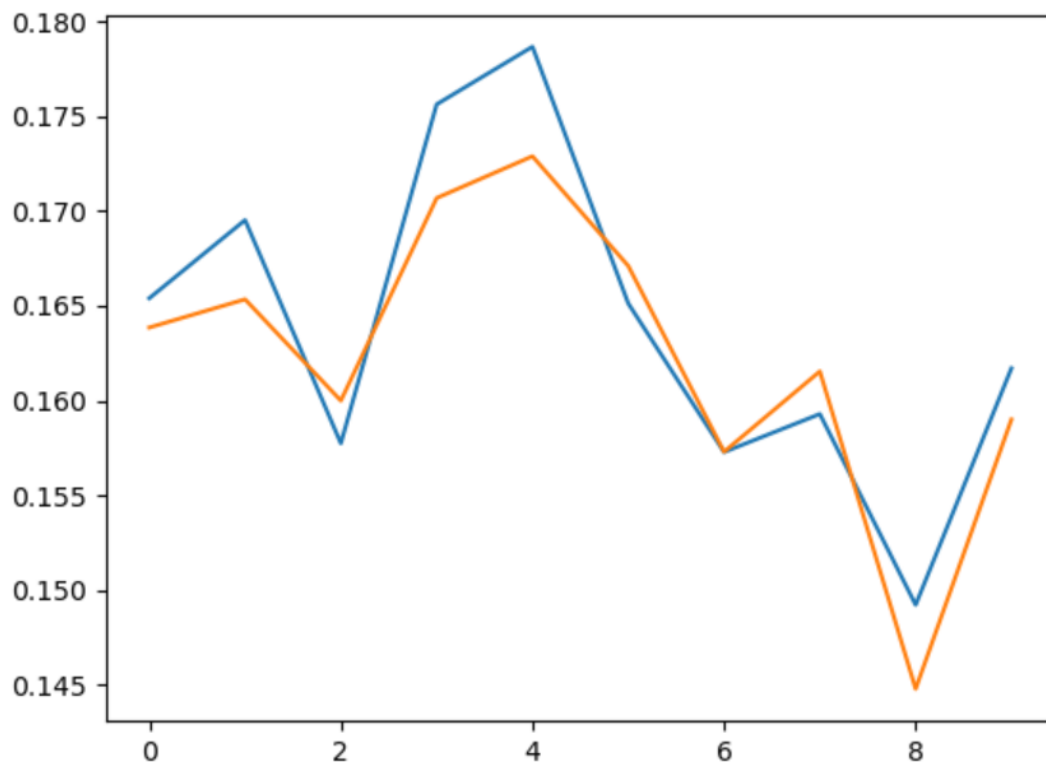
dmoshayof@gmail.com

31126668

המודל שבחרתי להשתמש בו היא רשת נירונים בעלת שכבה נסתרת אחת MLP1 תוך שימוש ב-  
בפונקציית אקטיבציה-ReLU על מנת להפוך את המודל ללא לינארי. בנוסף, השתמשתי בייצוג  
התוצאה ב-one hot vector תוך שימוש ב-softmax לשם כך. כדי להגיע לדיוק מירבי ניסיתי לשנות את  
ההיפרפרמטרים תוך צפייה בגרף המציג את ה- train accuracy (כחול) מול ה- dev accuracy.

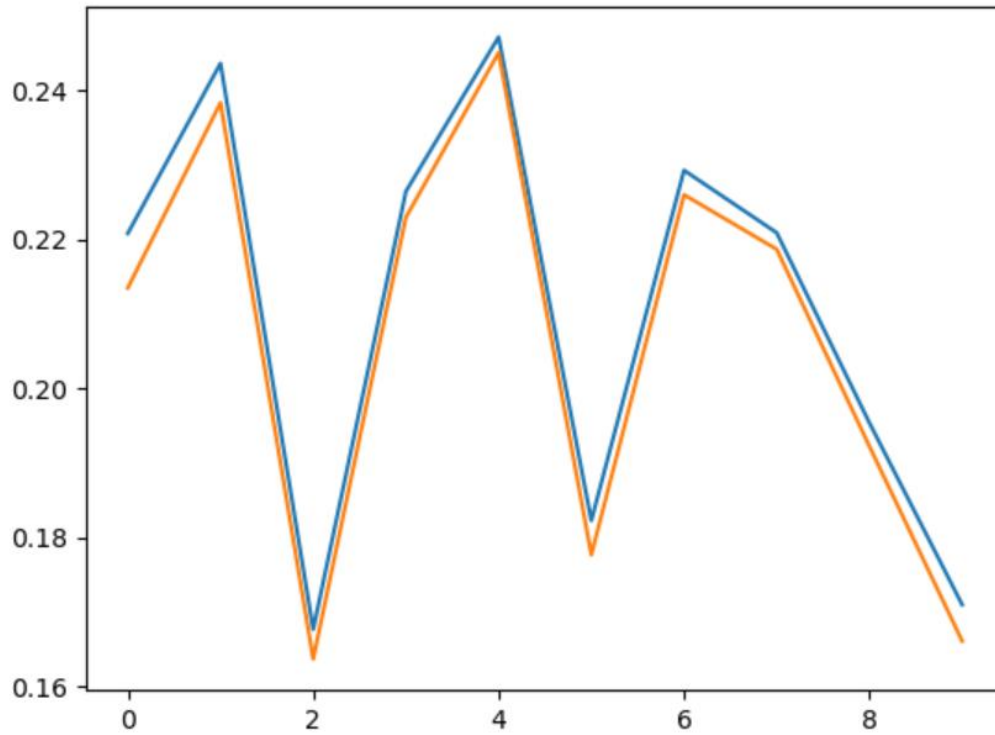
1. תחילה ניסיתי עבור ערכים התחלתיים נמוכים וראיתי כי הדיוק מאוד נמוך- 0.18 לכל היותר  
וזו לא המטרה\

```
HIDDEN = 50  
ETA = 0.1  
EPOCHS = 10
```

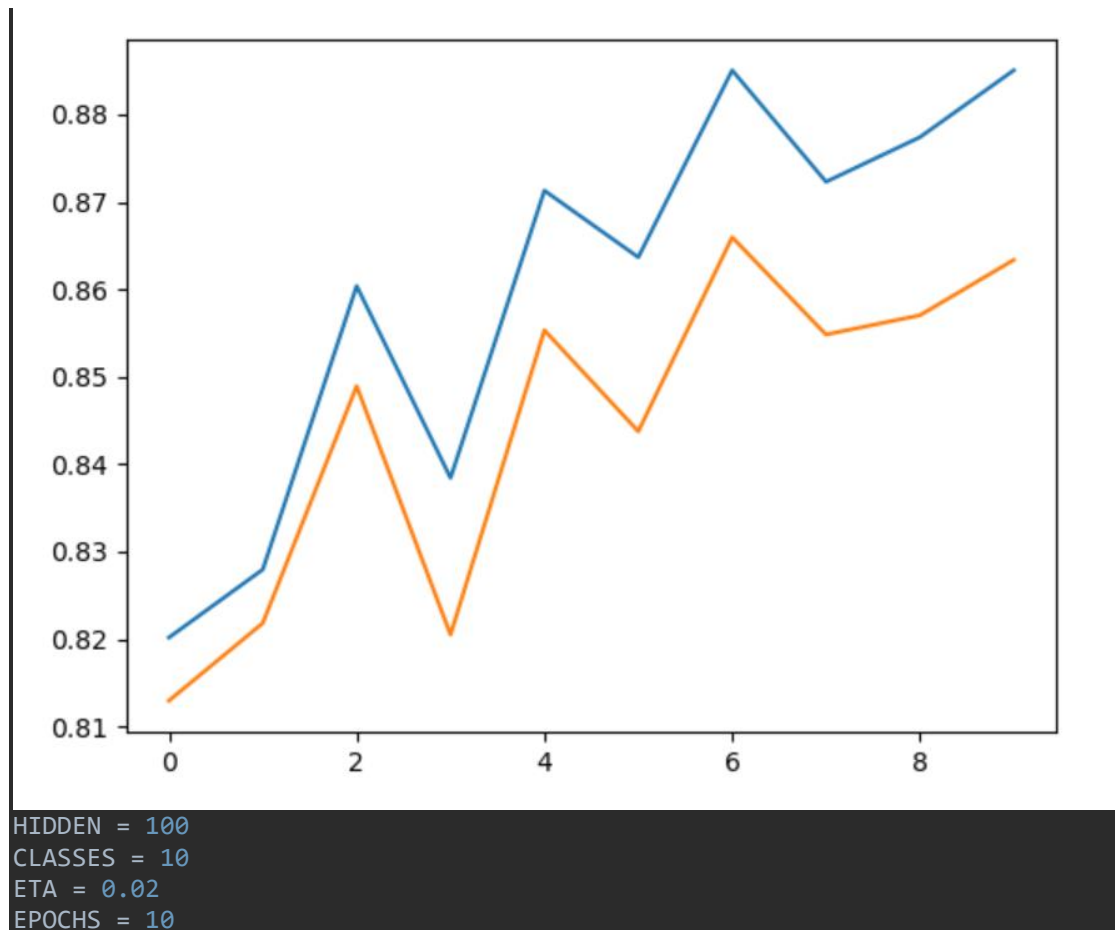


2. בניסיון השני הגדלתי את השכבה הנסתרת ונראה שאכן הדיוק עלה אך לא במידה מספקת. כיוון שגודל השכבה הנסתרת כבר גדול יחסית כנראה שלא קשור אליו ואנסה לשנות פרמטרים אחרים.

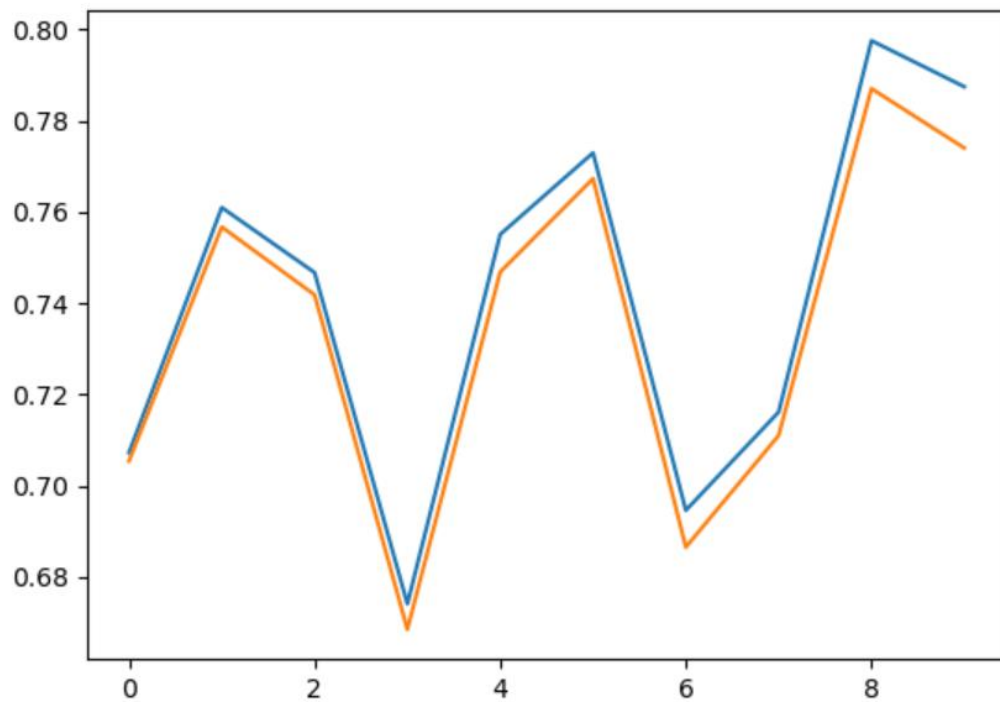
```
HIDDEN = 120  
CLASSES = 10  
ETA = 0.1  
EPOCHS = 10
```



3. כאשר הורדתי את הETA כלומר את learning rate נצפה שיפור משמעותי עם עלייה לדיוק של 0.88 ויחסים צמודים בין ה-dev and train כלומר שהלמידה נעשית בצורה טובה גם עבור דאטא שלא נלמד.

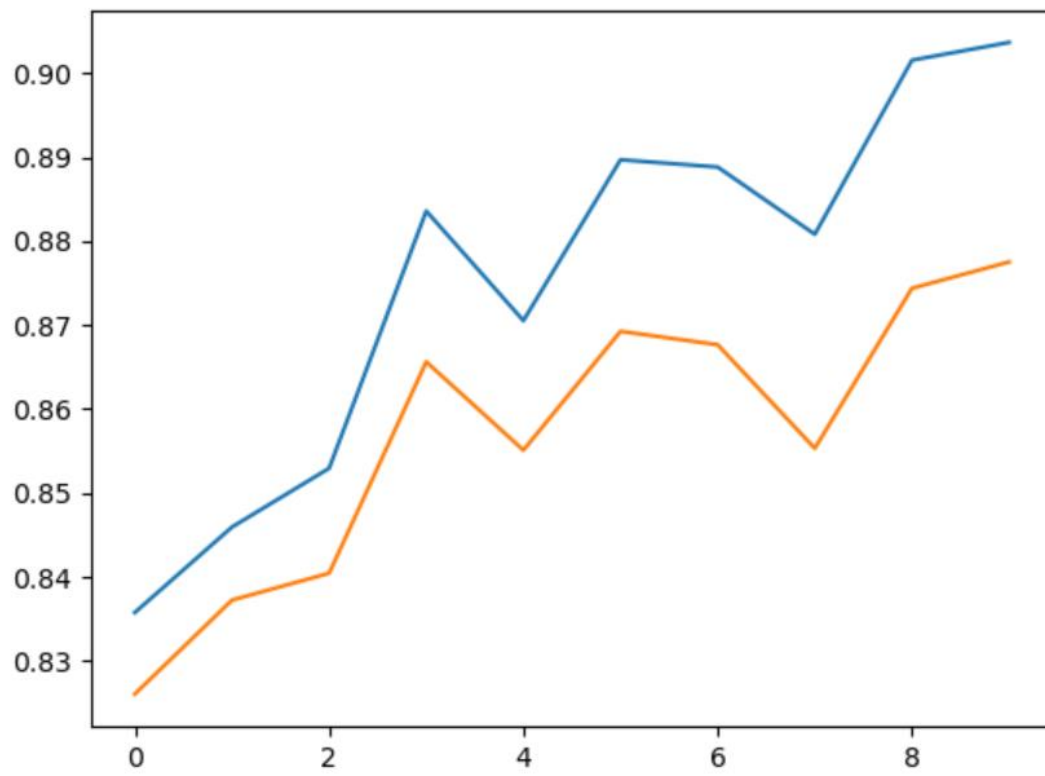


4. בניסיון ללמוד את השפעות learning rate שיניתי אותו לערך טיפה גדול יותר וראיתי ירידה בביצועים ולכן הבנתי כי 0.02 ומטה יהיו הערכים הטובים.



```
HIDDEN = 120  
CLASSES = 10  
ETA = 0.05  
EPOCHS = 10
```

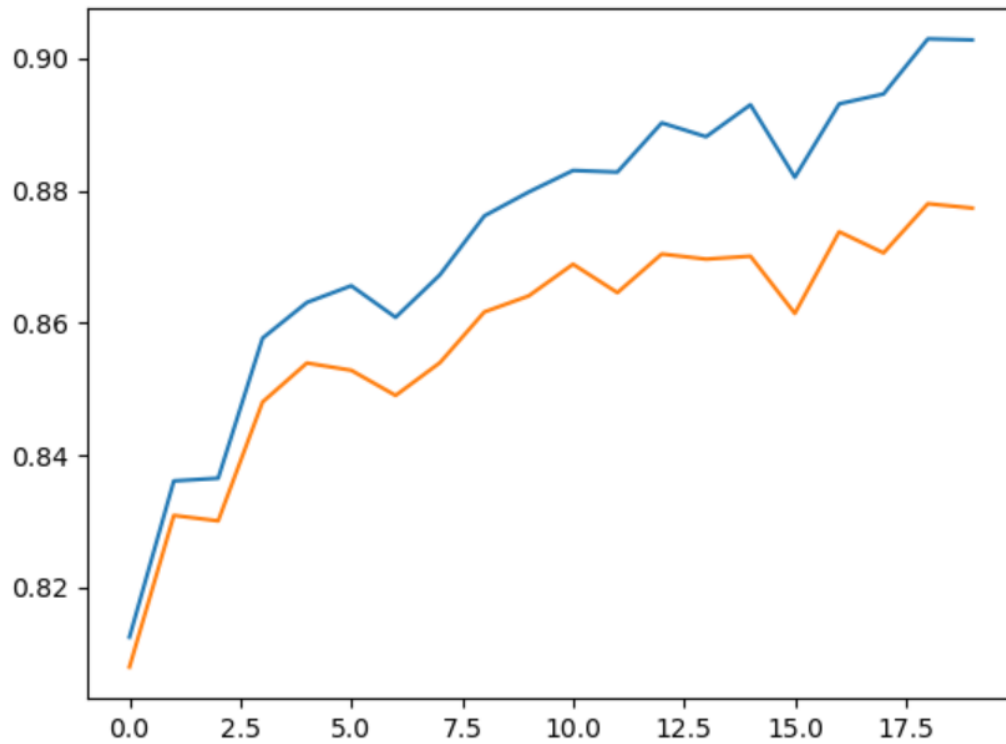
5. כאשר הקטנתי עוד את הETA נראו תוצאות טובות יותר מבחינת ה dev\_accuracy ולכן ניסיתי בהמשך להוריד אותו עוד יותר. בנוסף כדי לשפר את זמני הריצה ניסיתי להוריד את כמות השכבה הנסתרת בכל אופן.



```
HIDDEN = 120  
CLASSES = 10  
ETA = 0.01  
EPOCHS = 10
```

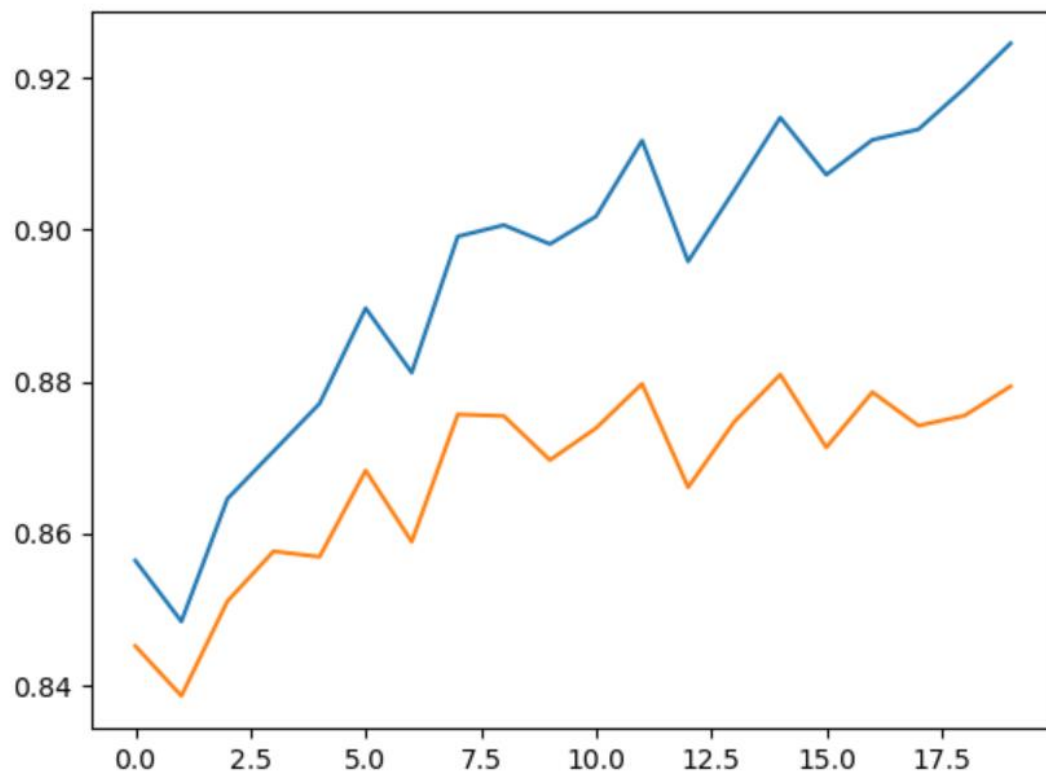
6. התוצאות הכי טובות עד כה נצפו כאשר שיניתי לנתונים הבאים כלומר learning rate נמוך  
הוא הגורם העיקרי שמשפיע בעוד שגודל השכבה הנסתרת פחות השפיעה, בנוסף, נראה כי  
מספר ה- EPOCHS השתנה כאן גם הוא והוא תרם לעלייה נוספת בדיוק לקראת הסוף ולכן  
בחרתי להמשיך עם מספר EPOCHS גדול יותר לעומת ה-10 שהיו עד כה.

```
HIDDEN = 80  
CLASSES = 10  
ETA = 0.001  
EPOCHS = 20
```

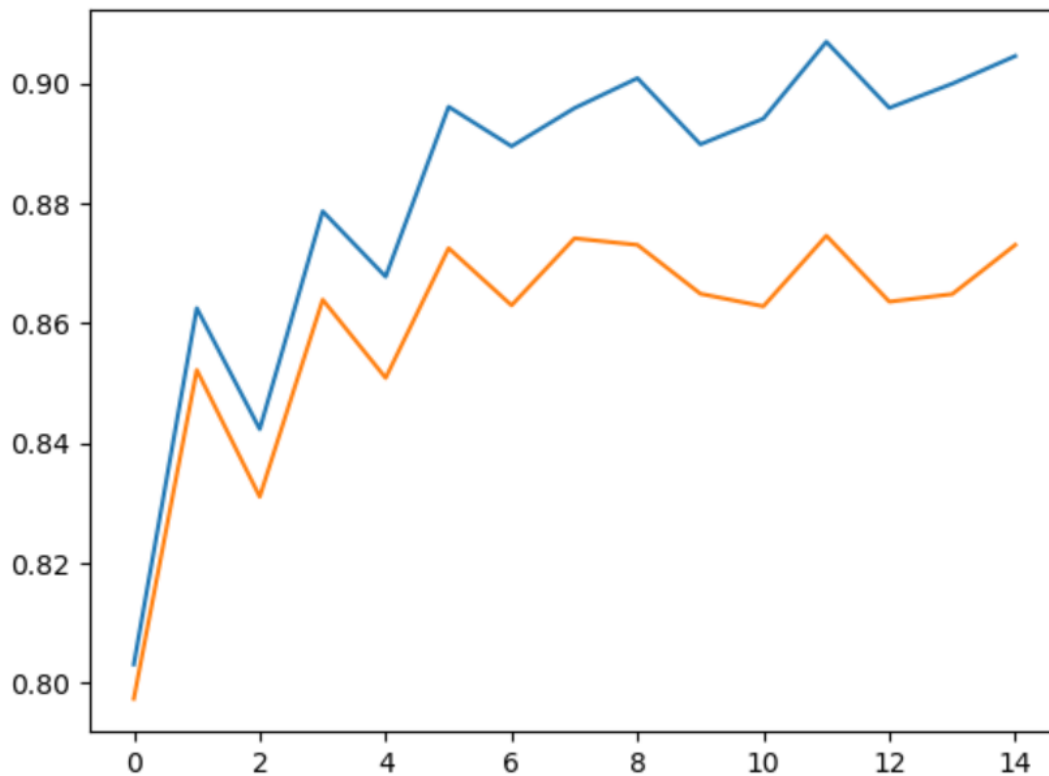


7. כאשר העלתי את ה learning rate נראה שכבר מתחיל לקראת ה-EPOCHS האחרונים overfitting על הדאטא כך שהדיוק של הtrain במגמת עליית בעוד ה dev מתחיל לרדת, פתרון אחרון שאציע לייתר ביטחון הוא להוריד את מספר ה-EPOCHS על אותם ערכים.

```
HIDDEN = 80  
CLASSES = 10  
ETA = 0.01  
EPOCHS = 20
```



8. בהורדת מספר ה-EPOCHS ל-15 נראה שיפור ומניעה של overfitting שכן שני ערכי הדיוק בעליה.



בהתחשב בכל הניסיונות שהצגתי לעיל (ובניסיונות נוספים שלא הצגתי שכן לא הראו שינוי) בחרתי להשתמש בערכים של דוגמה מספר 6 שכן שם נראית קורלציה מתאימה בין הדיוק על סט האימון מול סט הולידציה ועבור שניהם ערך הדיוק שהתקבל היה מירבי.



