**ניתוח ואיבוד מידע ספרתי - דוח פרויקט 17.2.2017**

הבעיה: סיווגת אותיות בעברית, מספרים וסימנים מתמטיים.

איסוף הנתונים:

מאגר התמונות שלנו כולל אוסף של כ 21000 תמונות שאספנו ושיתפנו בין הסטודנטים בקורס. התפלגות התמונות בין הספרות שווה (בקירוב). את אוסף התמונות חילקנו באופן הבא:

85% מהתמונות הלכו ללימוד, ו15% למבחן כך:

18000 תמונות – לימוד

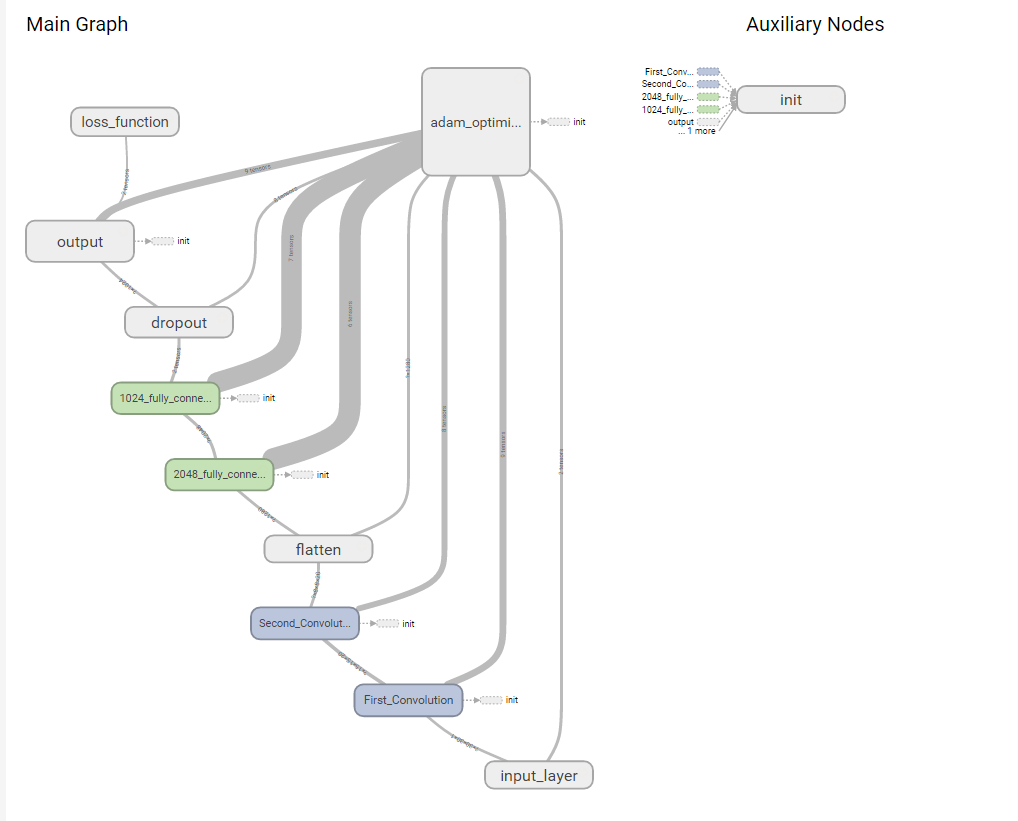
3000 תמונות - מבחן

פתרון בעזרת CNN:

בנינו מערכת למידה המבוססת על DEEP LEARNING ברשת נוירונים מסוג CNN.

הארכיטקטורה דומה לפתרון ש-TensorFlow מציעים עבור MNIST.

ארכיטקטורה:



שכבת כניסה לתמונות,

1. שכבת קונבולוציה – גודל 5X5 צעד 1X1
2. MAXPOOLING
3. שכבת קונבולוציה 2 - גודל 5X5 צעד 1X1
4. MAXPOOLING נוסף
5. ושכבת FULLY CONNECTED – גודל 2048
6. שכבה FULLY CONNECTED – גודל 1024
7. ואז ביצוע DROP OUT
8. מספר הOUTPUT שיש לנו ברמה האחרונה היא 50 כמספר המחלקות שיש ועוד 1, מצופה ממחלקה 22 להיות ריקה.

המערכת קולטת נתונים לזיכרון ומשנה את גודלם להיות לגודל אחיד, ולאחר מכן מריצה את הנתונים לתוך הCNN, ולבסוף מדפיסה לפלט את התוצאה.

האלגוריתם:

ממומש בPYTHON 3.5 בעזרת ספריית TENSORFLOW

1. קריאת קבצי התמונות לזיכרון
2. שינוי גודלם של התמונות לגודל אחיד של 30\*30
3. המרה של התמונות להיות שחור לבן
4. איתחול של המשתנים של הרשת, learning rate, batch size, number of epochs
5. הרצה בלולאה מספר הEPOCHS
   1. הרץ את BATCH SIZE של התמונות לתוך המערכת
   2. אם זו איטרציה המתחלקת ב10 – הרץ בדיקה על הטסט דאטה.
      1. הדפס את התוצאה

הניסויים שבוצעו במערכת:

1. שינויי LR
   1. <למלא>
2. שינוי מספר הBATCH SIZE
3. שינוי גודל התמונות:
   1. 30\*30
   2. 20\*20
   3. <למלא>

טבלת התאמה:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| א | ב | ג | ד | ה | ו | ז | ח | ט | י | כ | ל | מ | נ | ס | ע |
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| פ | צ | ק | ר | ש | ת |  | ך | ם | ן | ף | ץ | ( | ) | + | - |
| 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ∫ | <= | >= | ∞ | ∑ | ∩ | ≠ | ÷ | ͓ |
| 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 |

תוצאות הניסוי