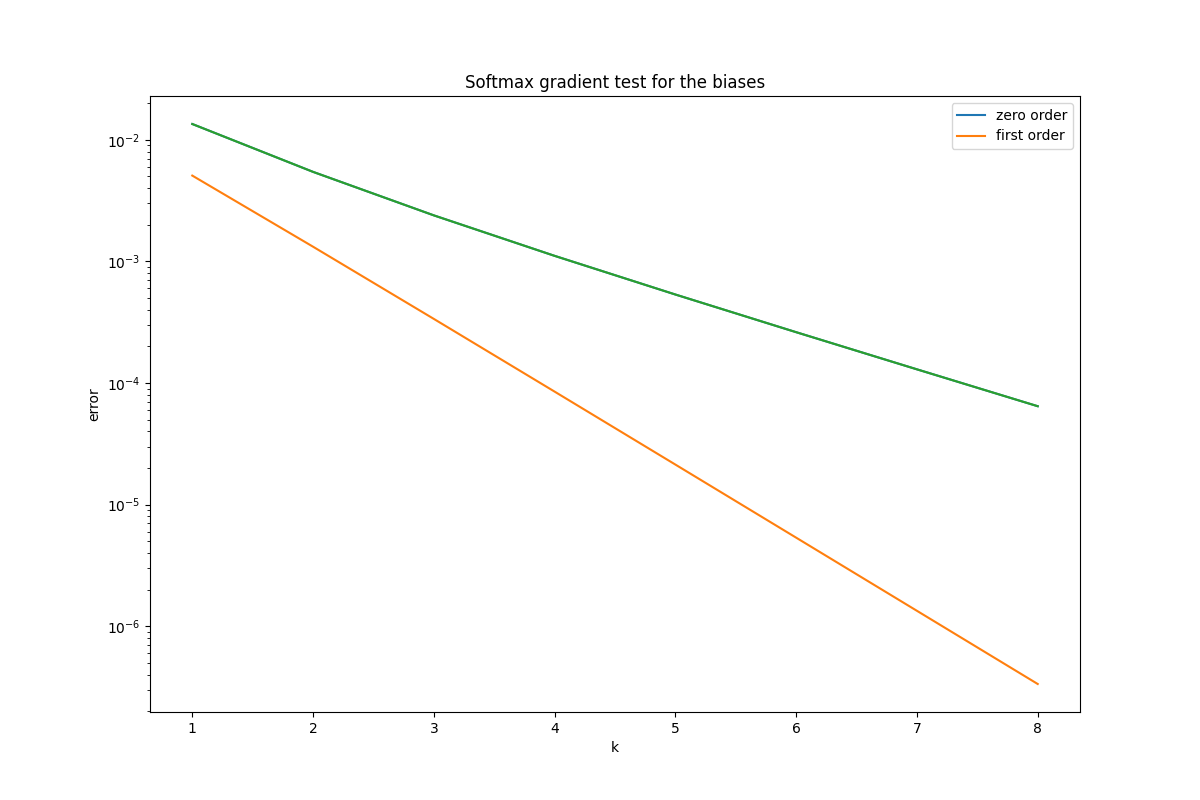
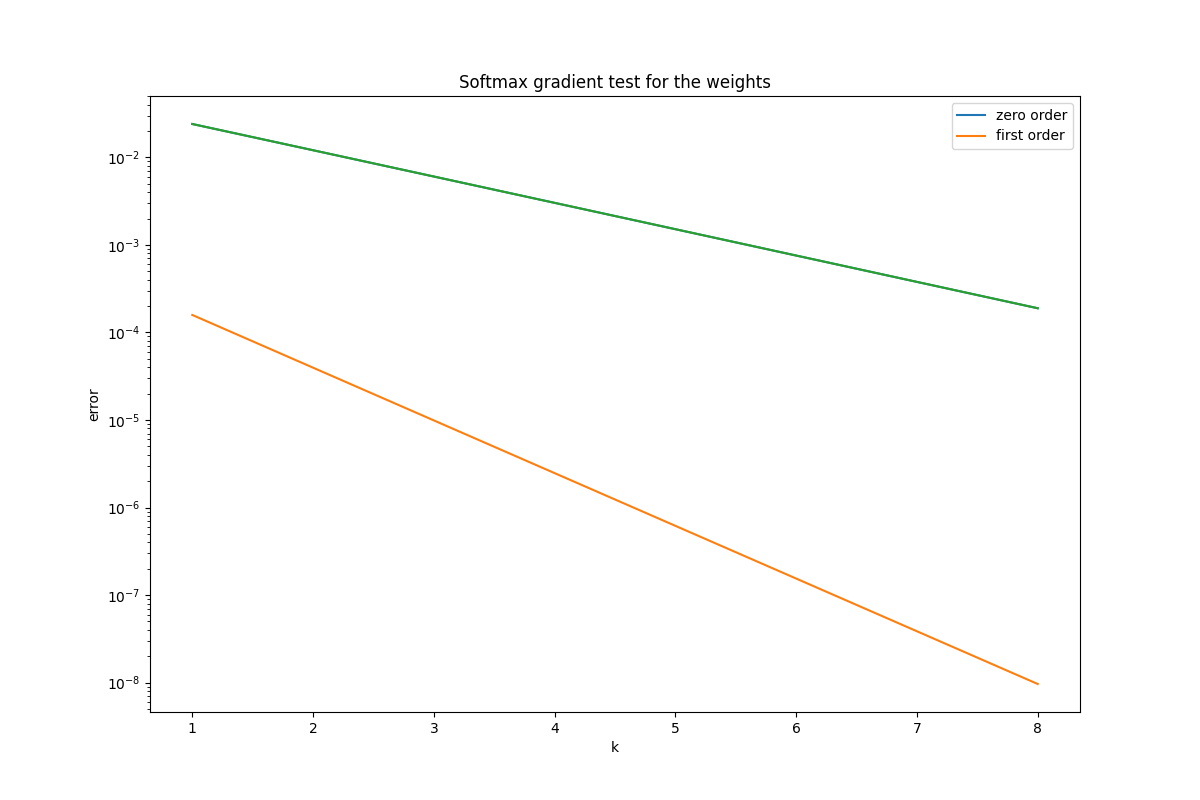
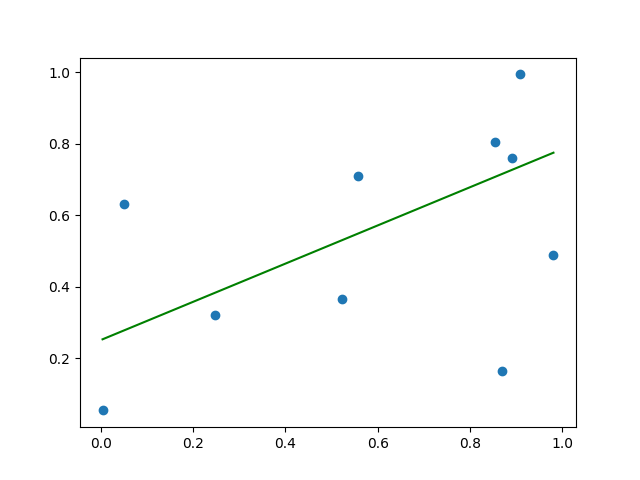
למידה עמוקה עבודה 1

**2.1**

1) הטסטים של הגרדיאנט ל-softmax regression לפי w ו-b :



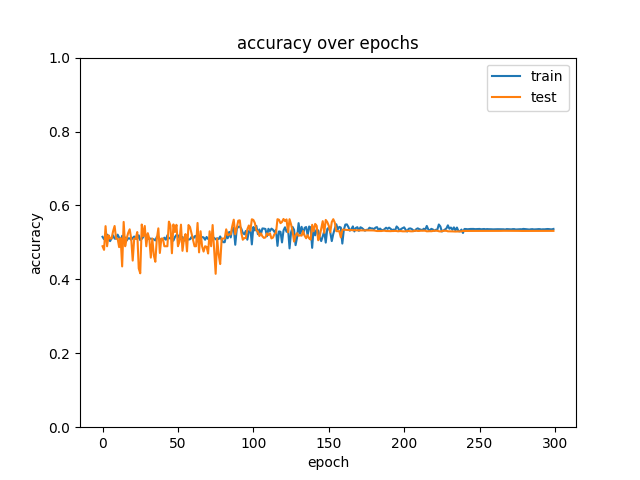
2) לקחנו 10 data-points רנדומליים והרצנו SGD על obejective function של least squares . הרצנו מאה אלף איטרציות עם learning rate של 0.1 . זאת התוצאה שקיבלנו:



3)

הגענו למסקנה שהמודל רק עם ה softmax לא מצליח ללמוד טוב, ללא תלות ב-lr וב-batch size

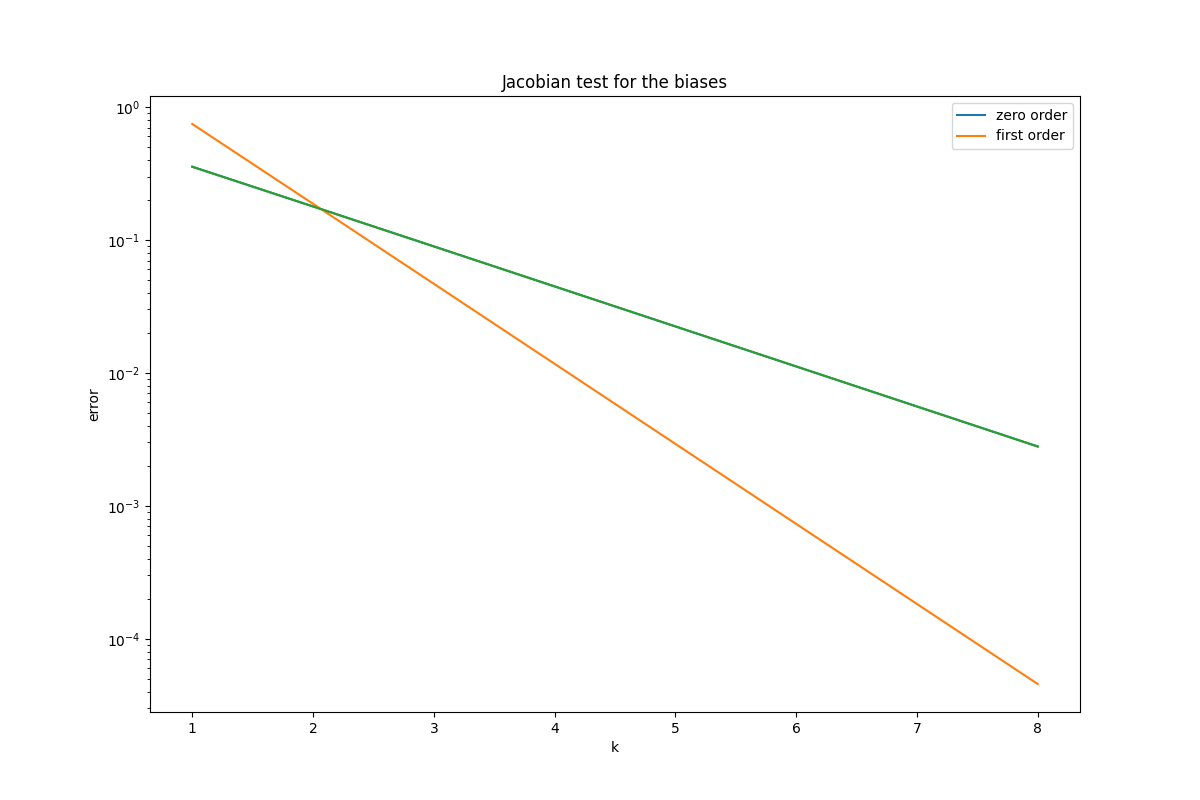
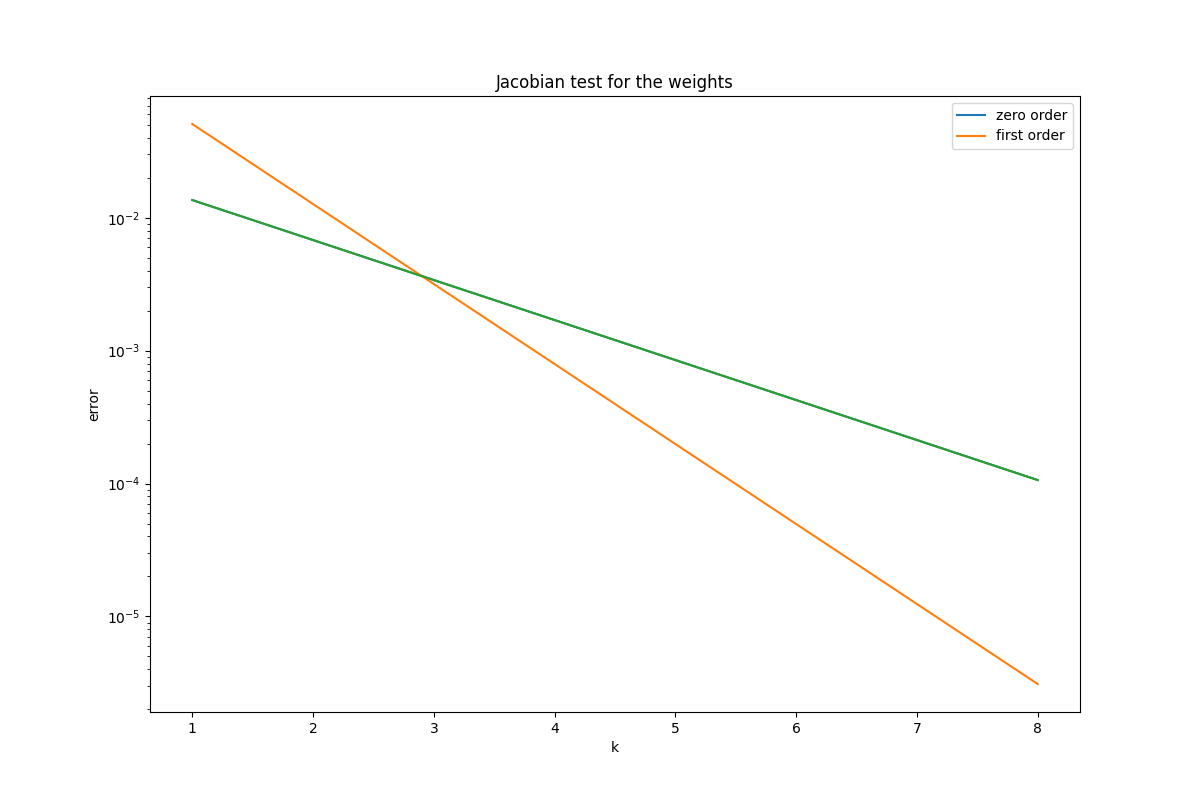
התוצאות שקיבלנו:



**2.2**

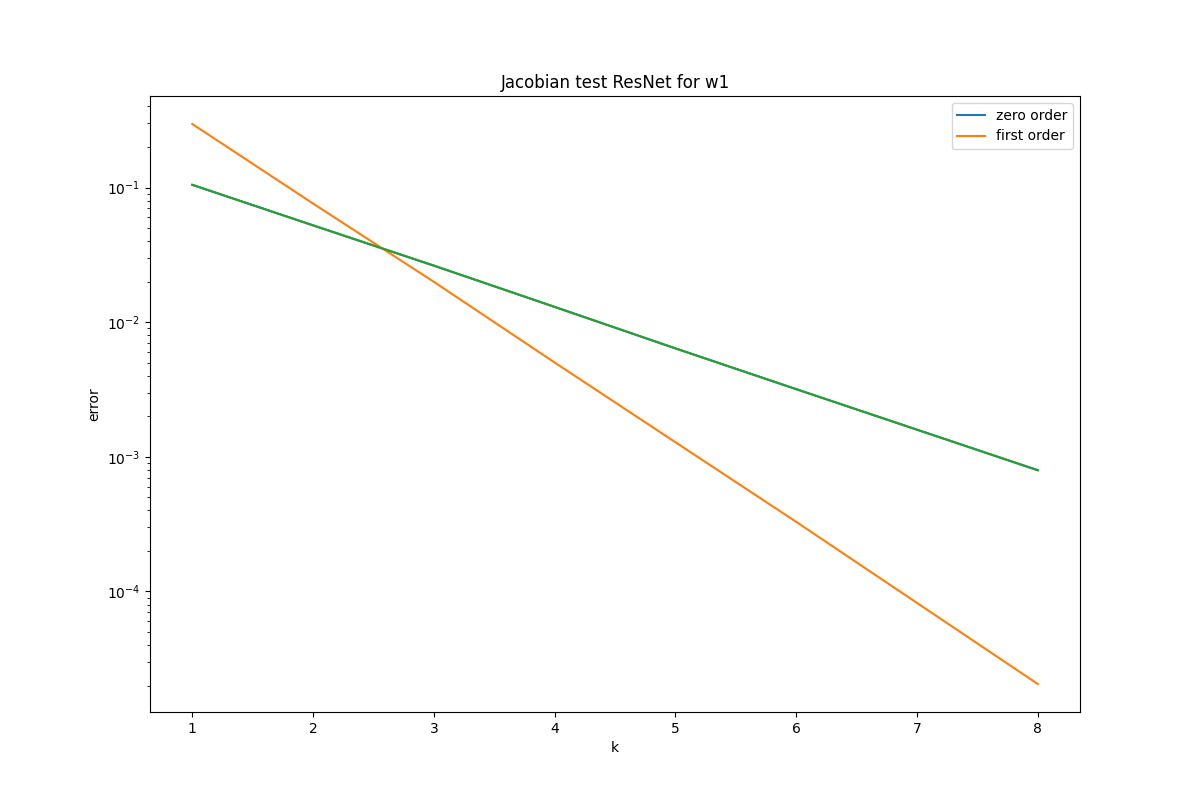
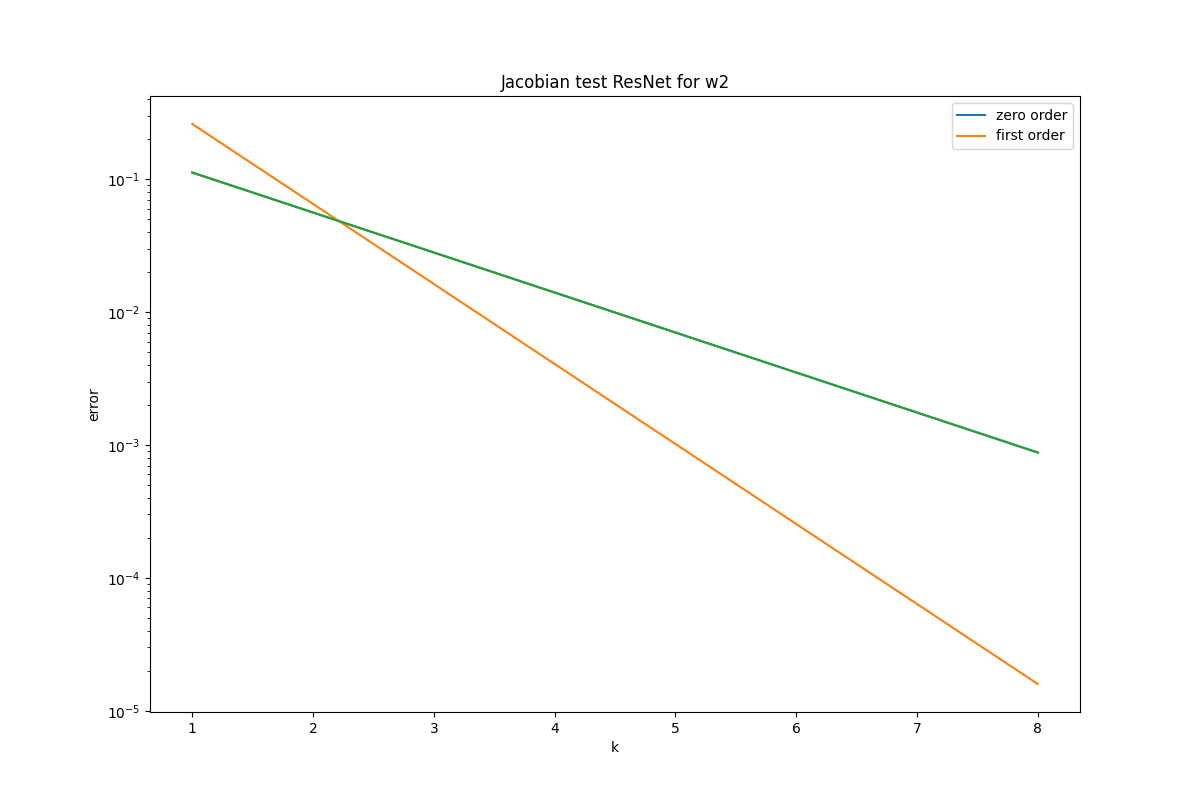
1)

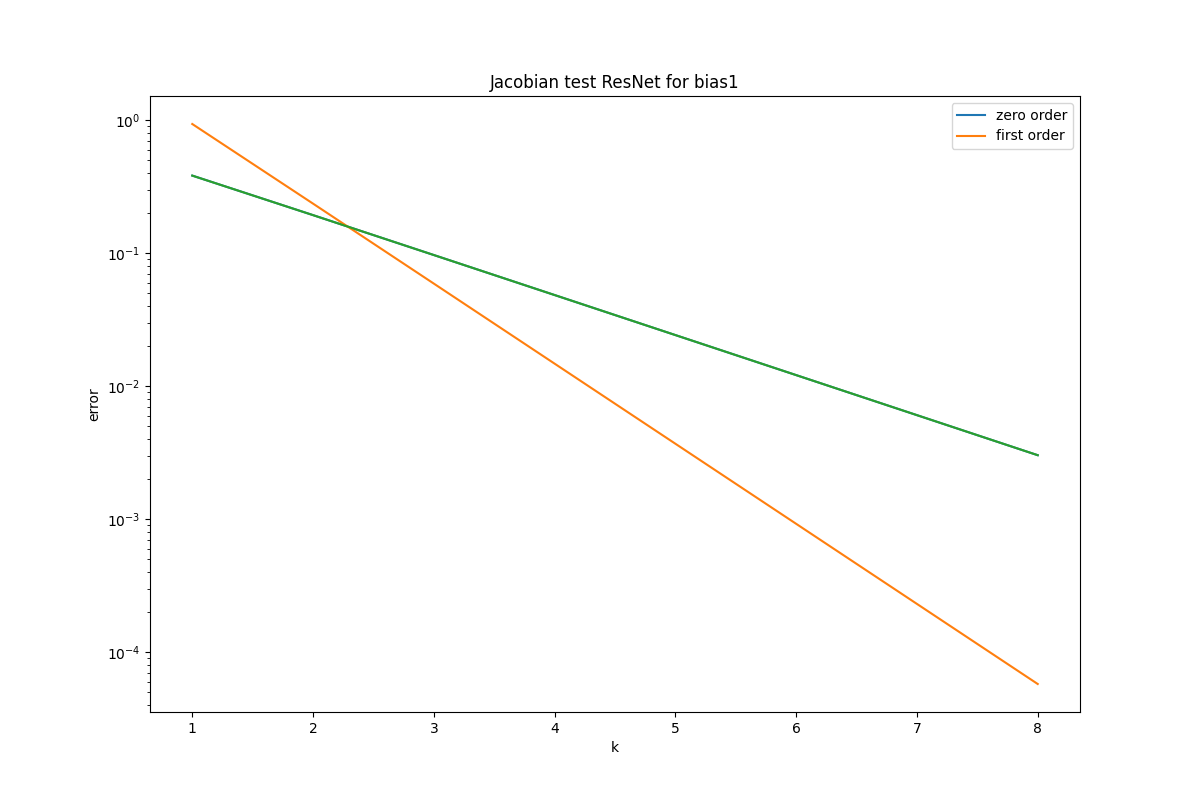
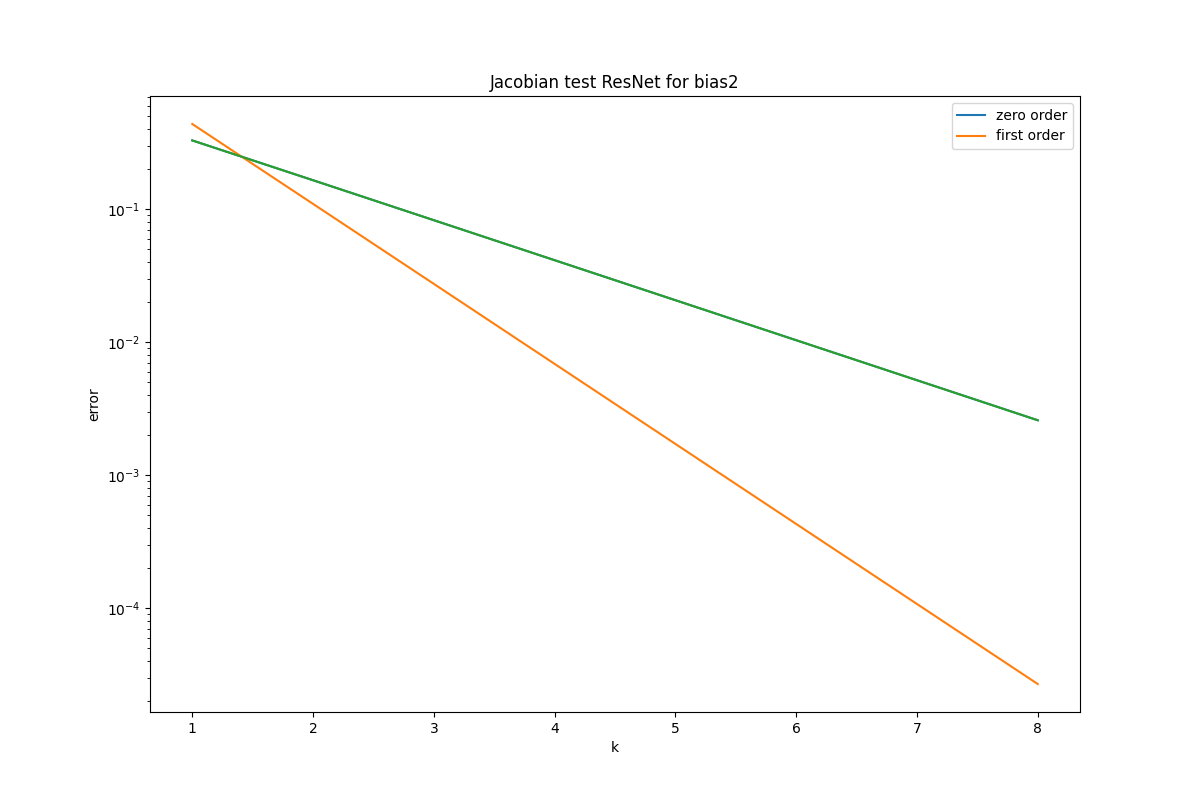
נצרף את ה-Jacobian tests של הרשת נוירונים שבנינו, לפי w ו-b :



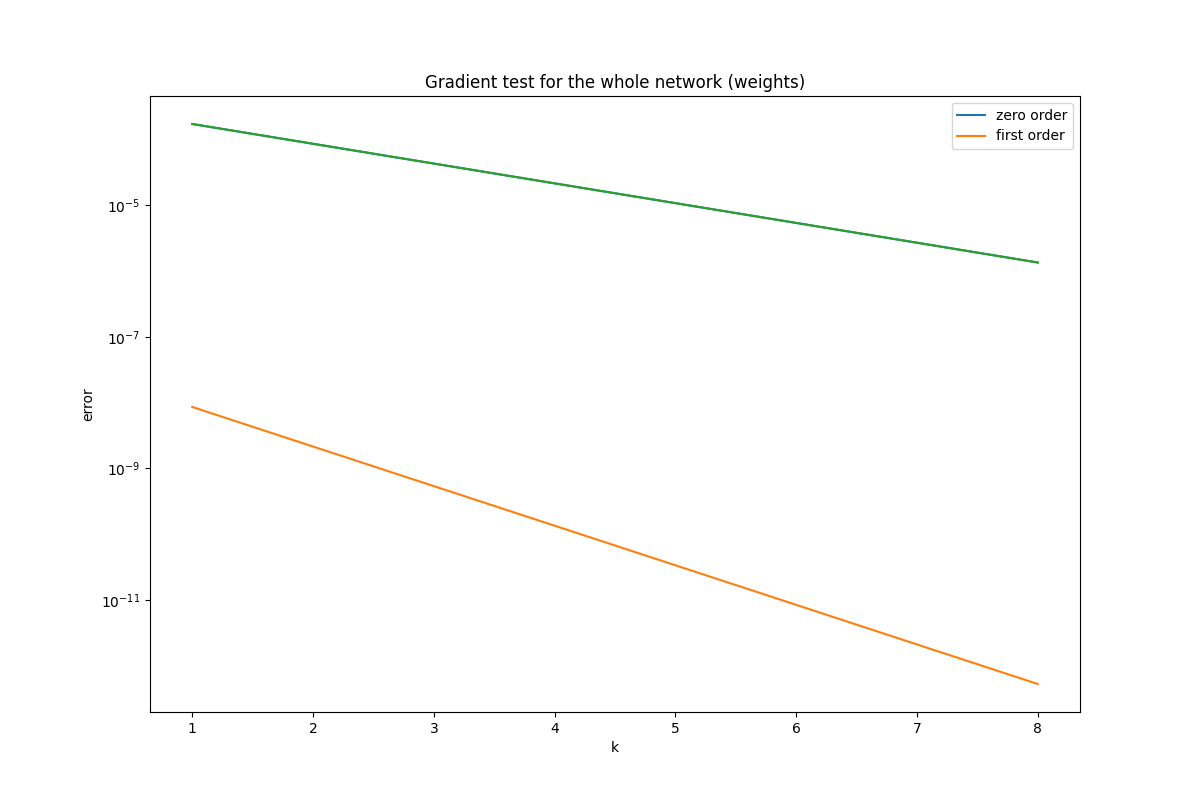
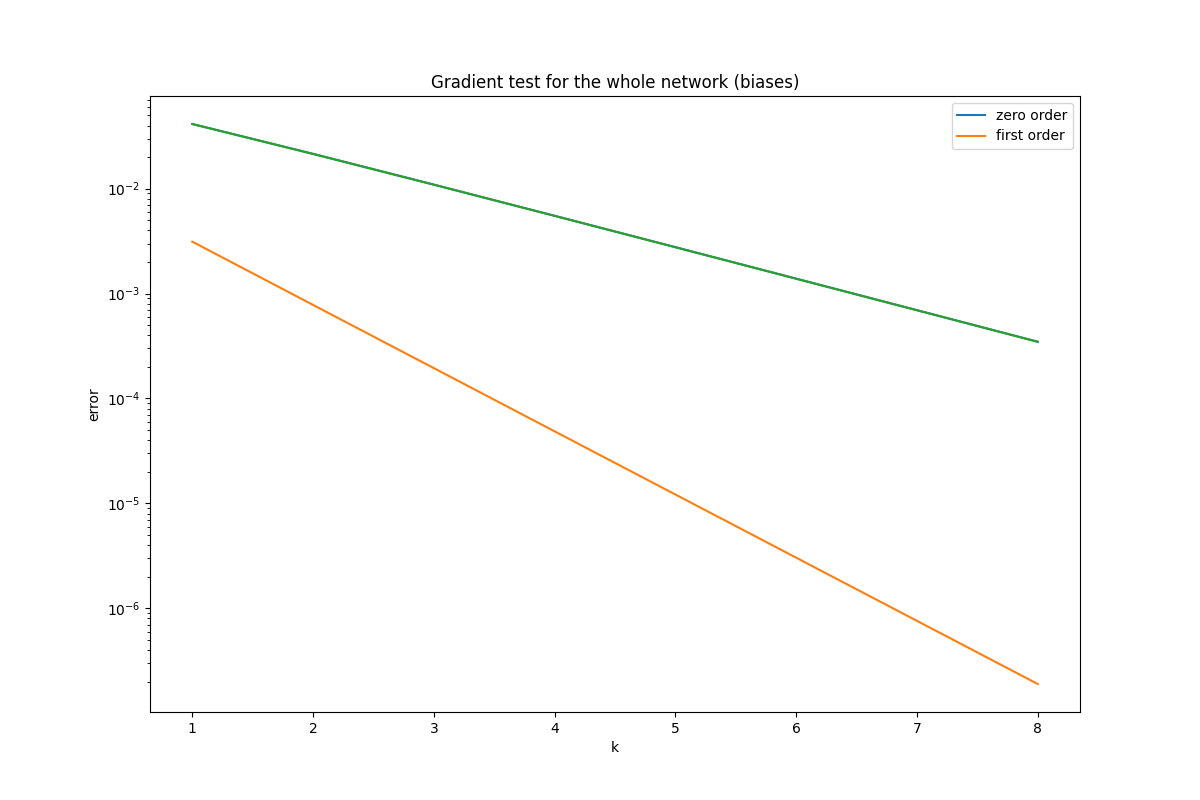
2)

נצרף את ה-Jacobian tests של הרשת נוירונים עם Residual block שבנינו, לפי w1,b2 ו-b1,b2 :

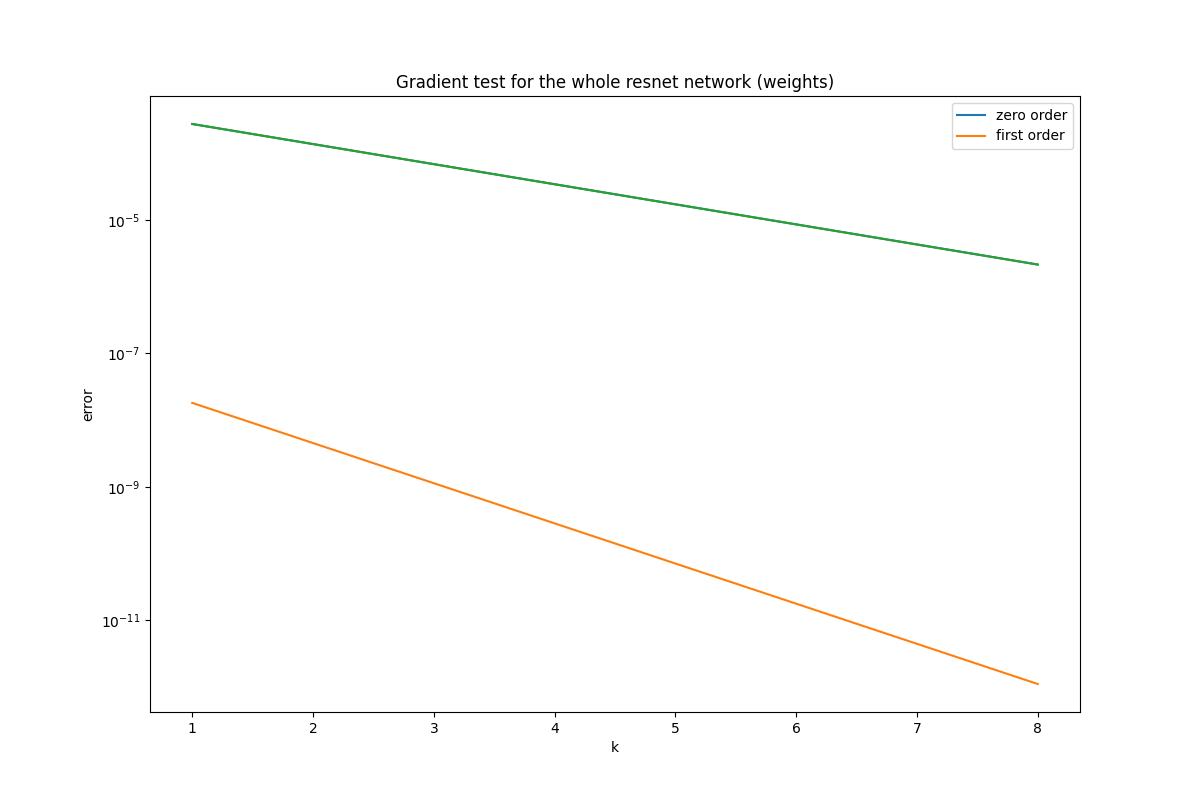
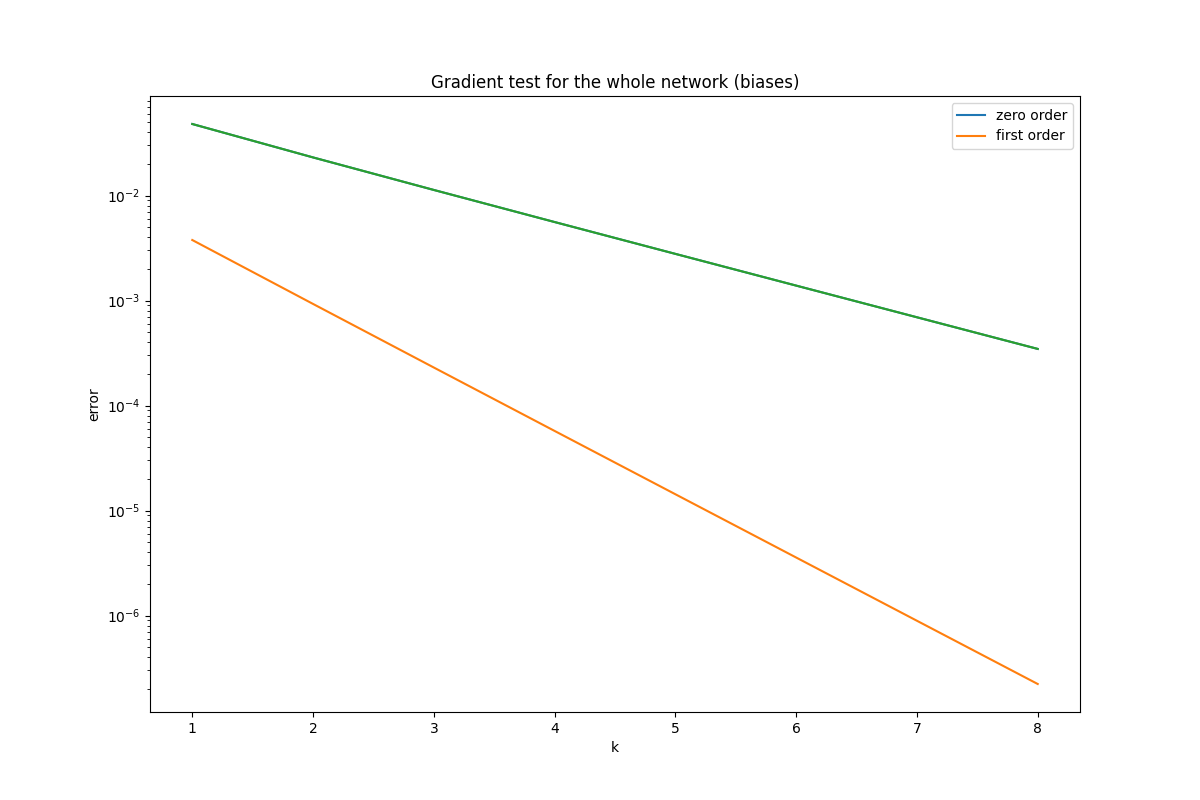
 

3) מצורף jacobian test עבור רשת נוירונים בסיסית:

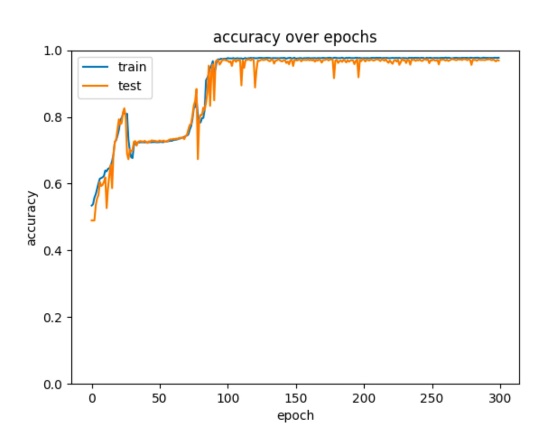
**** ****

עבור רשת נוירונים עם Residual blocks :

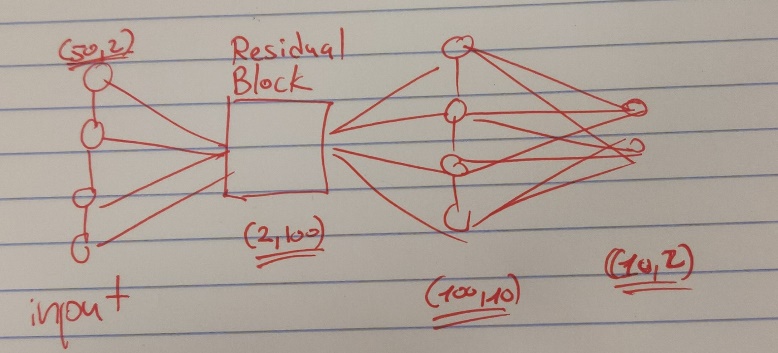
4)

אחרי משחק רב עם ההיפר-פרמטרים התוצאה הטובה ביותר שהגענו אליה היא 97% דיוק



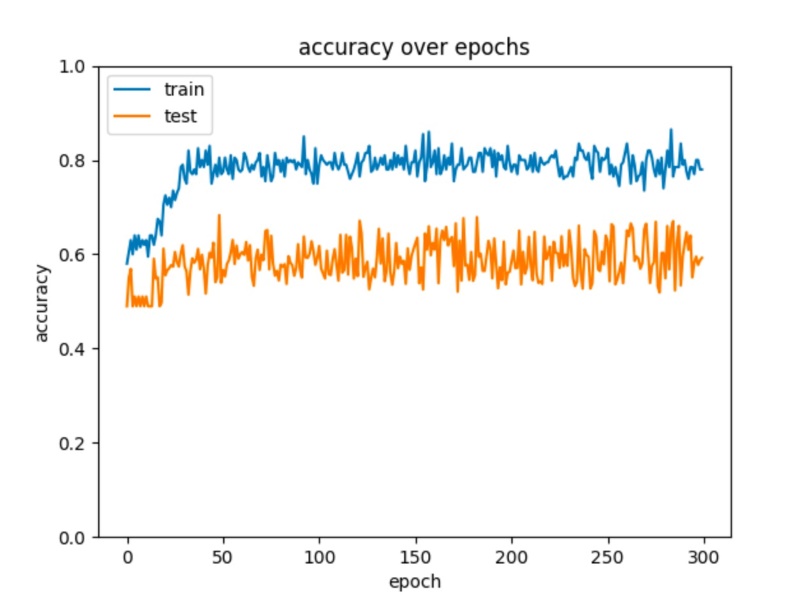
היפר פרמטרים: batch size = 100, lr=1, epochs = 300

המודל מורכב מresidual block ועוד hidden layer רגיל:



5)

כשהמודל מתאמן רק על 200 דוגמאות, הוא לא לומד מספיק טוב ונראה שהוא נכנס לover-fitting



ההיפר-פרמטרים הם בדיוק כמו בסעיף הקודם. (גם ששיחקנו איתם זה לא עזר יותר מדי)