

### תרגיל מס' 3: רשתות עמוקות

תאריך אחרון להגשה: 12.1.2020

יצאו את הקבצים בקובץ ex3.zip לתיקייה.

קוד המטלב הנתון מכיל סימולציה של רשת נוירונים המפעילה אלגוריתם למידה Back propagation. הקלט לרשת הוא תמונות של ספרות הכתובות בכתב יד כאשר הספרות הנתונות הן הספרות 0 ו 3 והרשת צריכה ללמוד לסווג תמונות לפי הספרה המופיעה בהן. התמונות הנתונות כקלט לרשת הן חלק מסט נתונים מאוד מפורסם הנקרא mnist, תוכלו לקרוא עליו כאן: [/http://yann.lecun.com/exdb/mnist](http://yann.lecun.com/exdb/mnist)

1. בקובץ doAll.m תמצאו את שורת הקוד להצגת התמונה הראשונה בסט האימון, הציגו את 10 הדוגמאות הראשונות בסט האימון והתרשמו מסט הנתונים.
  2. עתה הריצו את הקובץ doAll, הקובץ קורא לפונקציה בשם backprop המקימה רשת נוירונים ומריצה אותה מספר איטרציות (נתון כפרמטר). בסיום, הרשת מחזירה את שגיאת האימון (על סט האימון) וההכללה (על סט הבדיקה) כתלות במספר האיטרציות. ברשת יש שכבה חבויה בעלת מספר משתנה של נוירונים (תלוי בפרמטר לפונקציה) וכן שכבת פלט בעלת נוירון פלט אחד. כל הנוירונים ברשת הם נוירונים לוגיסטיים ופונקציית השגיאה היא Cross entropy. הלמידה מתבצעת באמצעות Stochastic Gradient Descent.
- הכנסו לקובץ המכיל את הפונקציה backprop, קראו את ההערה בראש הקובץ וודאו שאתם מבינים את מטרתם של הפרמטרים השונים. בנוסף, נסו להבין את הקוד שהפונקציה מריצה.

בתרגיל זה נחקור את ההשפעה של פרמטרים שונים על הרשת וכן נוסיף רגולריזציה לרשת:

#### 3. קבוע הלמידה:

נסו להריץ את הרשת עם ערכים שונים של קבוע למידה (אל תשנו את ערכי הפרמטרים האחרים): 0.1, 0.05, 0.01, 0.005, 0.001, 0.0005, 0.0001. איזה ערך נתן את התוצאות הטובות ביותר, הסבירו כיצד קבוע הלמידה משפיע על הלמידה ברשת.

#### 4. מספר הנוירונים בשכבה החבויה

שנו את קבוע הלמידה בחזרה לערך 0.01. עתה נסו לשנות את מספר הנוירונים בשכבה החבויה. נסו את הערכים הבאים: 10, 50, 100, 200.

- a. איזה מספר הראה את הביצועים הטובים ביותר?
- b. מה קרה לשגיאת האימון כאשר הגדלנו את מספר הנוירונים משמעותית? מה קרה לשגיאת ההכללה? כיצד נקראת תופעה זו?

#### 5. הוספת רגולריזציה

נרצה להוסיף רגולריזציה (דעיכת משקולות) לרשת. תזכורת: המטרה ברגולריזציה היא לשמור על משקולות קטנים, השיטה שבה נשתמש על מנת לבצע זאת היא "הענשה" על משקולות גבוהים בפונקציית השגיאה:

$$RegLoss = loss + \frac{1}{2} \alpha \|w\|^2$$

כאשר  $\alpha$  הוא פרמטר שקובע כמה דגש יש לשים על הקטנת המשקולות. שימו לב שתוספת הרגולריזציה באה לידי ביטוי בתוך ה Gradient שלנו.

$$\frac{\partial \text{RegLoss}}{\partial w_{ij}^k} = \frac{\partial \text{loss}}{\partial w_{ij}^k} + \alpha w_{ij}^k$$

ולכן כאשר נבצע צעד של *Gradient Descent* עדכון המשקולות צריך להיות:

$$\begin{aligned} (w_{ij}^k)^{\text{new}} &= w_{ij}^k + \Delta^{\text{reg}} w_{ij}^k = w_{ij}^k - \lambda \frac{\partial \text{RegLoss}}{\partial w_{ij}^k} = w_{ij}^k - \lambda \left( \frac{\partial \text{loss}}{\partial w_{ij}^k} + \alpha w_{ij}^k \right) \\ &= w_{ij}^k - \lambda \frac{\partial \text{loss}}{\partial w_{ij}^k} - \lambda \alpha w_{ij}^k = w_{ij}^k + \Delta w_{ij}^k - \lambda \alpha w_{ij}^k \end{aligned}$$

כאשר  $\Delta w_{ij}^k$  הינו העדכון במשקולות לפני הרגולריזציה.

- a. עדכנו את הפונקציה *UpdateWeights* כך שעדכון המשקולות יכלול את הרגולריזציה (שימו לב שהפרמטר *alpha* כבר קיים אך לא נעשה בו שימוש עד עתה).
- b. הגדילו את מספר האיטרציות שהאלגוריתם רץ ל 300 ושנו את מספר הנירונים בשכבה הנסתרת ל 50 נירונים. הריצו את הרשת עם ערכי *alpha* שונים (0.05, 0.01, 0.005, 0.001) ובחנו את ההשפעה על פעילות הרשת. כיצד משפיע ערך *alpha* על ביצועי הרשת? מהו ערך *alpha* הטוב ביותר?
- c. עתה בחרו מספר גדול יותר של נירונים (200), האם הצלחתם לשפר את ביצועי הרשת ביחס למצב ללא רגולריזציה? הסבירו מדוע, מה השפעת הרגולריזציה?

### הנחיות הגשה:

יש להגיש את הקבצים הבאים:

1. doAll.m

2. UpdateWeights.m

עם השינויים שביצעתם.

בנוסף יש לצרף דו"ח עם התשובות לשאלות כולל גרפים שיעזרו להמחיש את טענותיכם.

בהצלחה!