**תרגיל 3- ביולוגיה חישובית**

**מגישים: אלון לובושיץ (312115090) וגילי וולף (315144907)**

***הקדמה:***

מטרת המשימה הייתה לבנות רשת SOM שמבוססת על 10,000 תמונות של ספרות מתוך MNIST data set, בגודל 100 נוירונים, כך שכל ניורון בעל וקטור של 784 ערכים המייצגים את פיקסלי התמונה. את הרשת בנינו בעזרת מימוש על מטריצה הקסה-גונלית (כל ניורון בעל 6 שכנים), שאותחל כפי המפורט ב-[1]. לאחר מכן המודל אומן בעזרת 100 איטרציות או 90 שניות (הקצר מביניהם), כך שבכל איטרציה רצנו על כל הדוגמאות ולכל דוגמה:

א. מצאנו את הניורון הכי קרוב אליה בעזרת מרחק אוקלאידי

ב. עדכנו את ערכי הניורון והשכנים שלו כפי המפורט ב-[2]

לאחר אימון המודל, קיבלנו וקטור בגודל [1x10,000] כך שבכל אינדקס i, הערך בתא שקול לניורון המייצג של הדוגמה ה-i. הויזואליזציה נעשתה בעזרת צבעים, כך שכל ניורון נצבע בצבע המייצג את הספרה הדומיננטית שלו (לפי המקרא הנמצא בצד התמונה) , ואחוז ה'עומק' של הצבע מייצג את אחוז ההופעה שלה (ככל שהצבע עמוק יותר, כך אחוז ההופעה גבוה יותר). הנתונים נכתבו גם בתוך הניורונים עצמם.

חזרנו על התהליך כ-10 פעמים, עד אשר מצאנו את הפתרון הטוב ביותר (מופיע ב-[5]), לפי המפורט ב-[4].

***[1] אתחול ערכי הניורונים:***

ביצענו אנליזה מקדימה של אחוז ה'דלילות' (sparsity) של דוגמאות MNIST, וגילינו כי בממוצע כל דוגמה מכילה כ-80% ערכי אפס מתוך כלל ה-784 'פיקסלים'. בעקבות כך, החלטנו לאתחל את הניורונים בהתאם: 80% אפסים ו-20% ערכים רנדומליים בטווח [0,255].

***[2] עדכון ניורון ושכניו:***

לאחר מציאת הנוירון הקרוב ביותר לדוגמה מסוימת, עדכנו אותו ואת השכנים במעגל של 2 רמות ממנו.   
כלל העדכון שנבחר הינו:   
 - חישוב המרחק מהדוגמה לניורון ע"י חיסור ביניהם (וקטור בגודל 784)

- הוספת הוקטור המשוקלל [learning rate \* מרחק ] לערכי הניורון (וקטור בגודל 784)

עדכון הניורון נעשה על ידי : learning rate = 0.25

עדכון השכנים ברמה 1 (6 שכנים מסביבו) נעשה על ידי : learning rate = 0.16

עדכון השכנים ברמה 2 (12 שכנים מסביב לשכנים ברמה 1) נעשה על ידי : learning rate = 0.8

***[3] גודל הרשת:***

האם הגודל המוצע של הרשת מתאים לדעתכם לייצוג אוכלוסיית ספרות זו?

***[4] בחירת הפתרון:***

על מנת לבחור את הפתרון הטוב ביותר מבין הריצות השונות נעזרנו בשני מטריקות:

א. מדד הטופולוגי המציין את "דחיסות" הרשת. מדד זה בודק את שני הנציגים הכי קרובים לדוגמא מסוימת ומחשב את המרחק ביניהם. מכיוון שנוירון ברשת מיוצג ע"י תא במטריצה הקסגונלית נמיר את הקואורדינטה q,r קואורדינטות של קוביה - x,y,z ע"י הטרנספורמציה: x=q, y= -q -r, z=r ע"י התנאי x+y+z = 0. כעת המרחק בין שני קורדינטות כאלה יהיה מרחק מנהטן (מקסימום על המרחק האבסולוטי בכל ציר). המרחק המנהטי מתאר כמה "צעדים" צריך לקחת בין שני הקורנדיטות ולכן מרחק של 1 יהיה מינמלי - שכנים צמודים, 2 רמה שניה וכן הלאה.

ב. גרפים: הגרפים שהתקבלו, כל גרף מייצג את הרשת כפי שתואר בהקדמה כך שהצבע מעיד על הלייבל של הנוירון ואילו העומק על תדירות הקלטים שהוא מייצג. ככל שהגרף מחולק לצבעים בצורה מבודדת יותר (קלסטור טוב) ועומק הצבעים חזק יותר (קלסטור ייחודי) ככה הרשת יותר מדוייקת וטובה.

סה"כ נעזר במדד חישובי וגרפי על מנת לבחור את הפתרון הטוב ביותר.

מציאת הפתרון:

הגרלנו 10 רשתות שונות באמצעות הגרלת 10 learning rates שונים לכל רמת עדכון (נוריון, שכניו, שכני שכניו).

לעדכון הנוירון הגרלנו learning rates מטווח של: 0.2-0.8.  
לעדכון השכנים הראשונים הגרלנו learning rates מטווח של: 0.05-0.2

לעדכון השכנים האחרונים הגרלנו learning rates מטווח של 0.01-0.1

בנוסף, בדקנו שני מצבים בהם הlearning rate גבוה יותר לשכנים מאשר לנוירון.

לכל עשרת השילובים בנינו רשת וחשבנו את המדד הטופולוגי לכל נציג מלייבל מסוים. הנציגים זהים בין הרשתות על מנת שההשוואה תהיה צודקת. דווחנו על סכום המרחקים הכללי לכלל הנציגים ואילו לכל אחד בנפרד. בנוסף יצרנו גרף לכל רשת. מתוך הגרף וציון המדד בחרנו את הרשת בעלת הפרמטרים הטובים ביותר.

***[5] הצגת הפתרון הטוב ביותר:***

נציג שלוש פתרונות לצורך השוואה:

Net number 6:

Total topo score: 10

Topo scores: [1, 1, 1, 1, 1, 1, 1, 1, 1, 1]

Learning\_rates: (0.2666666666666667, 0.16666666666666669, 0.08)

Net number 7:

Total topo score: 17

Topo scores: [1, 1, 1, 1, 1, 1, 1, 2, 7, 1]

Learning\_rates: (0.33333333333333337, 0.08333333333333334, 0.030000000000000006)

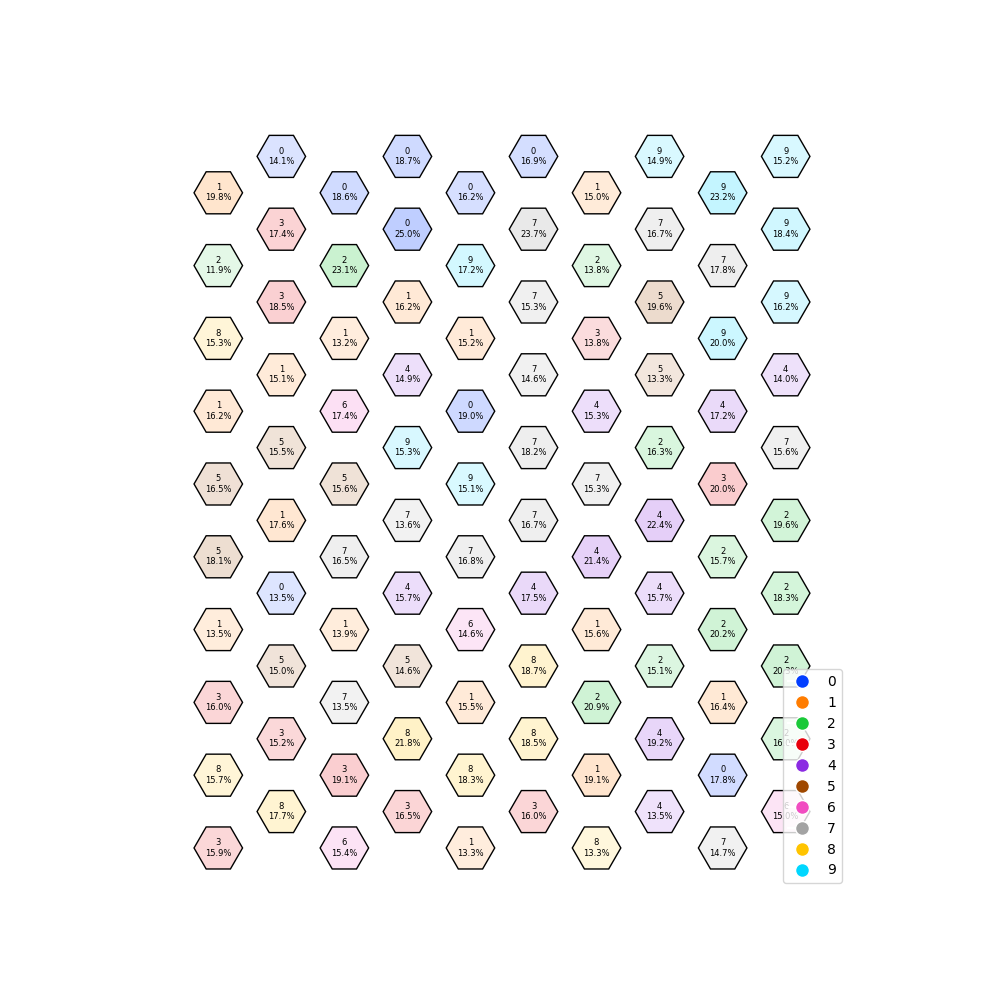
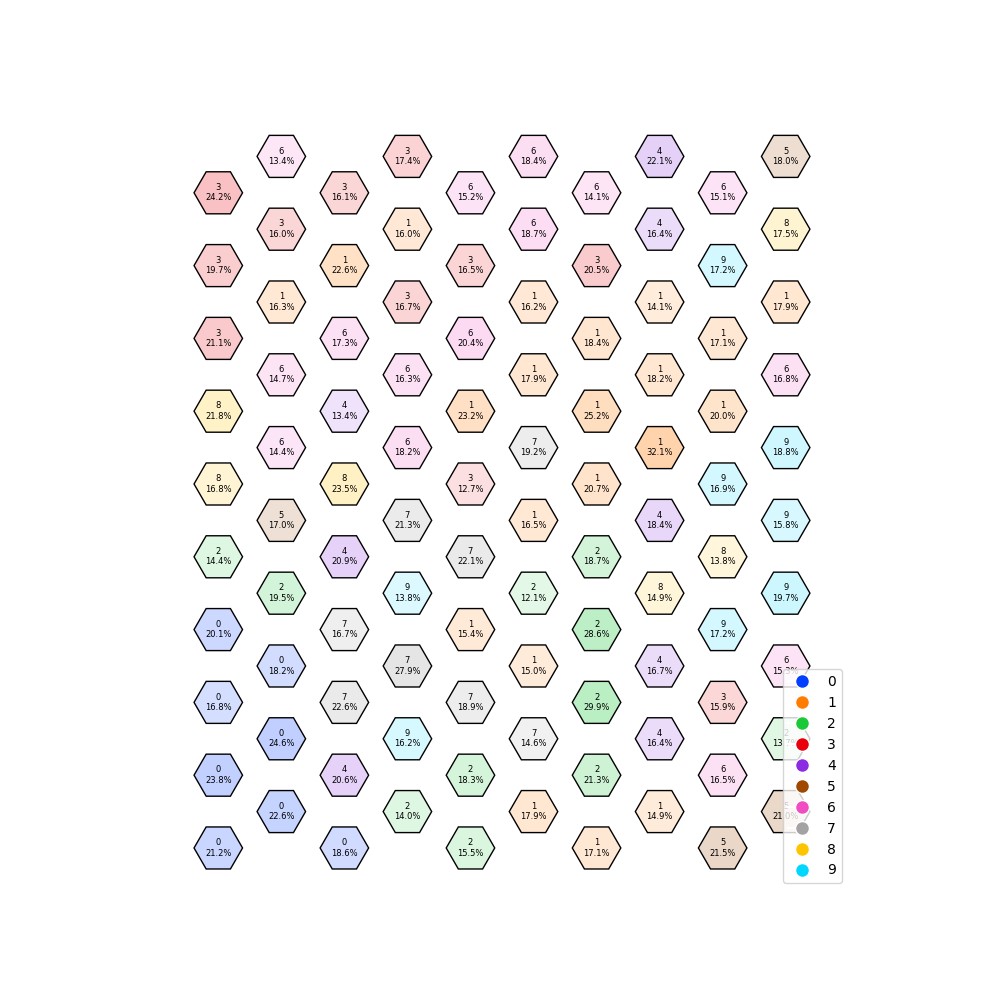
Net number 10:

Total topo score: 23

Topo scores: [1, 6, 3, 1, 3, 1, 1, 1, 4, 2]

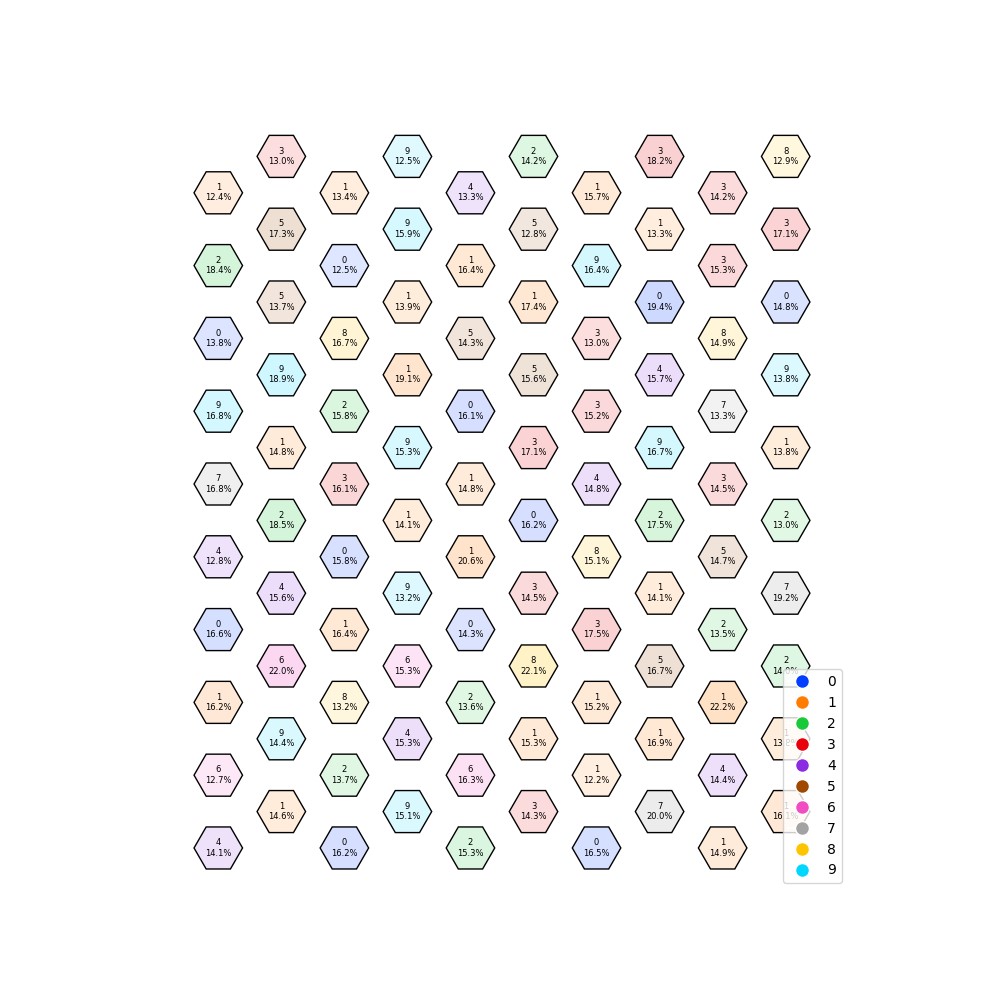
Learning\_rates: (0.09000000000000001, 0.18333333333333335, 0.5333333333333334)

נשים לב, רשת מס' 10 בעלת מקדמי למידה גדולים יותר לשכנים מאשר לנוירון. רשת מס' 6 בעלת ציון טופולוגי מינמאלי!

גרפים:

7

6

ניתן לראות ברשתות 6 ו-7 קלסטור טוב יותר מאשר ברשת 10. בנוסף ברשת 6 אחוזי התדירות ועומק הצבעים יותר חזקים ולכן בחרנו בפרמטרים לפי רשת 6.

10