**תרגיל בית 1**

# הקדמה

במסגרת תרגיל בית זה מימשנו רשת נוירונים Fully-connected (FC) לסיווג ספרות ממאגר הנתונים MNIST. ביצועי הרשת שאומנה על סט הבוחן הם 98.87 דיוק.

# תהליך הלמידה ווולידציה

בוצעה השוואה על מנת לבחור את המודל הטוב ביותר ולכן סט האימון חולק ונלקחו ממנו 8,750 תמונות לצורך וולידציה, בהתפלגות זהה לסט האימון הנותר.

הוולידציה בדקה ביצועי רשתות בעלות ארכיטקטורה שונה, קצבי לימוד ומקדמי רגולריזציה שונים.

נבחנו 7 ארכיטקטורות שונות:

* מודל 1 - רשת בעלת 3 שכבות FC עם [78, 39, 10] נוירונים. לאחר שתי השכבות הראשונות ישנה פונקציית אקטיבציה ReLU ולאחר השכבה האחרונה Softmax. סה"כ 64,711 משקלים. ראה איור 1.
* מודל 2 - רשת בעלת 12 שכבות FC כאשר 11 השכבות הראשונות בעלות 50 נוירונים שלאחר כל שכבה ישנה פונקציית אקטיבציה ReLU ולאחר השכבה האחרונה Softmax. סה"כ 42,310 משקלים במודל. ראה איור 2.
* מודל 3 - רשת שמבצעת חלוקה של תמונת הקלט לארבע רבעים. החלקים השונים מועברים ברשתות מקבילות זהות בצורתן - 3 שכבות FC בעלות [210, 105, 10] נוירונים ולאחר כל שכבה ישנה פונקציית אקטיבציה ReLU. היציאות נערמות (concatenate) ומועברות דרך שכבה נוספת בעלת 10 נוירונים שלאחריהן פונקציית Softmax. סה"כ 64,995 משקלים במודל. ראה איור 3.
* מודל 4 - רשת שמבצעת חלוקה של תמונת הקלט ל-16 ריבועים שווים (7X7 כל אחד). החלקים השונים מועברים ברשתות מקבילות זהות בצורתן - 2 שכבות FC בעלות [36, 18] נוירונים ולאחר כל שכבה ישנה פונקציית אקטיבציה ReLU. היציאות נערמות ומועברות דרך 3 שכבות FC נוספות בעלת [144, 72, 36] נוירונים שלאחריהן שיכבה אחרונה בעלת 10 נוירונים ופונקציית Softmax. סה"כ 60,850 משקלים. ראה איור 4.
* מודל 5 רשת שמבצעת חלוקה של תמונת הקלט ל-16 ריבועים שווים (7X7 כל אחד). החלקים השונים מועברים ברשתות מקבילות זהות בצורתן - 4 שכבות FC בעלות [120, 90, 60, 30] נוירונים ולאחר כל שכבה ישנה פונקציית אקטיבציה ReLU. כל 4 יציאות נערמות יחדיו ומועברות ברשתות מקבילות זהות בצורתן - 4 שכבות FC בעלות [120, 90, 60, 30] נוירונים ולאחר כל שכבה ישנה פונקציית אקטיבציה ReLU. לבסוף עורמים את ה-4 היציאות הנותרות דרך 4 שכבות FC נוספות בעלות [120, 90, 60, 30] נוירונים שלאחריהן שיכבה אחרונה בעלת 10 נוירונים ופונקציית Softmax. סה"כ 57,520 משקלים במודל. ראה איור 5.
* מודל 6 - רשת שמבצעת חלוקה של תמונת הקלט ל-16 ריבועים שווים (7X7 כל אחד). החלקים השונים מועברים ברשתות מקבילות זהות בצורתן – שכבה FC בעלת [18] נוירונים ולאחריה ישנה פונקציית אקטיבציה ReLU. היציאות נערמות ומועברות דרך 5שכבות FC נוספות בעלת [288, 144, 72, 36, 18] נוירונים שלאחריהן שיכבה אחרונה בעלת 10 נוירונים ופונקציית Softmax. סה"כ 56,440 משקלים במודל. ראה איור 6.
* מודל 7 - רשת שמבצעת חלוקה של תמונת הקלט ל-16 ריבועים שווים (7X7 כל אחד). החלקים השונים מועברים ברשתות מקבילות זהות בצורתן - שיכבה FC בעלת [285] נוירונים ולאחריה ישנה פונקציית אקטיבציה ReLU. היציאות נערמות ומועברות דרך שיכבה FC נוספות בעלת [4560] נוירונים שלאחריהן שיכבה אחרונה בעלת 10 נוירונים ופונקציית Softmax. סה"כ 59,860 משקלים במודל. ראה איור 7.

פונקציית ההפסד שנבחרה היא אנטרופיה צולבת (NLLL לאחר Softmax במוצא הרשתות) ואלגוריתם ADAM שימש לצורך אופטימיזציה שלה.

קצבי הלימוד, מקדמי הדעיכה (לצורך רגולריזציה) וגדלי תת-קבוצות האימון שנבחנו במסגרת תהליך הוולידציה הם . הוולידציה בוצעה עבור כל שילוב מודל-פרמטרים אפשרי למשך 100 epochs.

עיבוד מקדים של התמונות כלל החסרת ממוצע ונרמול לפי סטיית תקן בלבד.

התוצאות הטובות ביותר, כלומר - השגיאה היחסית על פני סט הוולידציה, מוצגות בטבלה 1 עבור כל המודלים יחד עם הפרמטרים המתאימים (lr, wd, bs, epoch).

|  |  |
| --- | --- |
| איור 1 - מודל 1 | איור 2 - מודל 2 |
| איור 3 - מודל 3 | איור 4 - מודל 4 |
| איור 5 - מודל 5 | |
| איור 6 - מודל 6 | איור 7 - מודל 7 |

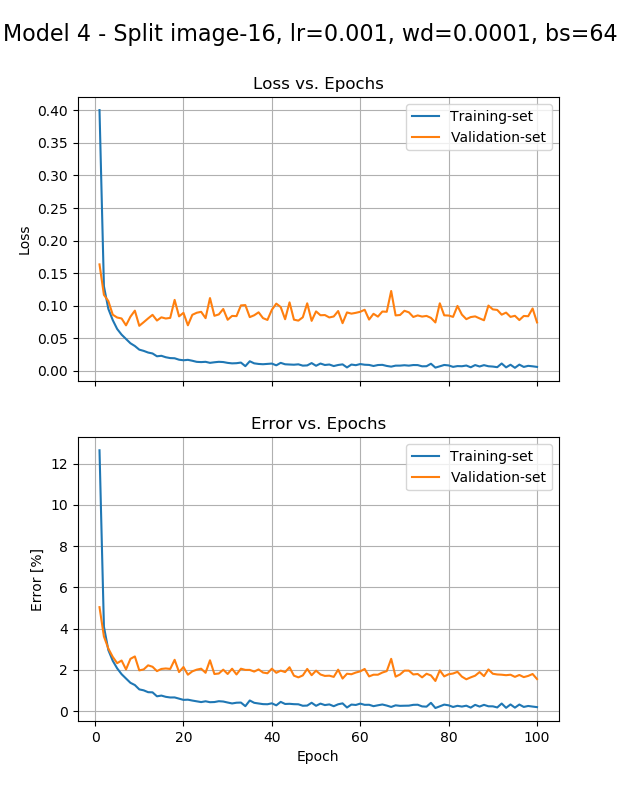
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **מודל** | **מספר משקלים ברשת** | **שגיאת וולידציה יחסית מינימלית במהלך האימון [%]** | **lr** | **wd** | **bs** | **epoch** |
| **1** | 64,711 | 2.240 | 1e-3 | 1e-4 | 32 | 63 |
| **2** | 42,310 | 2.940 | 1e-3 | 1e-5 | 32 | 93 |
| **3** | 64,995 | 1.727 | 1e-3 | 1e-5 | 64 | 44 |
| **4** | 57,520 | 1.458 | 1e-3 | 1e-4 | 64 | 77 |
| **5** | 60,850 | 1.715 | 1e-3 | 1e-4 | 64 | 97 |
| **6** | 56,440 | 1.727 | 1e-3 | 1e-4 | 64 | 56 |
| **7** | 59,860 | 1.913 | 1e-4 | 1e-4 | 32 | 94 |

טבלה 1 - השוואת טיב המודלים

# אימון המודל המיטבי

המודל שהניב את התוצאות הטובות ביותר הוא מודל 4, כפי שניתן לראות בטבלה 1. ערכי פונקציית ההפסד והשגיאה היחסית לפי ה- epoch על פני סט האימון (mini-batch) וסט הוולידציה מוצגים באיור 8. ניתן לראות שערך השגיאה ממשיך לרדת.

המודל הטוב ביותר אומן למשך 100 epochs נוספים על סט האימון והוולידציה יחד. והדיוק שהתקבל לאחר מכן על סט הבוחן הוא 98.87.

*איור 8 - הפסדים ושגיאות בזמן הלמידה*

# סיכום ומסקנות

התחלנו לנסות להתמודד עם הבעיה בשני מודלים פשוטים ונאיביים. מודל 1, רשת רדודה ושנועדה למקסם את מספר הפרמטרים ומודל 2, רשת עמוקה אך "רזה" בנוירונים. ראינו כי הראשונה נותנת תוצאות מיטביות (אחוז שגיאה קטן ב-20%) והחלטנו להמשיך לחקור רשתות רדודות.

הכיוון השני אליו פנינו נבע מהסתכלות על תמונות במאגר: רוב המידע נמצא במרכז התמונה ועל כן הקצוות לא רלוונטיים ומכניסים לנו רעש. כדי ללמד את המערכת לשים דגש על מרכז התמונה חילקנו את התמונות, תחילה ל-2X2 ואז ל-4X4 בכוונה שהנוירונים שמקשורים לאותם פיקסלים ריקים "יושתקו". מודל 3 הוא הראשון שישם את עקרון זה והציג שיפור משמעותי מאלו שלפניו (אחוזה שגיאה קטן ב-20%).

בשלב זה התחלנו להתמקד באופטימיזציה של תהליך החלוקה: ניסינו מספר שכבות לפני איחוד המידע (concatenate) כמו במודלים 4 ו-5, ניסינו איחוד בשלבים (מודל 5), וניסינו עומקים שונים לאחר האיחוד כמו במודלים 6 ו-7. תוצאה הטובה ביותר קיבלנו כאשר הצבנו מספר נמוך של שכבות לפני האיחוד ומספר רב לאחר העיבוד. לגבי איחוד בשלבים, נראה שבשל הפשוטות היחסית של הקלט, פרמטרים "מתבזבזים" על שלבי הביניים ועל כן התוצאות מעט פחות טובות.

נציין כי הניסויים נעשו עבור קצבי הלימוד, מקדמי הדעיכה וגדלי תת-קבוצות האימון שונים והורצו מספר פעמים כדי לקבל אתחול מיטבי.

לסיכום עבור מאגר MNIST ומערכת ללא קונבולציה:

* בשל פשטות יחסית של הבעיה (תמונות בשחור לבן וברזולוציה נמוכה ומספר סופי ונמוך של קלסיפיקציות) מערכת עמוקה לא נותנת תוצאות טובות יותר מאשר מערכת רדודה.
* רוב המידע של נמצא במרכז התמונה ועל חלוקה של התמונות אפשרה למערכת להתמקד באזורים ה"חשובים".
* חיבור מספר נוירונים לכל פיקסל אפשר לייצר מסלולים שונים למערכת להבין את אותו הפיקסל (בעצם מעין פילטר) מה שאפשר לנו לשפר את התוצאות.