**תרגיל בית 3**

מגישים:

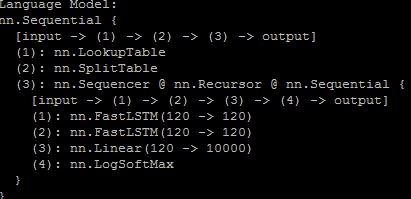
רם יזדי 305246308

גיא לורברבום 302773338

ספי לנגמן 305183527

קישור לgithub:

<https://github.com/sefilang/machine-learning>

הארכיטקטורה של המודל

input

FastLSTM

FastLSTM

Linear

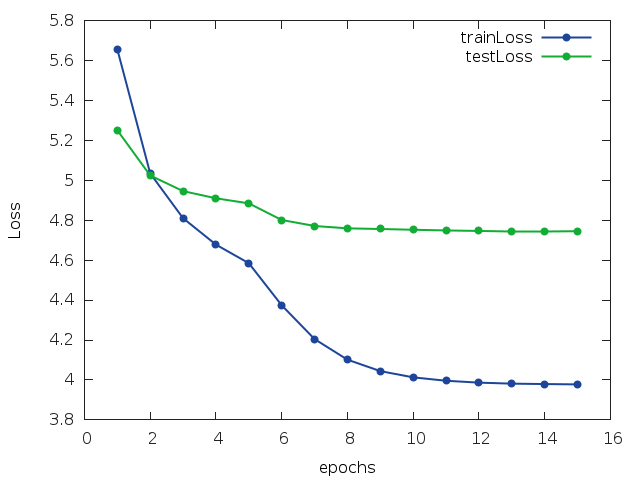
LogsoftMax

Output

**הסבר:** במהלך האימון המודל מקבל את ה-input – המשפטים שמרכיבים את מדגם האימון. כל מילה במדגם האימון הופכת ממילה לווקטור המייצג אותה על ידי פונקציית הlookup table.

לאחר מכן המודל מכניס 20 וקטורים (כגודל הsequence שהגדרנו) כאשר כל ווקטור אורכו 120. על הווקטורים האלו מריץ את מודל ה-FastLSTM פעמיים. המודל מנבא את המילה הבאה במשפט והמשפט שהתקבל ממשיך להתקדם בשני דרכים- האחד דרך השכבה הליניארית, שכבת ה- LogsoftMax ויוצא כשכבת output לשלב ה-t והשני כקלט לשכבה ה- t+1.

הקריטריון שאנו ממזערים הוא מסוג ClassNLLCriterion.

להלן גרף המתאר את הloss שהתקבל על מדגם האימון והמבחן:

ה-perplexity עבור מדגם המבחן אשר הושג באמצעות המודל (נבחר המודל הטוב ביותר על validation-set) הינו 114.2. הרצנו 15 אפוקים לטובת אימון הרשת והתוצאה הטובה ביותר הושגה לאחר האפוק ה-14. (לאחר ניסוי וטעיה מצאנו כי מתחילה התאמת-יתר החל מסביבות האפוק ה-15)

להלן משפטי ההמשך שהרשת ייצרה עפ"י התפלגות מולטינומית על פני וקטור ההסתברויות לכל מילה להופיע.

