שלום בודק יקר,

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

**מודל pos:**

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

1. **Layer size** – זהו גודל השכבה הhidden שלנו. המטרה שלנו היא שהמודל יצמצם את הבעיה בהדרגתיות עד למימד kX1, כאשר k זה מספר המחלקות, כך שנוכל לתייג את הדוגמא, כלומר לחשב את y\_hat. גודל השכבה הראשונה שלנו הוא 100.
2. **Learning rate** – מקדם הלמידה שלנו. בקוד יצרנו משתנה בשם " LEARN\_RATE" שהוא מקדם הלמידה שלנו. נזכיר את כלל העדכון:

כלומר רשת הנוירונים בכל שלב מעדכנת את המשקולות. קצב הלמידה, כלומר רמת העדכון שלנו מושפע מפרמטר זה כפי שניתן לראות בנוסחת העדכון שהצגתי לעיל. מקדם הלמידה שלנו הינו: 0.01.

1. **Epochs** – מספר האפוקים, מספר הפעמים שאנו עוברים על כל הדוגמאות בסט האימון, חשוב מצד אחד שהמודל יעבור מספיק פעמים על הדוגמאות, כך שיוכל ללמוד כמו שצריך. אך מצד שני חשוב גם שלא נעשה יותר מידיי אפוקים, כדי שלא יהיה מצב של overfitting.

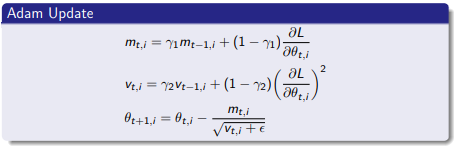
לכן, בהתאם לדרישות התרגיל ביצעתי 3 אפוקים.

1. **גודל batch** – בחרנו להגדיר גודל באטצ' 1024 עבור הtrain set , כך שהמודל יעדכן משקולות לאחר מעבר על 1024 דוגמאות וחישוב הממוצע שלהם לצורך העדכון.
2. **פונקציית אקטיבציה** – אנו רוצים להכניס אלמנט של אי-לינאריות למודל, לפיכך במעבר בין שכבות אנו משתמשים בפונקציות לא לינאריות. כפי שהוצג בהרצאה יש מספר אופציות לפונקציות לא לינאריות שיכולות לשמש אותנו: סיגמואיד, tanh, relu ועוד.

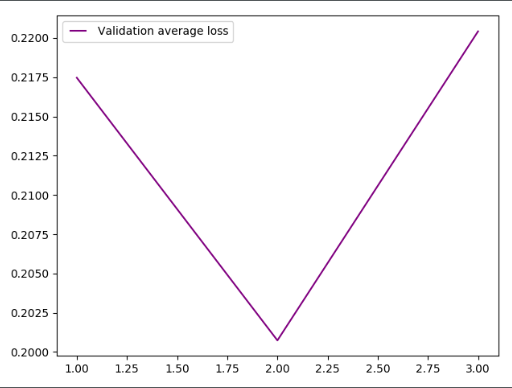
כפי שהוצג בדרישות התרגיל השתמשתי בפונקציה tanh.

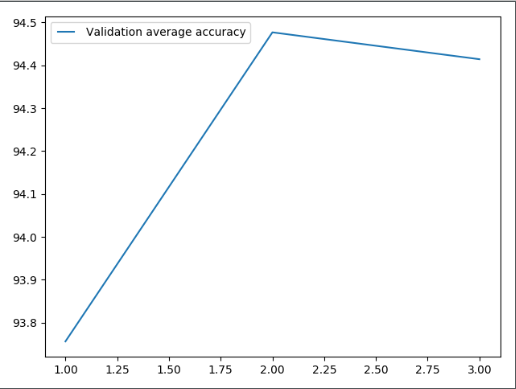
בנוסף השתמשנו באופטימייזר adam.

כפי שלמדנו בלמידת מכונה, הוא מוגדר באופן הבא:



**גרפים רלוונטים למודל זה:**





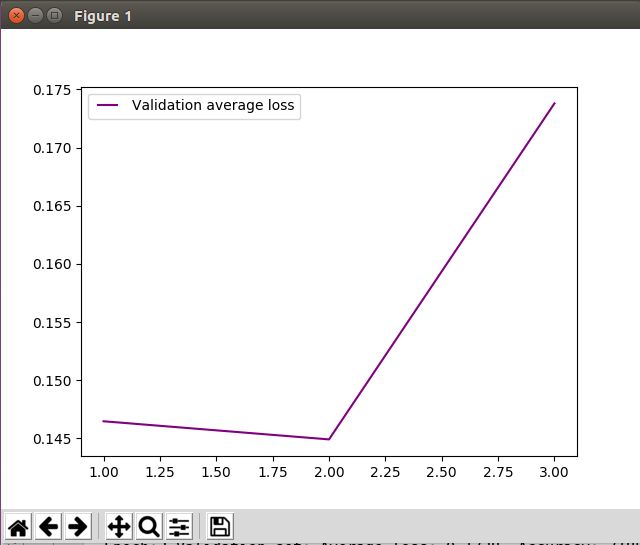
**מודל ner:**

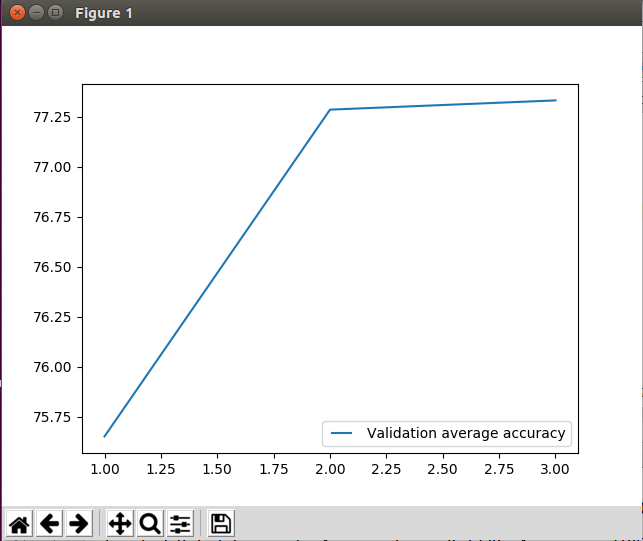
**היפר פרמטרים (רלוונטים למודל שלנו):**

נציג הפעם את הנתונים בצורת טבלה כדי להימנע מחזרות מיותרות:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Layer size** | **Learning rate** | **Epochs** | **גודל batch** | **פונקציית אקטיבציה** | **oprimizer** |
| 100 | 0.05 | 3 | 1024 | tanh | adam |

**גרפים רלוונטים למודל זה:**





**תשובות לconsiderations:**

**השאלות שהוצגו בתרגיל:**

1. **איך התמודדנו עם מילים חדשות שלא ראינו בשלב הtrain?**
2. **איך יצרנו חלון עבור המילים הראשונות והאחרונות של כל משפט?**

**תשובות:**

1. מילים שלא ראינו בשלב הtrain מופו למילה unk, שמייצגת את המילים הלא ידועות, כלומר מילים שלא מוכרות למודל שלנו. כל המילים הנ"ל ממופים לאותו וקטור במטריצת הword embedding של המודל.
2. נדרשנו בתרגיל לייצר חלון באורך חמש עבור כל מילה (שתי מילים שלפני המילה ושתי מילים שאחרי המילה). עבור שתי המילים הראשונות ושתי המילים האחרונות במשפט, כמובן שיהיה בעייתי לייצר חלון בגודל 5. לפיכך הוספנו לתחילת כל משפט שתי מילים של start, ולסוף כל משפט הוספנו שתי מילים של end כך שנוכל ליצור את החלונות. בנוסף, הוספנו את המילים start וכן end לרשימת המילים שלנו, והמודל אתחל עבורם word embeddings מתאימים.