

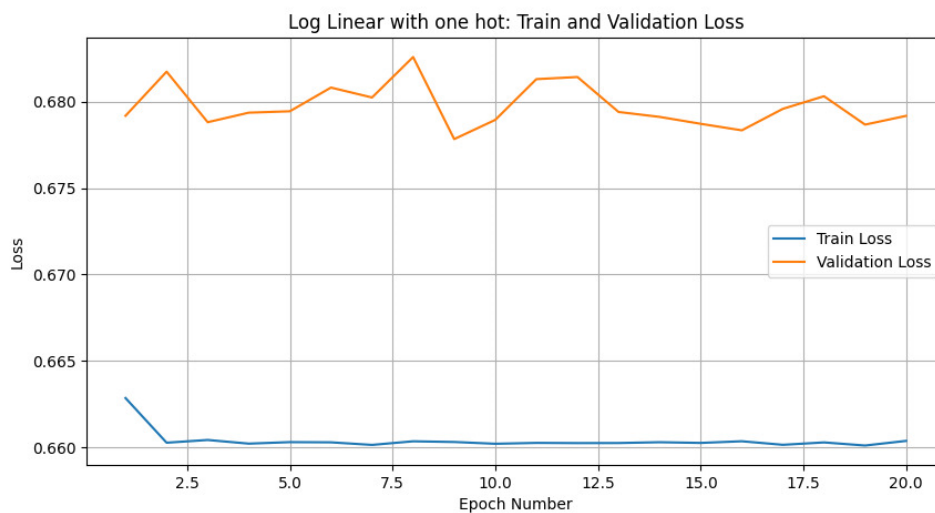
עיבוד שפה טבעית – תרגיל 3

אליה חסון 208845032, שיר שבח 322407701

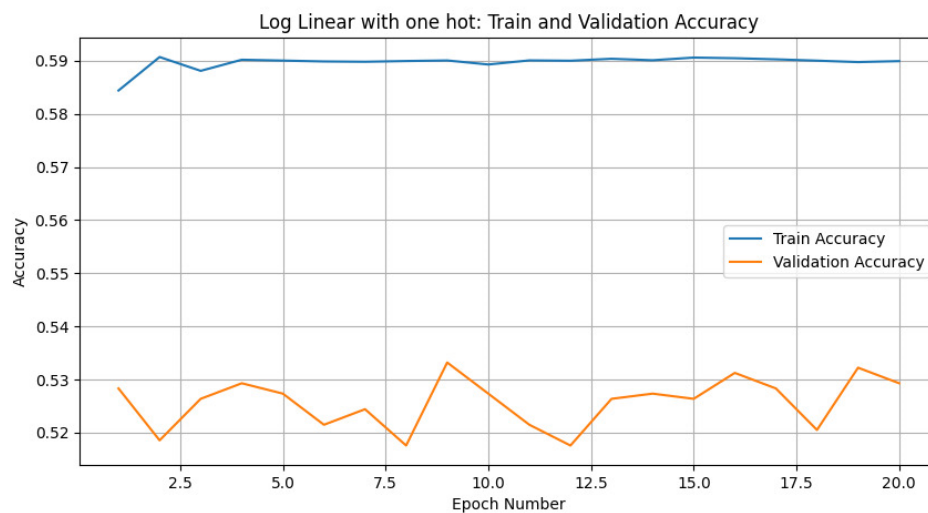
תוצאות המודלים השונים:

מודל one hot log-linear:

גרף ה-loss עבור train ו-validation:



גרף ה-accuracy עבור train ו-validation:

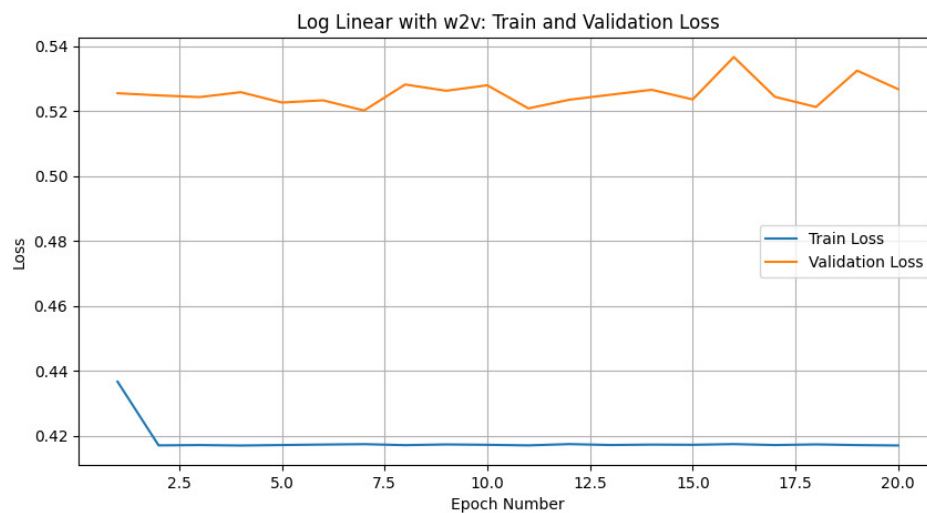


ותוצאות ההרצה עבור test set:

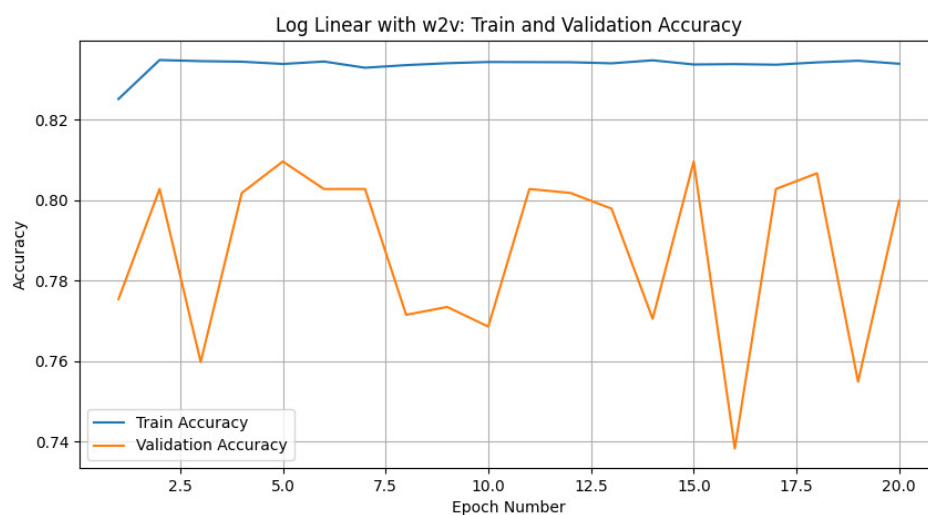
```
Test set loss: 0.6734655191394268
Test set accuracy: 0.5634765625
Test negate set loss: 0.6966124810278416
Test negate set accuracy: 0.4838709533214569
Test rare set loss: 0.7298804162442685
Test rare set accuracy: 0.2800000011920929
```

מודל word2Vec log-linear:

גרף ה-loss עבור train ו-validation:



גרף ה-accuracy עבור train ו-validation:

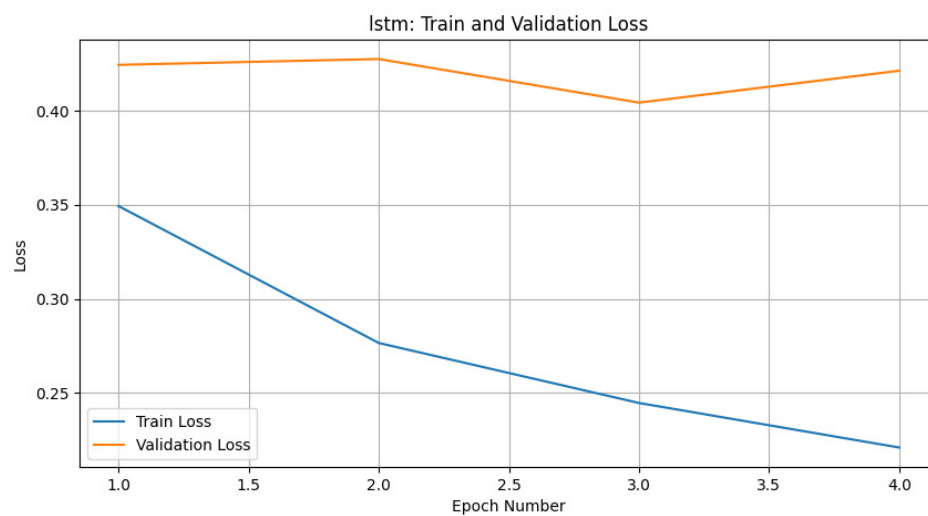


ותוצאות ההרצה עבור test set:

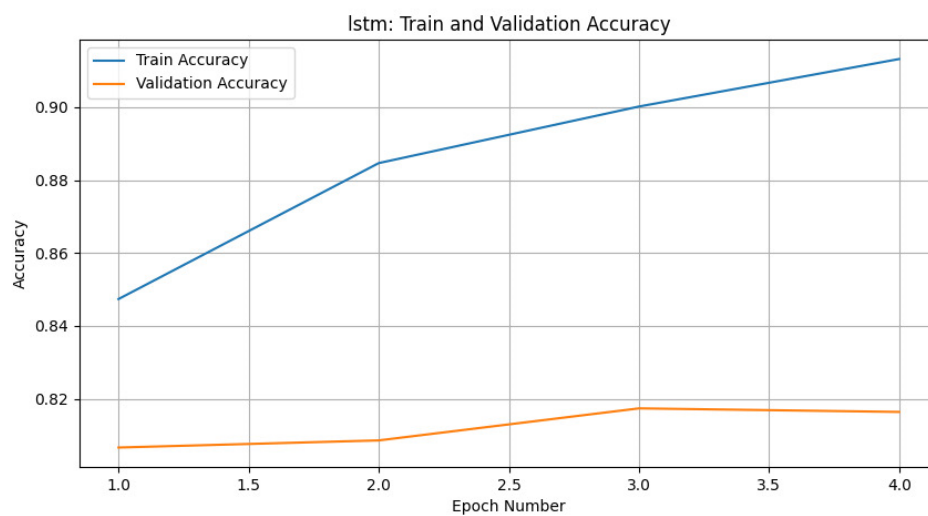
```
Test set loss: 0.4898636393841116
Test set accuracy: 0.828125
Test negate set loss: 0.730816989085607
Test negate set accuracy: 0.5645161271095276
Test rare set loss: 0.633424941636622
Test rare set accuracy: 0.7200000286102295
```

מודל LSTM:

גרף ה-loss עבור train ו-validation:



גרף ה-accuracy עבור train ו-validation:



ותוצאות ההרצה עבור test set:

```
Test set loss: 0.30682598891775115
Test set accuracy: 0.873046875
Test negate set loss: 0.8032960572520331
Test negate set accuracy: 0.7096773982048035
Test rare set loss: 0.4870263536716811
Test rare set accuracy: 0.8399999737739563
```

השוואות בין המודלים השונים:

1. נשווה בין התוצאות שקיבלנו במודל הראשון לבין התוצאות שקיבלנו במודל השני. במודל הראשון ניתן לראות כי ערכי ה-accuracy של ה-validation נעים בין 0.52-0.53 וערך ה-accuracy של ה-test הוא ~ 0.56 . בהשוואה למודל השני שבו ניתן לראות כי ערכי ה-accuracy של ה-validation נעים בין 0.74-0.81 וערך ה-accuracy של ה-test הוא ~ 0.83 . ניתן להסיק שמודל word2Vec log-linear משיג תוצאות טובות יותר ממודל ה-one-hot log-linear. שוני זה יכול לנבוע מין העובדה כי במודל W2V מילים דומות מיוצגות על ידי ווקטורים קרובים זה לזה. כלומר, המודל לוכד קשרים סמנטיים בין מילים. מידע סמנטי זה אינו נלכד בקידוד על ידי ווקטור ה-one-hot כאשר כל מילה מיוצגת כווקטור ללא קשר מובנה בין מילים דומות. לכן, ניתן להסיק כי הלמידה תהיה אפקטיבית יותר כאשר קידוד המילים שלנו נותן יותר מידע עבור כל מילה והקשרה למילים אחרות.
2. נשווה בין תוצאות שני המודלים הראשונים לתוצאות של מודל LSTM. מודל ה-LSTM מקבל ערכי accuracy של ה-validation הנעים בין 0.8-0.82 וערך ה-accuracy של ה-test הוא ~ 0.87 . ניתן לראות שוב כי בדומה לסעיף הקודם, גם W2V log linear וגם LSTM משיגים תוצאות טובות יותר ממודל ה-one-hot log linear encoder לפי שוני הקידוד המילים בין המודלים. בין מודל W2V למודל LSTM קיים שוני, אולם קטן אבל ניתן לראות בבירור כי LSTM משיג תוצאות מעט טובות יותר. שוני זה יכול לנבוע מאופן בניית רשת הניורונים של מודל זה. במודל LSTM בניגוד למודל W2V, המודל לומד בצורה עמוקה – עם שכבות נסתרות ועם למידה שהיא bi-directional. באמצעות למידה זו נוכל לתפוס את הקשר המילה משני צידי המשפט. בניגוד למודל W2V שבו יש שכבה לינארית אחת, וללא שכבות נסתרות.
3. נשווה בין תוצאות כל המודלים עבור תתי הסטים המיוחדים: עבור תת-הסט המכיל את כל המשפטים שבהם רוב המשפט מכיל מילים בהקשר אחד (חיובי/שלילי), למרות שהמשפט כולו הוא בהקשר ההפוך, מודל ה-LSTM קיבל את מידת הדיוק הגבוהה ביותר - ~ 0.71 . בנוסף, ניתן לראות כי מודל W2V מקבל מידת דיוק נמוכה יותר (~ 0.56) ממידת הדיוק

של מודל ה-LSTM, בניגוד להבדל הקטן שהיה בין שני המודלים עבור הרצת test set הכללי. כאן ניתן להסיק כי ההבדל נובע משום שעבור משפטים מיוחדים כמו בתת סט הנוכחי, מודל שהולך "הלוך ושוב" על המשפט, יכול להסיק את הקשר המשפט בצורה טובה יותר, כאשר במקביל הוא בוחן את המשפט גם מתחילתו וגם מסופו, מאשר מודל W2V. כמו כן, במודל One-hot משיג רמת דיוק נמוכה של ~ 0.48 , בהתאם לתוצאות הסעיפים הקודמים.

עבור תת-הסט המכיל את כל המשפטים המכילים מילים נדירות או שלא נראו בtrain-set, גם כאן מודל ה-LSTM מקבל את רמת הדיוק הגבוהה ביותר (~ 0.84) ומודל W2V נמוך ממנו אך קרוב (~ 0.72) ואינו קרוב בכלל (בניגוד לתת הסט הקודם) למודל one-hot שמקבל רמת דיוק נמוכה ביותר (~ 0.28). פה ניתן לראות בבירור את היתרון המשמעותי של קידוד ווקטור על ידי word embedding. זאת משום, שאם מילה נדירה דומה סמנטית למילה אחרת שכן מופיעה הרבה פעמים בטקסט, הן יקבלו ווקטור המיוצג באופן קרוב במרחב. ועל כן, למרות שהיא לא מופיעה הרבה בtrain-set, יש לנו עדיין אינדיקציה גבוהה על המילה, ועל כן המודלים יחזו את הקשרה בצורה טובה. בניגוד למודל one-hot שבו אין הקשר בין המילים, ולכן אם המילים נדירות או שלא הופיעו בtrain-set (ונתנו לה את ווקטור האפס) למודל אין מספיק נתונים כדי לחזות את הקשר המשפט.