

חישוביות וקוגניציה – תרגיל 4

נועה בן דרור

316163260

שאלה 1

1. במקרה כזה, בו כל המשקולות מאותחלות ל-0, ופונקציית האקטיבציה מקיימת $g(0)=0$, הרשת לא תצליח ללמוד דבר. הקלט לכל שכבה הוא כפל של מטריצת הקשרים המתאימה, בשכבה שלפניה. בגלל שכל מטריצות הקשרים הן מטריצות אפסים, כפל כזה יביא לנו לכל שכבה קלט של אפסים, ולאחר אקטיבציה $g(0)=0$, נישאר עם אפסים (חוץ משכבת הקלט). ואז, לפי נוסחת עדכון המשקולות יתקיים:

$$\Delta W^l = -\eta \cdot \underbrace{\delta^l \cdot (s^{l-1})^T}_{\nabla_{W^l} \mathcal{E}}$$

נשים לב ש-SI-1 הוא וקטור האפס לפי מה שהסברנו הרגע, ולכן המשקולות יישארו תמיד מטריצות של אפסים. לכן, לא תתבצע למידה ברשת.

2. נניח שקבענו את הארכיטקטורה וברצוננו ללמוד את המשקולות עצמן. במקרה כזה, של פרספטרונים בינאריים, פונקציית האקטיבציה שנפעיל היא פונקציית sign. באלגוריתם BackPropagation בשלב ה backward pass משתמשים בנגזרת של פונקציית האקטיבציה כדי לחשב את סיגנלי השגיאה ואת עדכון המשקולות. אבל הנגזרת של פונקציית ה sign היא 0 לכל נקודה שונה מ-0, ואינה מוגדרת בנקודה 0. לכן, אם נפעל לפי האלגוריתם:

$$1. \text{ סיגנל השגיאה בשכבת הפלט } \delta^M = \nabla_{s^M} \mathcal{E} \odot g'(h^M)$$

$$2. \text{ עבור } l = M-1, M-2, \dots, 1:$$

$$\delta^l = ((W^{l+1})^T \delta^{l+1}) \odot g'(h^l)$$

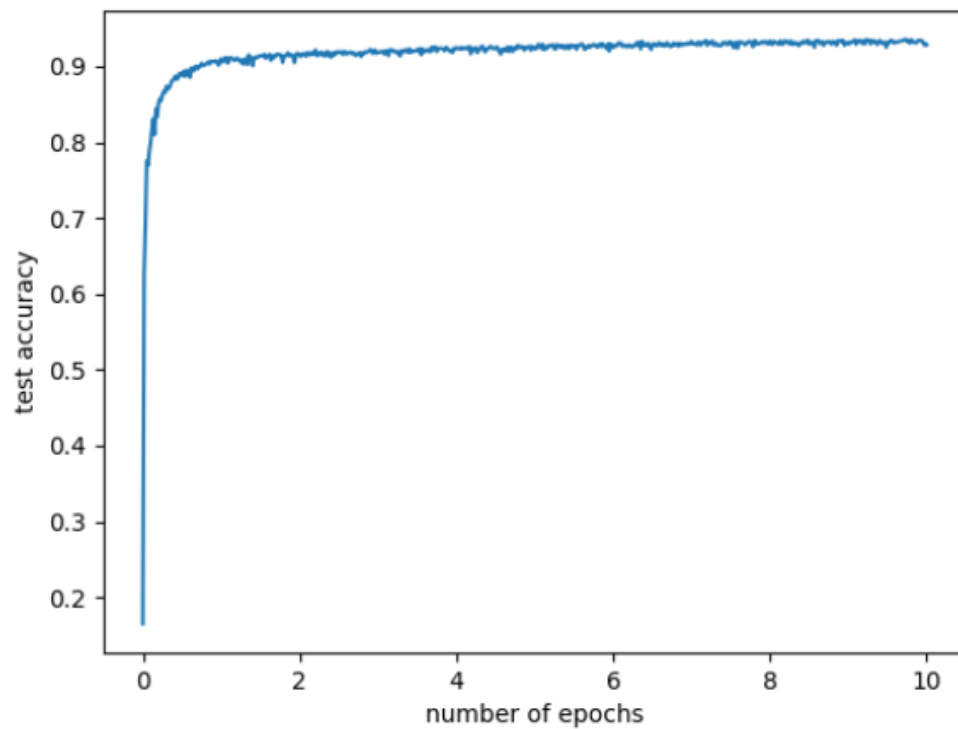
3. עדכון המשקולות בכיוון הנגדי לגרדיאנט:

$$\Delta W^l = -\eta \cdot \underbrace{\delta^l \cdot (s^{l-1})^T}_{\nabla_{W^l} \mathcal{E}}$$

נשים לב שנקבל שכל סיגנל שגיאה הוא 0 (אם h^l שונה מ-0, ואז $g'(h^l)=0$), או לא מוגדר (אם h^l שווה ל-0). לכן המשקולות יעודכנו להיות 0, ולכן לא ניתן להשתמש ב BackPropagation כדי לאמן רשת רב-שכבתית כזו של פרספטרונים בינאריים.

שאלה 2

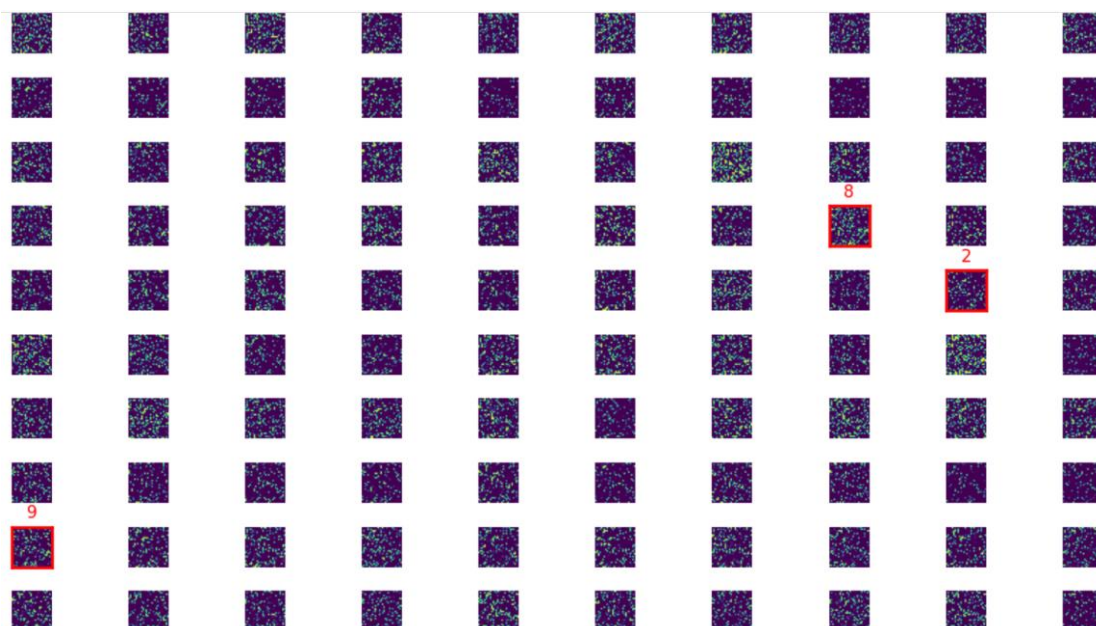
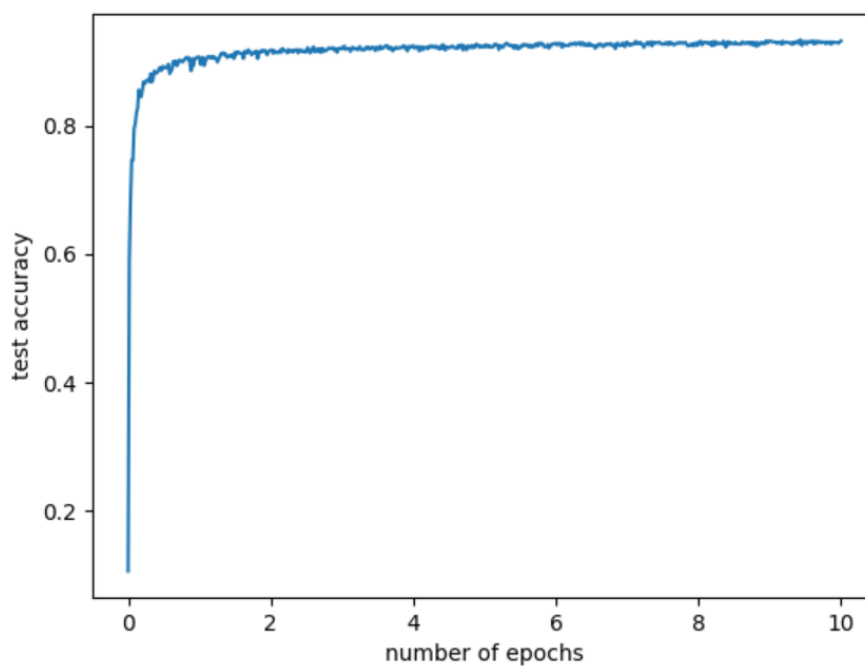
1.



2.

0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2
3	3	3	3	3 ²	3 ⁵	3	3	3	3
4	4	4 ⁷	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6 ³	6	6
7	7	7	7	7	7	7	7	7	7 ²
7 ⁵	8	8	8	8	8	8	8 ⁰	8	8
8	9	9	9	9	9	9	9	9	9

אחוז הטעויות הוא קטן, כלומר הרשת למדה בצורה טובה. ניתן לראות כי כאשר הרשת מבצעת טעויות, אלו טעויות שנראות "הגיוניות", כלומר, אפשר להבין את הדמיון בין הפלט הרצוי לפלט המצוי. נראה שהסיבה לכך היא העובדה שהתייחסנו לכל תמונה כווקטור חד ממדי באורך 784, ואולי איבדנו במובן מסוים את המשמעות של קווים אופקיים בתמונה.



נשים לב שבגרף של עקומת הלמידה, קיבלנו שהרשת לומדת באותו אופן, כפי שלמדה לפני הפרמוטציה. כלומר היא אינה לומדת טוב יותר או טוב פחות לאחר הפרמוטציה. נראה ששיבוש הסדר של התמונות לא השפיע על יכולת הרשת ללמוד. הסיבה לכך היא- ביצענו פרמוטציה זהה על כל התמונות, כלומר – שיבשנו את הסדר הפנימי של כל וקטור שמייצג תמונה. בצורה הזו אנו, בני אדם, לא מסוגלים לזהות איזו ספרה מופיעה בתמונה. עם זאת, הרשת לא באמת יודעת איך ספרה נראית, היא מבחינה מקבלת וקטור באורך 784, והיא לומדת את הדוגמאות הללו (כאשר שוב נדגיש שמתבצעת אותה פרמוטציה לכל הדוגמאות, ולכן כל תמונה משתנה "באותו אופן" כמו התמונות האחרות).