תרגיל בית 3 – נושאים מתקדמים במערכות מידע

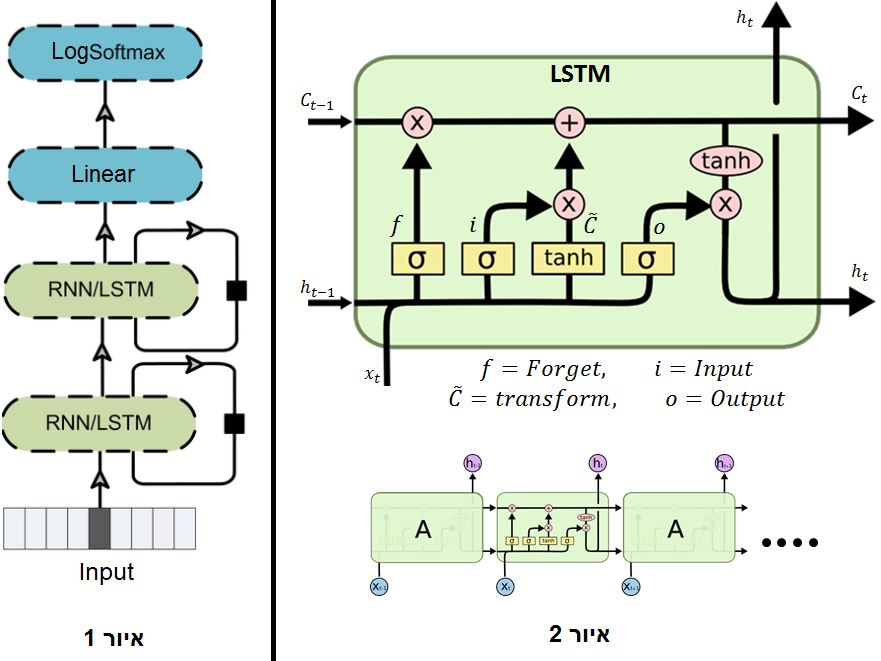
*שמעון ארזואן 037036209*

*שגיא חכמון 365656870*

*גיא רוזין 204097984*

**תיאור ארכיטקטורה**

הרשת שבנינו מורכבת מ-4 שכבות עיקריות. 2 שכבות LSTM, שכבה לינארית ולבסוף שכבת LogSoftMax כמתואר באיור 1. כל שכבת LSTM בנויה באופן המתואר באיור 2.



פרמטרים שלנו שכיילנו (לאחר ניסוי וטעיה):

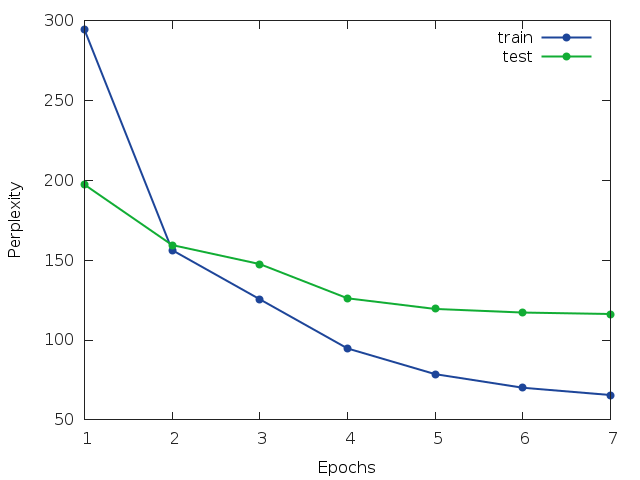
|  |  |
| --- | --- |
| פרמטר | ערך |
| גודל Batch | 20 |
| אורך רצף | 20 |
| גודל RNN | 133 |
| Dropout | ללא |
| מספר Epochs | 7 |
| מספר פרמטרי המודל | 2,979,440 |

**שיטת האימון**

ראשית טענו את הנתונים מ-3 קבצי הטקסט. איפסנו את המצבים, בנינו את הרשת לפי הארכיטקטורה המתוארת בסעיף לעיל ואימנו אותה בעזרת קובץ סט האימון. כיילנו את הפרמטרים של הרשת: עומק הרשת (הכוונה ל Recurrent), אורך רצף וגודל Batch עד לקבלת התוצאה הרצויה (Perplexity < 120) במספר הפרמטרים הרצוי (קטן מ-3 מיליון). למזלנו דיי מהר קבלנו את התוצאה הרצויה ולא היה צורך להרבה שינויים ברשת. לכן גם לא ניסינו שכבה 1 או 3 או 4 של LSTM במקום ה-2 שקבענו.

בנוסף הכנסנו לרשת קובץ קלט של המשפט שאותו נדרשנו להשלים וקבלנו השלמה של המשפט.

**גרפים**



ניסיונות: בדקנו learning rate, batch size, decay

<https://github.com/sagi180/Deep-Learning>