

תרגיל בית 3

מגישים :

רם יזדי 305246308

גיא לורברבום 302773338

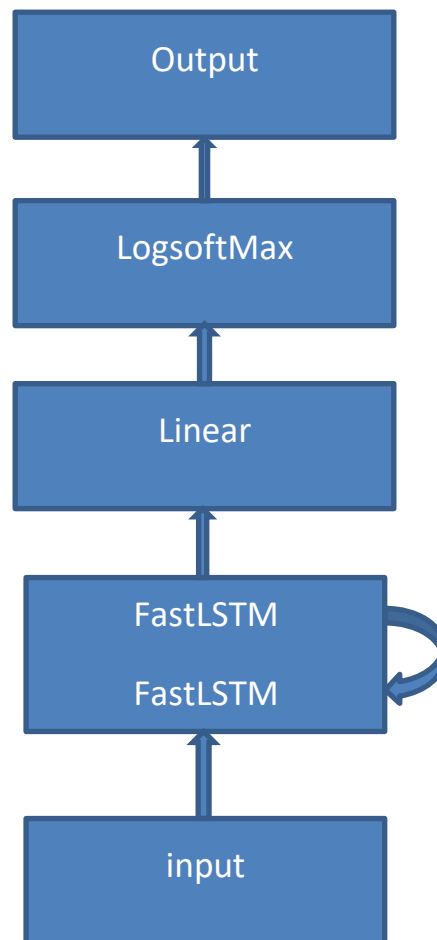
ספי לנגמן 305183527

קישור לgithub :

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

הארכיטקטורה של המודל

```
Language Model:
nn.Sequential {
  [input -> (1) -> (2) -> (3) -> output]
  (1): nn.LookupTable
  (2): nn.SplitTable
  (3): nn.Sequencer @ nn.Recursor @ nn.Sequential {
    [input -> (1) -> (2) -> (3) -> (4) -> output]
    (1): nn.FastLSTM(120 -> 120)
    (2): nn.FastLSTM(120 -> 120)
    (3): nn.Linear(120 -> 10000)
    (4): nn.LogSoftMax
  }
}
```

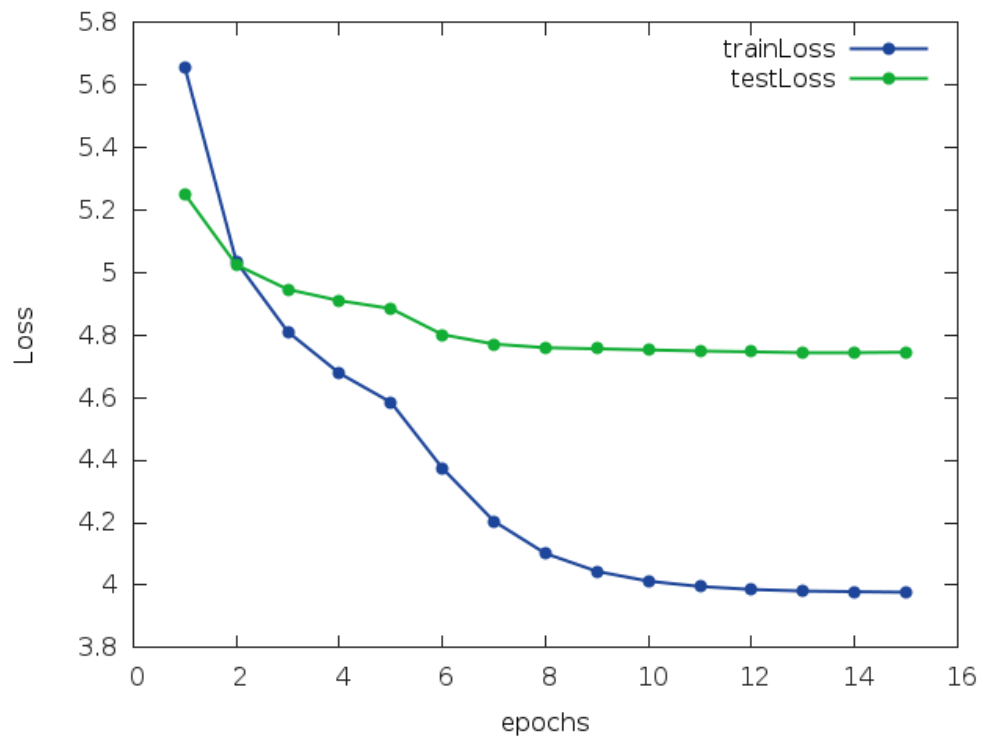


הסבר: במהלך האימון המודל מקבל את ה-input – המשפטים שמרכיבים את מדגם האימון. כל מילה במדגם האימון הופכת ממילה לווקטור המייצג אותה על ידי פונקציית lookup table.

לאחר מכן המודל מכניס 20 וקטורים (כגודל sequence שהגדרנו) כאשר כל וקטור 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 ה loss 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.

```
Buy low, sell high is the stock capacity on on listed length trading problem policy volume
Buy low, sell high is the wars effect for software said <unk> troubles government most device
Buy low, sell high is the move business them time actively past scientific political impact test
Buy low, sell high is the mill asset <unk> more about computer position legislature optical second
Buy low, sell high is the fine service from attempt happening holding fed closed soviet consumer
```