**חלק 4:**

שלום בודק יקר,

לצורך הכנת הtest.pred בחרנו להפעיל את המודל יחד עם ההיפר פרמטרים הבאים:

**מודל pos בלי pre-trained embeddings:**

**היפר פרמטרים (רלוונטים למודל שלנו):**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Layer size** | **Learning rate** | **Epochs** | **גודל batch** | **פונקציית אקטיבציה** | **oprimizer** |
| 100 | 0.01 | 3 | 1024 | tanh | adam |

**גרפים רלוונטים למודל זה:**

**מודל pos עם pre-trained embeddings:**

**היפר פרמטרים (רלוונטים למודל שלנו):**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Layer size** | **Learning rate** | **Epochs** | **גודל batch** | **פונקציית אקטיבציה** | **oprimizer** |
| 100 | 0.01 | 3 | 1024 | tanh | adam |

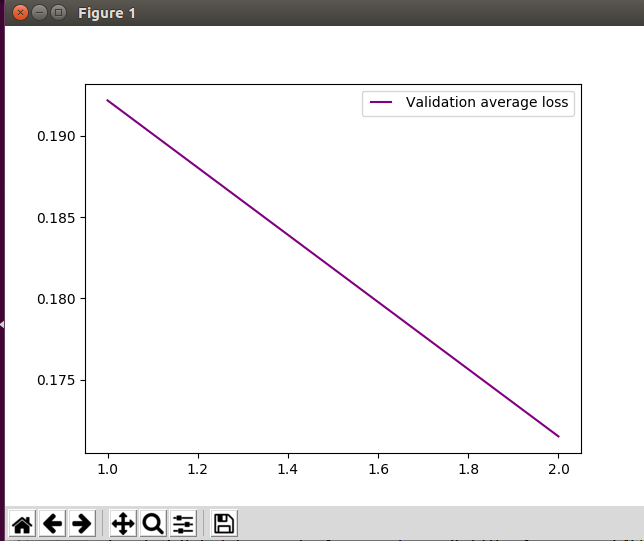
**גרפים רלוונטים למודל זה:**

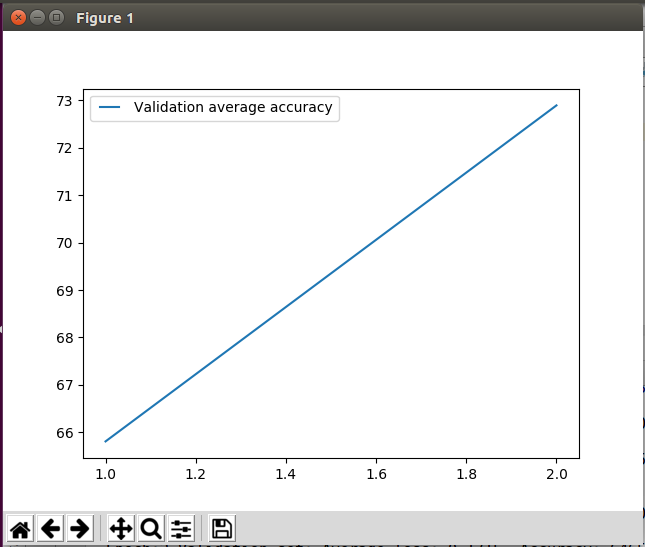
**מודל ner בלי pre-trained embeddings:**

**היפר פרמטרים (רלוונטים למודל שלנו):**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Layer size** | **Learning rate** | **Epochs** | **גודל batch** | **פונקציית אקטיבציה** | **oprimizer** |
| 100 | 0.01 | 2 | 1024 | tanh | adam |

**גרפים רלוונטים למודל זה:**



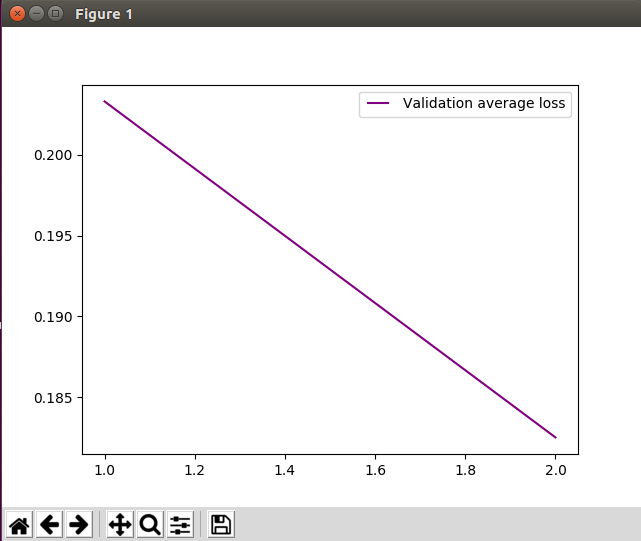


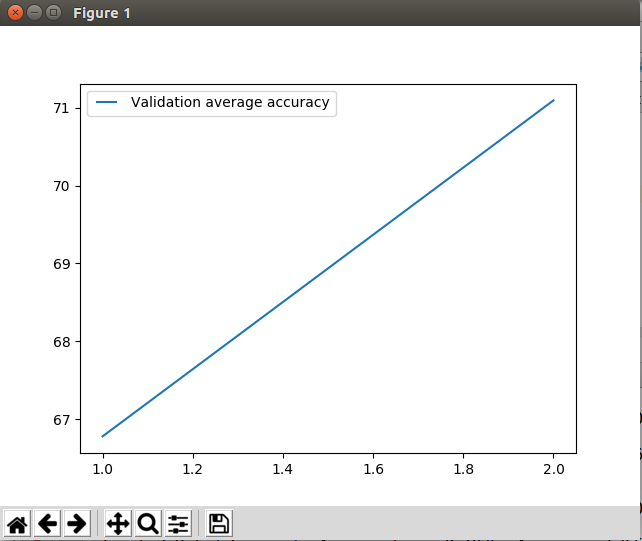
**מודל ner עם pre-trained embeddings:**

**היפר פרמטרים (רלוונטים למודל שלנו):**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Layer size** | **Learning rate** | **Epochs** | **גודל batch** | **פונקציית אקטיבציה** | **oprimizer** |
| 100 | 0.01 | 2 | 1024 | tanh | adam |

**גרפים רלוונטים למודל זה:**





**ניתוח התוצאות:**

יצרנו עבור הsuffixes וכן הprefixes 2 מטריצות embeddings כאשר כל וקטור במטריצות הנ"ל מייצג רישא או סיפא של מילים מתוך ה train set. נזכיר שבחרנו למפות מילים (בשלב ה validation וה test ) שלא מוכרות לנו משלב ה train למילה unk (“UUUNKKK”). ולכן יכלנו לחלץ רישא וסיפא של מילים לא מוכרות.

מהתוצאות שיצאו ראינו שהתוצאות על הner במודל זה פחות טובות ביחס לתוצאות של מודל של tagger2 ששם לא השתמשנו בwordEmbeddings עבור sub words. בשימוש ללא pre trained word embeddings הגענו ל 73% אחוזי דיוק ועם שימוש ב pre trained word embeddings הגענו במקסימום ל 71% דיוק. עושה רושם שהשימוש ב suffixes ו prefixes לא תרם כאן. על אף שאנו אמורים לכאורה לקבל תוצאות טובות יותר כי מסופק כאן יותר מידע על המילה שאנו רוצים לתייג , אנו משערים שהמבנה הבסיסי של המודל כנראה לא מאפשר לנו לנצל את כל המידע הזה ולעבד אותו בצורה טובה. המודל הבסיסי בתרגיל זה מורכב רק משכבה נסתרת אחת ומפונקציית אקטיבציה פשוטה יחסית, tanh. ייתכן ששכבה אחת לא מספיקה כדי לעבד ולתמצת את כל המידע באופן שבו נראה תוצאות טובות כפי שציפינו, ואף ייתכן ש"העמסת מידע" נוסף על המילה ברשת לא עמוקה מפריע לה לבצע פרידקציה בצורה טובה ומוביל לביצועים פחות טובים.

כמו כן, אנו לא מרגישים שה pre-trained וה sub-words משלימים אחד השני שכן השימוש בשניהם הוביל את המודל שלנו לתוצאות הנמוכות ביותר לעומת שימוש רק באחד מהם.

במידה דומה זאת עבור מודל הpos ראינו שאין הבדל משמעותי מבחינת אחוזי דיוק וה Loss כאשר השתמשנו ב pre trained vectors וב sub words לעומת המודלים הקודמים בתרגיל ואפילו ניתן לראות שהשימוש בשניהם יחד מצריך מאיתנו יותר מעברים על סט האימון כדי להגיע לתוצאות אליהן הגיע המודל tagger2. גם כאן כמובן שציפינו ששימוש במידע נוסף על המילה (רישא וסיפא), במיוחד כאשר מדובר בחלקי דיבר, יוביל לתוצאות טובות יותר אך אנו משערים שוב שהמבנה הבסיסי של המודל בתרגיל לא מאפשר לנו לעבד היטב את המידע הנוסף הזה.