

דו"ח הגשה – תרגיל 4, למידת מכונה

בדו"ח זה נציג עבור כל אחד מ-6 המודלים שלנו (מודל A – מודל F) את המרכיבים הבאים:

- גרף עבור ה-loss הממוצע בכל epoch, עבור ה-training set וה-validation set.
- גרף עבור ה-accuracy הממוצע בכל epoch, עבור ה-training set וה-validation set.
- אחוז הדיוק המתקבל עבור ה-test set.
- הסבר אודות ההיפר-פרמטרים השונים בקוד שלנו.

הערה חשובה: בחלק מהגרפים קיבלנו עליה קטנה שמלווה בירידה קטנה או להפך. הסיבה לכך – זה קרה במודלים שאחוז הדיוק ההתחלתי שלהם (ב-epoch הראשון) היה גבוה במיוחד, קרוב ל"נקודת הרוויה". כמו כן, אם אחוז הדיוק ההתחלתי של המודל גבוה, אז באופן טבעי ה-Loss יהיה נמוך כבר ב-epoch-ים הראשונים, ולכן ראינו מגמה דומה גם בחלק מהגרפים של ה-Loss.

בנוסף לכך, בתרגיל זה הונחנו להשתמש במספר קבוע מראש של epoch-ים (10). לכן, לא נפרט על ההיפר-פרמטר הזה תחת חלק ד' של המודלים השונים.

הערה לגבי סעיף ג' – אחוז הדיוק על גבי ה-test set: כמובן שאת סעיף זה הרצנו על גבי ה-test set מתוך ספריית ה-pyTorch, כדי להשוות את הפרדיקציות שקיבלנו אל מול הלייבלים המתאימים.

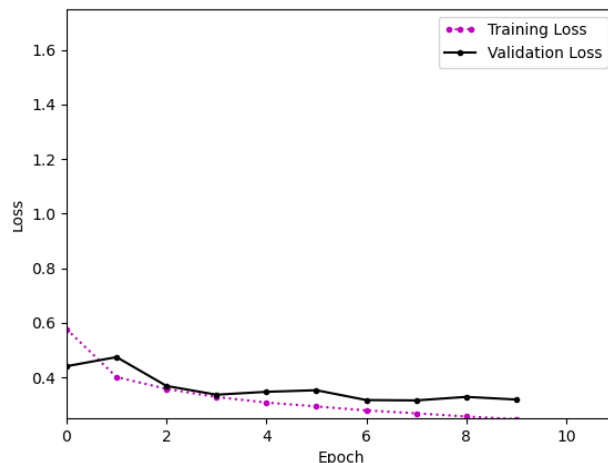
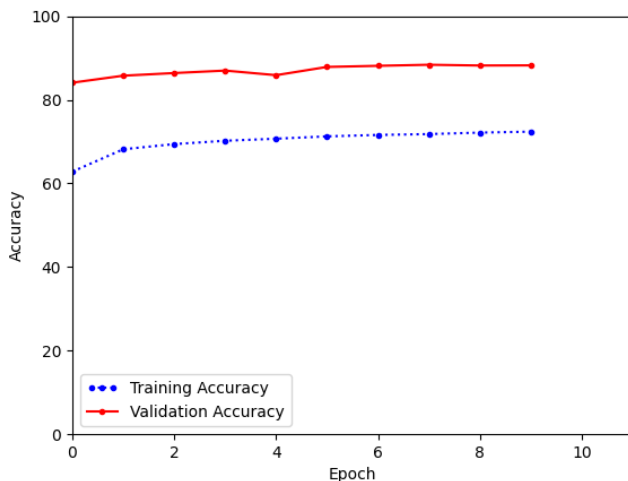
כמובן שאחוז הדיוק השתנה בחלק מהמודלים בין ההרצות השונות, בסדר גודל של אחוז.

בחרנו בקוד שלנו להשאיר את **מודל B כמודל העיקרי**, למרות שהיו מספר מודלים נוספים שהראו אחוזים יפים של דיוק – למשל מודל A ומודל E.

מודל A:

זוהי רשת נוירונים המורכבת מ-2 שכבות פנימיות. השכבה הראשונה בגודל 100 נוירונים והשכבה השנייה בגודל 50. שכבת הפלט היא בגודל 10 נוירונים (ככמות הלייבלים). פונקציית האקטיבציה שביצענו על השכבות הפנימיות היא ReLu, שיש לה מימוש ב-pyTorch. על שכבת הפלט הפעלנו את פונקציית ה-log_softmax. האימון על רשת זו התבצע בעזרת האופטימיזר SGD.

- גרף עבור ה-loss הממוצע בכל epoch, עבור ה-training set וה-validation set: **מימין**.
- גרף עבור ה-accuracy הממוצע בכל epoch, עבור ה-training set וה-validation set: **משמאל**.



ג. אחוז הדיוק המתקבל עבור ה-test set:

```
Test Set: Average loss: 0.3661, Accuracy: 8749/10000 (87%)
```

ד. היפר-פרמטרים:

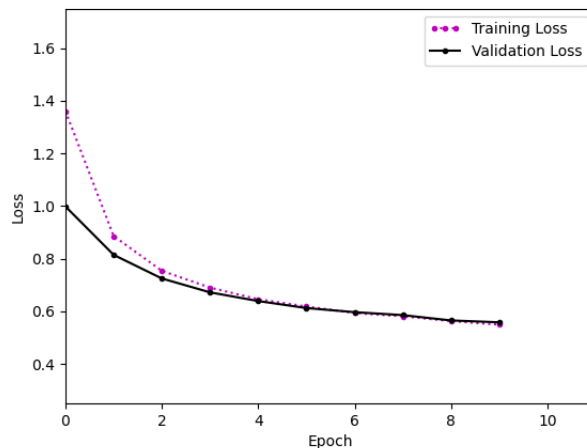
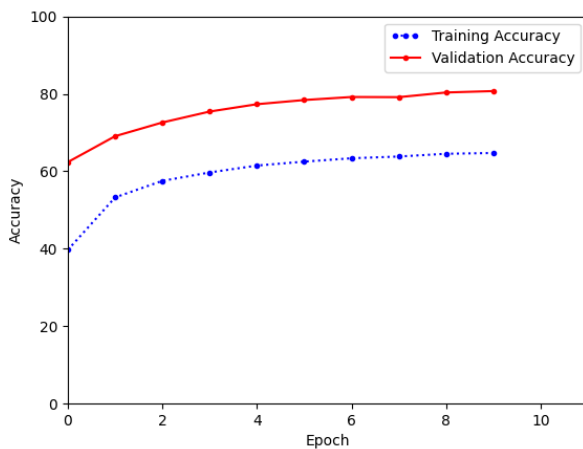
עבור ערך ה-learning rate בדקנו מספר ערכים – 0.01, 0.1, 0.0005, 0.0009, 0.0015, 0.001. לאחר מספר בדיקות, ראינו ש-0.1 מניב את אחוז הדיוק (accuracy) הגבוה ביותר ואת ה-Loss הנמוך ביותר, ולכן בחרנו בו.

מודל B:

זוהי רשת נוירונים המורכבת מ-2 שכבות פנימיות. השכבה הראשונה בגודל 100 נוירונים והשכבה השנייה בגודל 50. שכבת הפלט היא בגודל 10 נוירונים (ככמות הלייבלים). פונקציית האקטיבציה שביצענו על השכבות הפנימיות היא ReLu, שיש לה מימוש ב-pyTorch. על שכבת הפלט הפעלנו את פונקציית ה-log_softmax. האימון על רשת זו התבצע בעזרת האופטימיזר ADAM.

א. גרף עבור ה-loss הממוצע בכל epoch, עבור ה-training set וה-validation set: **מימין**.

ב. גרף עבור ה-accuracy הממוצע בכל epoch, עבור ה-training set וה-validation set: **משמאל**.



ג. אחוז הדיוק המתקבל עבור ה-test set:

```
Test Set: Average loss: 0.3433, Accuracy: 8849/10000 (88%)
```

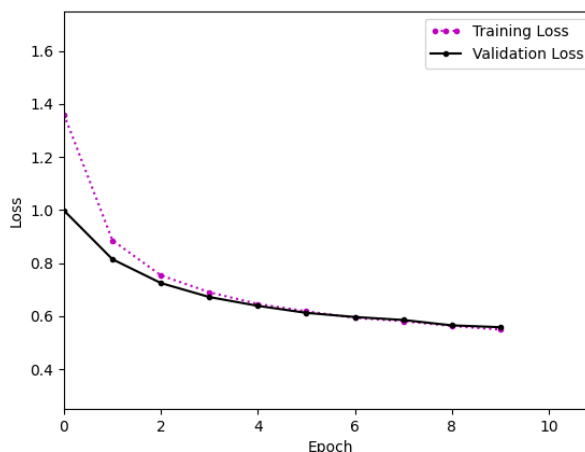
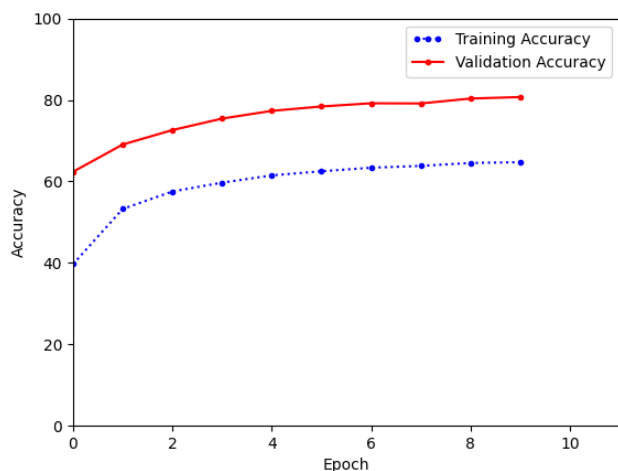
ד. היפר-פרמטרים:

עבור ערך ה-learning rate בדקנו מספר ערכים – 0.01, 0.1, 0.0005, 0.0009, 0.0015, 0.001. לאחר מספר בדיקות, ראינו ש-0.001 מניב את אחוז הדיוק (accuracy) הגבוה ביותר ואת ה-Loss הנמוך ביותר, ולכן בחרנו בו.

מודל C:

זוהי רשת נירונים המבוססת על מודל A עם שינוי יחיד – ביצענו Dropout על השכבות הפנימיות של רשת הנירונים. כל יתר הפרטים נותרו זהים.

- א. גרף עבור ה-loss הממוצע בכל epoch, עבור ה-training set וה-validation set: **מימין**.
ב. גרף עבור ה-accuracy הממוצע בכל epoch, עבור ה-training set וה-validation set: **משמאל**.



- ג. אחוז הדיוק המתקבל עבור ה-test set:

```
Test Set: Average loss: 0.5754, Accuracy: 8043/10000 (80%)
```

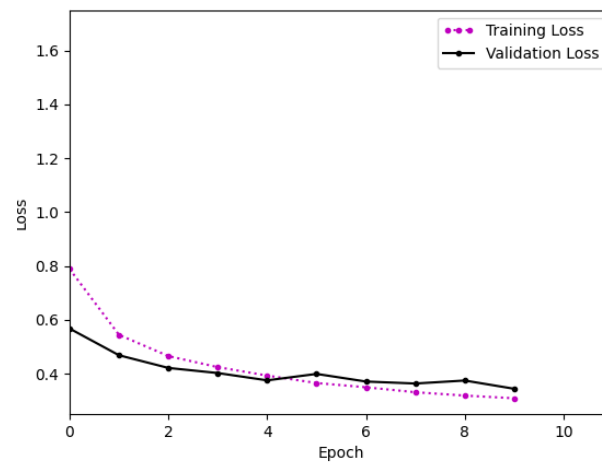
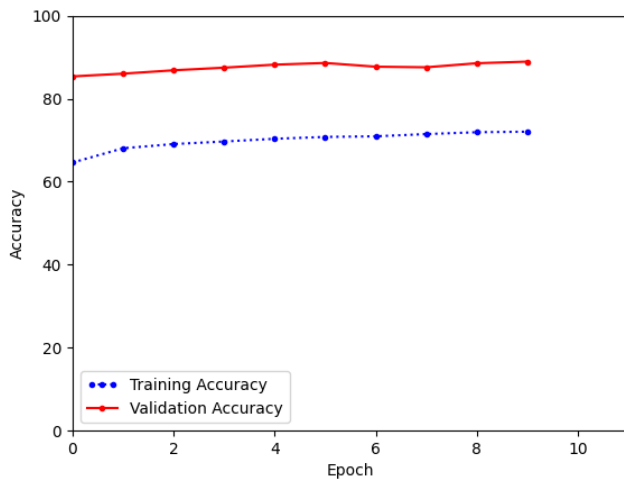
- ד. היפר-פרמטרים:
עבור ערך ה-learning rate בדקנו מספר ערכים – 0.01, 0.1, 0.0005, 0.0009, 0.0015, 0.001. לאחר מספר בדיקות, ראינו ש-**0.01** מניב את אחוז הדיוק (accuracy) הגבוה ביותר ואת ה-Loss הנמוך ביותר, ולכן בחרנו בו.

כאמור, במודל C הונחנו לבצע *Dropout*. פונקציה זו מקבלת כקלט ערך p שהוא ההסתברות לקחת את הנירון המסוים מתוך רשת הנירונים (או לא לקחת אותו). ראינו בהרצאה שערך ה- p המיטבי הוא בין 0.4 ל-0.6, והגענו למסקנה שכאשר $p=0.5$ אנחנו מקבלות את אחוזי הדיוק הטובים ביותר.

מודל D:

זוהי רשת נוירונים המבוססת על מודל A עם שינוי יחיד – ביצענו Batch Normalization לפני הפעלת פונקציות האקטיבציה. כל יתר הפרטים נותרו זהים.

- א. גרף עבור ה-loss הממוצע בכל epoch, עבור ה-training set וה-validation set: **מימין**.
ב. גרף עבור ה-accuracy הממוצע בכל epoch, עבור ה-training set וה-validation set: **משמאל**.



ג. אחוז הדיוק המתקבל עבור ה-test set:

```
Test Set: Average loss: 0.3687, Accuracy: 8737/10000 (87%)
```

ד. היפר-פרמטרים:

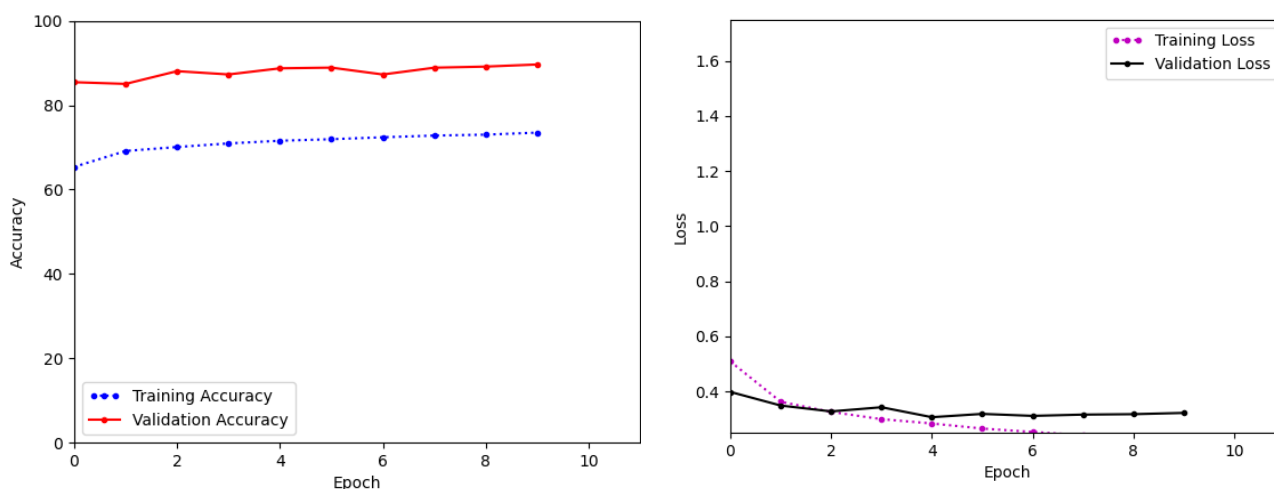
עבור ערך ה-learning rate בדקנו מספר ערכים – 0.001, 0.0015, 0.0009, 0.0005, 0.1, 0.01. לאחר מספר בדיקות, ראינו ש-**0.01** מניב את אחוז הדיוק (accuracy) הגבוה ביותר ואת ה-Loss הנמוך ביותר, ולכן בחרנו בו.

מודל E:

זוהי רשת נוירונים המורכבת מ-5 שכבות פנימיות, בגדלים – 128,64,10,10,10. שכבת הפלט היא בגודל 10 נוירונים (ככמות הלייבלים). פונקציית האקטיבציה שביצענו על השכבות הפנימיות היא ReLU, שיש לה מימוש ב-pyTorch. על שכבת הפלט הפעלנו את פונקציית ה-log_softmax.

במודל זה לא נאמר עם איזה אופטימיזר לאמן את רשת הנוירונים. לאחר שבדקנו מספר אופטימיזרים שונים, הגענו למסקנה שהאופטימלי הוא ADAM. לפיכך, האימון על רשת זו התבצע בעזרת האופטימיזר ADAM.

- א. גרף עבור ה-loss הממוצע בכל epoch, עבור ה-training set וה-validation set: **מימין**.
ב. גרף עבור ה-accuracy הממוצע בכל epoch, עבור ה-training set וה-validation set: **משמאל**.



ג. אחוז הדיוק המתקבל עבור ה-test set:

Test Set: Average loss: 0.3491, Accuracy: 8802/10000 (88%)

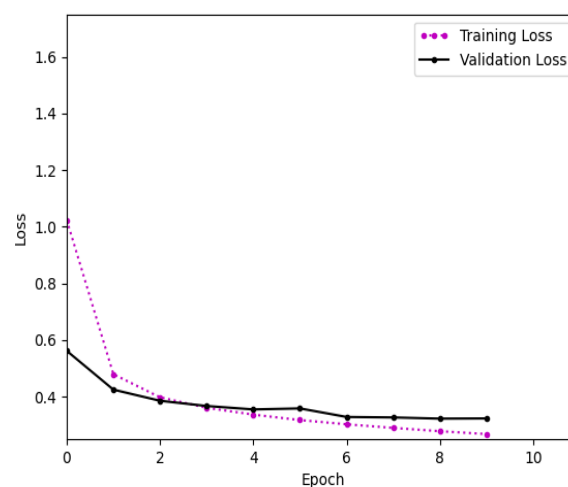
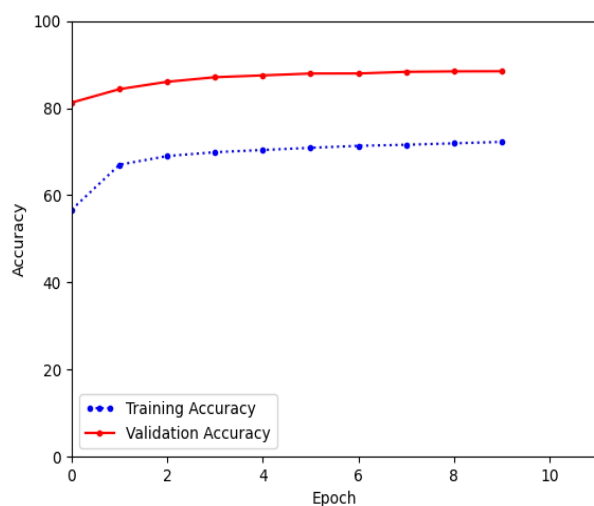
- ד. היפר-פרמטרים:
עבור ערך ה-learning rate בדקנו מספר ערכים – 0.001, 0.0015, 0.0009, 0.0005, 0.1, 0.01. לאחר מספר בדיקות, ראינו ש-0.0009 מניב את אחוז הדיוק (accuracy) הגבוה ביותר ואת ה-Loss הנמוך ביותר, ולכן בחרנו בו.

מודל F:

זוהי רשת נוירונים המורכבת מ-5 שכבות פנימיות, בגדלים – 128,64,10,10,10. שכבת הפלט היא בגודל 10 נוירונים (ככמות הלייבלים). פונקציית האקטיבציה שביצענו על השכבות הפנימיות היא Sigmoid, שיש לה מימוש ב-pyTorch. על שכבת הפלט הפעלנו את פונקציית ה-log_softmax.

במודל זה לא נאמר עם איזה אופטימיזר לאמן את רשת הנוירונים. לאחר שבדקנו מספר אופטימיזרים שונים, הגענו למסקנה שהאופטימלי הוא ADAM. לפיכך, האימון על רשת זו התבצע בעזרת האופטימיזר ADAM.

- א. גרף עבור ה-loss הממוצע בכל epoch, עבור ה-training set וה-validation set: **מימין**.
ב. גרף עבור ה-accuracy הממוצע בכל epoch, עבור ה-training set וה-validation set: **משמאל**.



- ג. אחוז הדיוק המתקבל עבור ה-test set:

```
Test Set: Average loss: 0.3582, Accuracy: 8728/10000 (87%)
```

- ד. היפר-פרמטרים:
עבור ערך ה-learning rate בדקנו מספר ערכים – 0.001, 0.0015, 0.0009, 0.0005, 0.1, 0.01. לאחר מספר בדיקות, ראינו ש-0.0005 מניב את אחוז הדיוק (accuracy) הגבוה ביותר ואת ה-Loss הנמוך ביותר, ולכן בחרנו בו.