

APML ex2

השוואת המודלים:

לפי גודל סט האימון מכלל הדאטא (באחוזים) ועבור טסט על 10% שלא נצפו על ידי המודל:

baseline accuracy : 0.7901405413177096	• 10%
HMM accuracy : 0.8736026742388843	
MEMM accuracy : 0.8353752930249015	
baseline accuracy : 0.8543276661514683	• 20%
HMM accuracy : 0.9157830415872902	
MEMM accuracy : 0.849175293624108	
baseline accuracy : 0.8962204809316703	• 50%
HMM accuracy : 0.9404047302397469	
baseline accuracy : 0.9014413572481219	• 90%
HMM accuracy : 0.9434150461881313	

דוגמה לדגימת רצף מילים ממודל ה HMM:

(HMM) Sampling from a Generative Model

sample : ['There', 'reject', 'distorted', 'one-quarter', 'happened', '','', 'checks', '22.25', 'cultures', 'ought', 'distance', 'Each', 'formal', 'claim']

sample pos: [['EX', 'VBP', 'JJ', 'NN', 'VBD', '','', 'VBZ', 'CD', 'NNS', 'MD', 'VB', 'DT', 'JJ', 'NN']]

נראה שאין היגיון מבחינת האנגלית למרות שיש קישורים שקצת מתאימים בשפה, בכל מקרה נראה שיש היגיון לא רע מבחינת התגיות כמצופה.

מודל ה MEMM:

תחילה אימנתי את המודל על פונקציית phi הבסיסית שראינו בהרצאה שמשמשת במאפיינים של מודל HMM על ידי מתן התייחסות להסתברויות ה transmission וה- emission .

התוצאה שקיבלתי היתה: 74%

לאחר מכן כתבתי פונקציה חדשה (מופיעה בקוד תחת השם phi_2)

הפונקציה מוסיפה למדדי ה emission, transmission גם התייחסות ל:

- אות גדולה בתחילת מילה
- אורך מילה קטן מ 3 בשאיפה שמילים מהסוג הזה יהיו מילות יחס קישור וכו'
- שלוש סיומות שונות שמאפיינות שם עצם, שלוש סיומות לשם תואר וסיומת אחת לפועל
- מילה המכילה מספרים

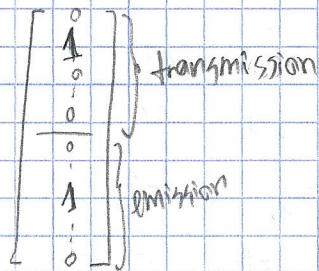
התוצאה עבור הפונקציה הנ"ל : 77%

** את האימון ביצעתי על 1500 משפטים וטסט של 100 משפטים שהמודל לא ראה בתהליך הלמידה.

EX 2 PART A

① HMM מודל MEMM

המודל מציג את המרחב המשותף בין המילים והמילים המצויות בהן.



המודל מציג את המרחב המשותף בין המילים והמילים המצויות בהן. המודל מציג את המרחב המשותף בין המילים והמילים המצויות בהן.

המודל מציג את המרחב המשותף בין המילים והמילים המצויות בהן.

המודל מציג את המרחב המשותף בין המילים והמילים המצויות בהן.

המודל מציג את המרחב המשותף בין המילים והמילים המצויות בהן.

$$P(y_{1:n}, x_{1:n}) = \prod_{i=1}^n \frac{e^{w_1 y_i + w_2 x_i}}{Z(y_{1:n}, x_{1:n})}$$

המודל מציג את המרחב המשותף בין המילים והמילים המצויות בהן.

$$\prod_{i=1}^n \frac{P(y_i | y_{i-1}) \cdot P(x_i | y_i)}{Z}$$

המודל מציג את המרחב המשותף בין המילים והמילים המצויות בהן.

המודל מציג את המרחב המשותף בין המילים והמילים המצויות בהן.

