עבודה 5 – מבוא למודלים גרפים ולמידה עמוקה

# שאלה 1:

הרשת הלינארית עם שכבות בגודל 512, 256 בשלב מסוים באימון מתחילה לרדת בדיוק על ה-validation set לכן היא overfitted, והגיע לדיוק של 0.75 על ה-validation set בסוף האימון. לעומתה הרשת הלינארית עם השכבות בגודל 64, 32 לא סובלת מ-overfitting והגיעה לדיוק של 0.81.

הרשת קובולוציה עם קרנל בגודל 10 גם עשתה overfitting אבל למרות זאת הגיעה לדיוק של 0.89 שזה טוב יותר משני הרשתות הלינאריות. לעומתה הרשת עם קרנל בגודל 1 הגיעה לדיוק של רק חצי על ה-validation set, כך שהיא גרועה יותר משאר הרשתות.

# שאלה 2:

זה לא מחייב שהמודל שנבחר על סמך ה- validation הוא יהיה המודל המוצלח ביותר על ה-test data, אנחנו מניחים שה- validation נבחר באופן רנדומלי בפילוג אחיד, כמו שה-test data נבחר, ובנוסף אנחנו לא מאפשרים למידה של המודל על ה- validation כדי לחכות את ההתנהגות של ה-test data.   
ולכן המודל שמתפקד הכי טוב על מידע שהוא validation יכול לתת לנו אינדיקציה טובה לגבי יעילות המודל.

# שאלה 3:

בחרנו ב-Finetuning CNN מהסיבה ש-Resnet18 מאומנת על תמונות ריאליסטיות, ולא על תמונות שמייצגות סמלים או אותיות (המידע שאנחנו רוצים שהיא תלמד לסווג זה מספרים), ולכן אנחנו ללמד את הרשת בצורה יותר עמוקה, כלומר לשנות יותר משכבה אחת אחורה.

# שאלה 4:

הביצועים ב- Finetuning CNNעלו בצורה משמעותית, וזאת מכיוון שבמקרה זה אנחנו לא מתחילים מ"אפס", וכבר בהתחלה יש לנו בסיס טוב ואנחנו רק צריכים ללמד את הרשת להתאים את עצמה לצרכים שלנו. למדנו בכיתה שיש הרבה דמיון בין השכבות הראשונות ברשתות שמאומנת על תמונות, ולכן במקרים מסוימים יש יתרון לשימוש ברשתות שמאומנות מראש.  
ועוד סיבה היא שהרשת Resnet18 פשוט מורכבת בצורה הרבה יותר חכמה.

# שאלה 5:

רואים שקצב למידה גבוה לא יעיל בכלל, יוצר קפיצות גדולות בדיוק המודל ולא וודאית לעומת זאת עם קצב למידה נמוך אפשר לבזבז הרבה זמן, כלומר צריך למוצא את נקודת האיזון בין קצב גבוה לנמוך.