**פרויקט סיום – Deep Learning**

**שמות המגישים:**

משה

מוריה דסה 205412786

דניאל לוי 204001291

**רשת מספר 1: CIFAR CNN**

**שאלה מספר 1:**

**פירוט שכבות הרשת** (עם ובלי פרמטרים) **:**

1. **שכבת קונבולוציה** **-** בעלת 32 קרנלים, גודל כל קרנל 3X3, הקלטים מהשכבה הראשונה בממד 3. בתוספת 32 bias אחד לכל קרנל. סך כל הפרמטרים בשכבה זו : 896.
2. **שכבת RELU** **-** פונקציית אקטיבציה שמטרתה לאפס ערכים שליליים ולהחזיר ערכים חיוביים כמו שהם.
3. **שכבת קונבולוציה** - בעלת 32 ערכים בווקטור מיפוי המידע (feature map). ערך אחד בעבור כל קרנל מתוצאת השכבה הקודמת, 32 קרנלים בגודל 3X3, בתוספת 32 bias אחד לכל קרנל. סך כל הפרמטרים בשכבה זו : 9248.
4. **שכבת אקטיבציה - RELU**.
5. **שכבת MaxPooling -** הפרמטרים אינם נלמדים ,מדובר בהורדת רזולוציה מרחבית. ברשת השתמשו בפי שתיים בכל ציר, כלומר בוחרים בפיקסל בעל הערך הגבוה אחד מתוך כל ארבעה.
6. **שכבת Dropout -** למניעת overfitting נבחרים רנדומלית פרמטרים שייעלמו ונתעלם מהם, בשכבה זו נעלים 0.25 אחוז ברשת שלנו.
7. **שכבת קונבולוציה -** בעלת 32 ערכים בווקטור מיפוי המידע (feature map). ערך אחד בעבור כל קרנל מתוצאת השכבה הקודמת, 64 קרנלים בגודל 3X3, בתוספת 64 bias אחד לכל קרנל. סך כל הפרמטרים בשכבה זו : 18496.
8. **שכבת אקטיבציה - RELU**
9. **שכבת קונבולוציה -** בעלת 64 ערכים בווקטור מיפוי המידע (feature map). ערך אחד בעבור כל קרנל מתוצאת השכבה הקודמת, 64 קרנלים בגודל 3X3, בתוספת 64 bias אחד לכל קרנל. סך כל הפרמטרים בשכבה זו : 36928.
10. **שכבת אקטיבציה - RELU**.
11. **שכבת MaxPooling** **-** ברשת השתמשו בפי שתיים בכל ציר, כלומר בוחרים בפיקסל בעל הערך הגבוה אחד מתוך כל ארבעה.
12. **שכבת Dropout -** בשכבה זו נעלים 0.25 אחוז ברשת שלנו.
13. **שכבת Flatten -** שתפקידה להעביר לווקטור (חד ממדי),גודל הווקטור 2304 (6X6X64) ובנוסף bias אחד.
14. **שכבת dense fully connected -** אשר מקבלת 2305 קלטים ו 512 נוירונים בשכבה. סך כל הפרמטרים בשכבה זו : 1180160.
15. **שכבת אקטיבציה -  RELU**.
16. **שכבת Dropout -** בשכבה זו נעלים 0.5 אחוז מרשת הנוירונים.
17. **שכבת dense fully connected -** אשר מקבלת 512 נוירונים משכבה קודמת  + bias אחד, נכפול אותם בעשרה נוירונים של שכבת היציאה (10 classים אפשריים). סך כל הפרמטרים בשכבה זו : 5130.
18. **שכבת Softmax -** המייצרת וקטור המייצג הסתברויות בין 0-1 בעבור כל קלט נקבל ערך הסתברותי .

**סך כל הפרמטרים ברשת זו:**

1,250,858

**פירוט נוסף :**

לרשת שלנו פונקציית loss מסוג categorial cross entropy , האופטימיזציה בה אנו משתמשים הינה RMSprop אשר מחלקת את שיעור הלמידה על ידי ממוצע של gradients בריבוע,

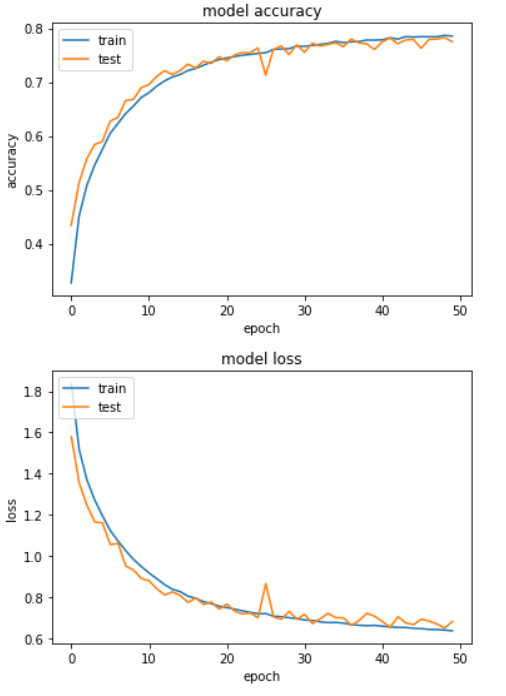
Batch size 32 - חלוקת המידע לסט קבוצות של 32 דוגמאות בכל קבוצה,

מספר epochs 50 בעצם במשך 50 הרצות אנחנו לומדים את הדוגמאות.

**תוצאות הביצועים המושגים:**

Test loss: 0.680613421535492

Test accuracy: 0.7753

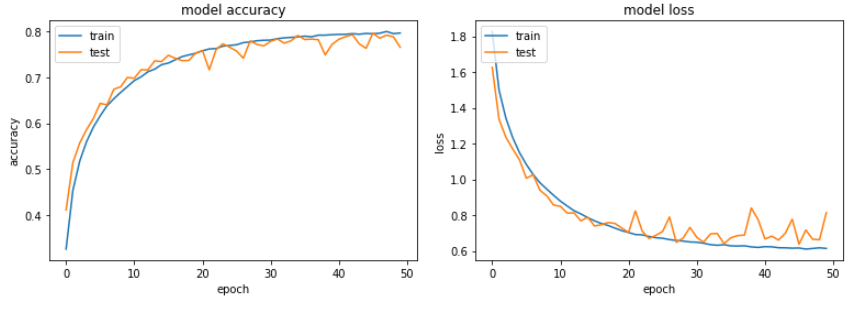


**שאלה מספר 2:**

**סעיף מספר 1:**

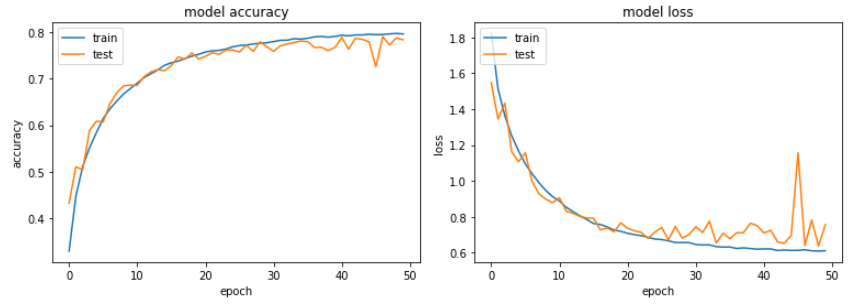
1. הוספת שכבה אחת לרשת: הוספנו שכבת קונבולוציה דו מימדית עם 64 קרנלים בגודל 3X3 ושכבת אקטיבציה .RELU

הוספת שכבה לרשת יכולה לעזור במשיכת פיצ'רים רבים ומובהקים יותר, שיקלו על אימון הרשת. חידוד הפיצ'רים לבסוף גרם אצלנו ככל הנראה להיווצרות של Over-Fitting: במהלך האימון פספסנו פרמטרים רלוונטיים והתמקדנו בפיקסלים ספציפיים מדיי.



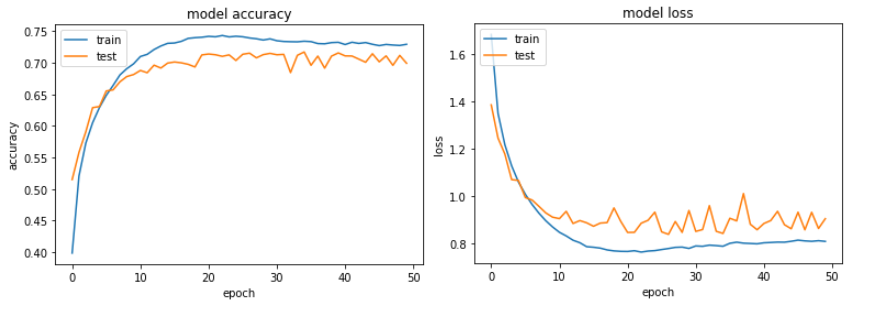
1. הוספת שתי שכבות לרשת: הוספנו שכבת קונבולוציה דו מימדית עם 32 קרנלים בגודל 3X3 ושכבת אקטיבציה RELU, לאחר מכן הוספנו שכבת קונבולוציה דו מימדית עם 64 קרנלים בגודל 3X3 ושכבת אקטיבציה RELU.

הרעיון בהוספת שתי שכבות לרשת זהה לזה של הוספת שכבה אחת. כאן נקבל פיצ'רים אף יותר מובהקים וזה יכול להועיל אך גם להזיק ולגרום ל Over-Fitting. אצלנו הוספת שכבה שניה אמנם שיפר את התוצאות בהשוואה להוספת שכבה אחת, אך גם התוצאות הללו לא מהוות שיפור מהרשת הנוכחית

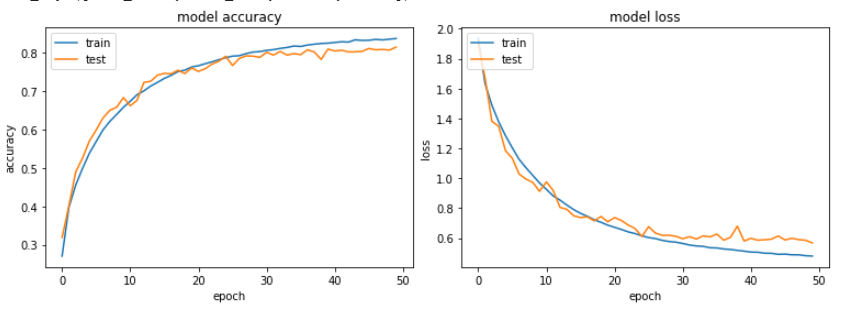


1. השינויים שביצענו:

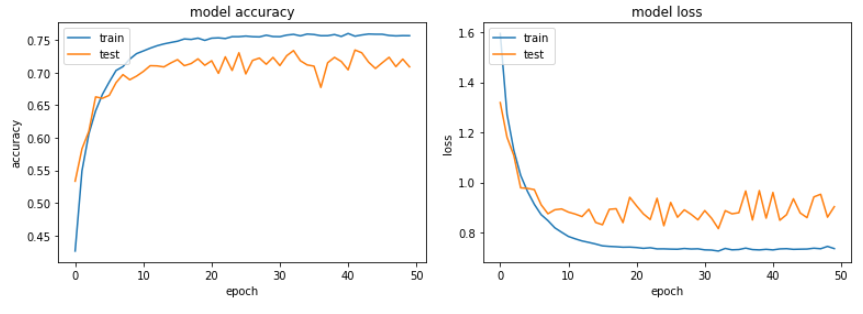
* הסרת מספר שכבות מהרשת: הסרנו את שכבות הקונבולוציה הדו ממדיות עם 64 קרנלים בגודל 3X3 ושכבות האקטיבציה RELU שאחריהן, וכן את שכבת ה Max Pooling וה Dropout. זאת על מנת לבחון איך תגיב מערכת יותר פשוטה. שינוי זה פגם בביצועים: במדד ה Loss וכן גם הוריד 10% מה Accuracy של קבוצת המבחן.



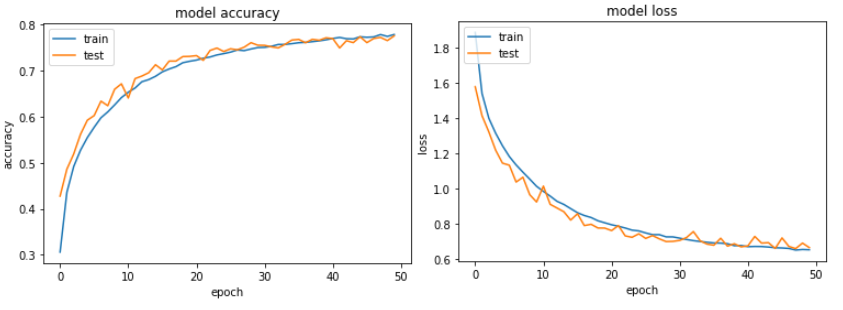
* הוספת מספר שכבות לרשת: הוספנו את שכבות הקונבולוציה הדו ממדיות עם 128 קרנלים בגודל 3X3 ושכבות האקטיבציה RELU שאחריהן, וכן את שכבת ה Max Pooling וה Dropout. תוספת זו שיפרה את הביצועים: במדד ה Loss וכן גם ה Accuracy של קבוצת המבחן והאימון.



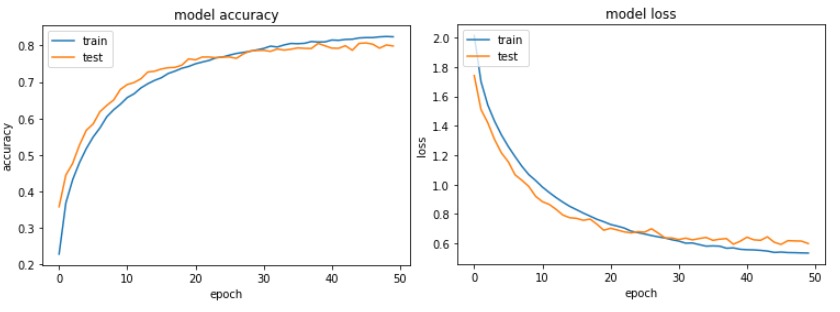
* הסרת מספר שכבות מהרשת: הסרנו את שכבות הקונבולוציה הדו ממדיות עם 32 קרנלים בגודל 3X3 ושכבות האקטיבציה RELU שאחריהן, וכן את שכבת ה Max Pooling וה Dropout. שינוי זה פגם בביצועים: במדד ה Loss ווה Accuracy של קבוצת המבחן וכך גם בקבוצת האימון.



* שינוי של מספר שכבות ברשת: הפכנו את שכבות הקונבולוציה הדו ממדיות עם 64 קרנלים בגודל 3X3 שבסקשן השני לשכבות קונבולוציה דו ממדיות עם 32 קרנלים בגודל 4X4. רצינו בכך להפחית את כמות הפרמטרים הנלמדים וכך להוריד את הסיכוי ל over fitting. שינוי זה שיפר את הביצועים במעט: במדד ה Loss וה Accuracy של קבוצת המבחן וכך גם בקבוצת האימון.



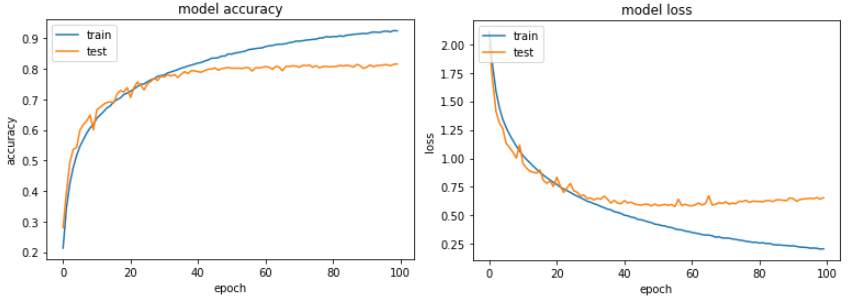
* הוספת ושינוי של מספר שכבות ברשת: הוספנו את סקשן נוסף של שכבות קונבולוציה דו ממדיות עם 64 קרנלים בגודל 4X4 ושכבות אקטיבציה RELU אחריהן, וכן שכבת Max Pooling ו Dropout. וכן שינינו את הפרמטרים בשאר הסקשנים הקודמים לקונבולוציה דו ממדיות עם 64 קרנלים בגודל 2X2 ו 3X3 לפי סדר עולה. שינוי זה שיפר את הביצועים בצורה הטובה ביותר: נראה כי רשת מסובכת יותר מביאה למשיכת פיצ'רים רבים ו\או מדויקים יותר, ובכך אנו מקלים על אימון הרשת.



**סעיף מספר 2:**

1. אלגוריתם SGD:

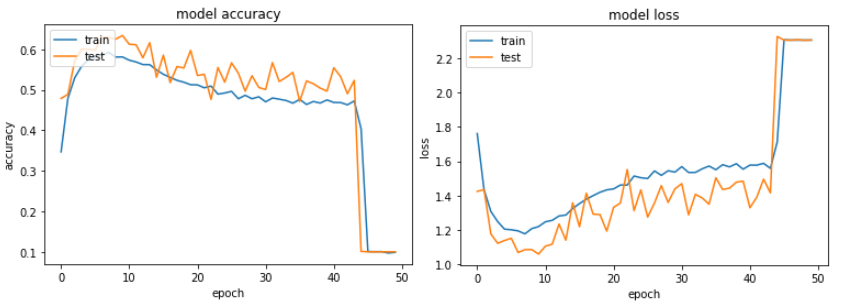
* הרצת האלגוריתם SGD - חל שיפור קטן מאוד בתוצאות ה Loss וה Accuracy.
* הרצת האלגוריתם SGD עם learning rate 0.1 - פגם בתוצאות ה Loss וה Accuracy של האימון.
* הרצת האלגוריתם SGD עם מספר epochs = 25 - חל שיפור קטן מאוד בתוצאות ה Accuracy של קבוצת המבחן.
* הרצת האלגוריתם SGD עם מספר epochs = 100 - שינוי זה הביא את השיפור המשמעותי ביותר לקבוצת המבחן, של כ- 80% במדד ה Accuracy (וכן גם לקבוצת האימון קיים שיפור ניכר בתוצאות). \* נראה כי לאחר 40 אפוקים ויותר מדד ה Accuracy מפסיק להשתפר וכך גם תהליך הלמידה.



* הרצת האלגוריתם SGD עם מספר epochs = 30 ו learning rate 0.05 - חל שיפור כמעט בלתי מורגש בתוצאות ה Accuracy של קבוצת המבחן.

1. אלגוריתם SGD עם MOMENTUM:

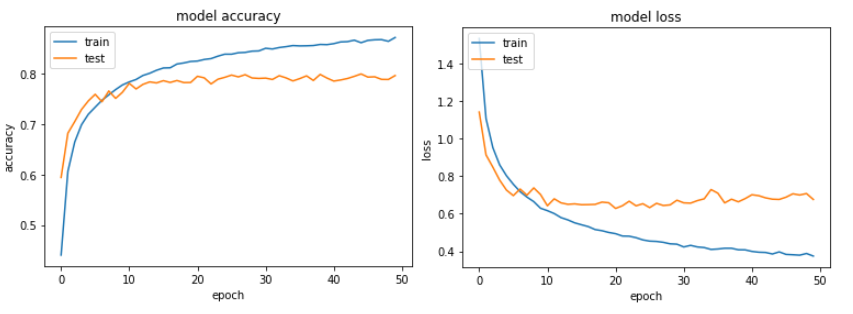
* הרצת אלגוריתם SGD עם Momentum = 0.9 - חלה ירידה קלה ביותר במדד ה Accuracy.
* הרצת אלגוריתם SGD עם Momentum = 0.9 ו Nesterov = true - כמו בסעיף הקודם, ניתן להבחין בירידה קלה במדד ה Accuracy.
* הרצת אלגוריתם SGD עם Momentum = 0.9, Nesterov = true ו learning rate = 0.02 - התוצאה הכי פחות טובה שהייתה עד כה, עם מדד Accuracy של עשרה אחוזים בלבד. ניתן לראות כי במהלך ההרצה הרשת למדה בדיוק של כ 50%, אך בסביבות האפוק ה 40 קרתה ירידה משמעותית בתוצאות.



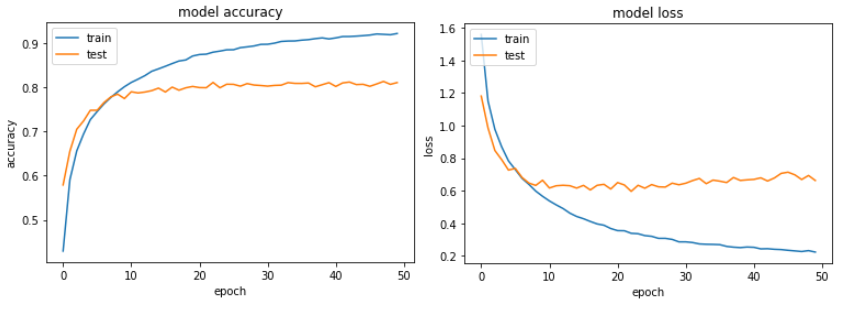
* הרצת אלגוריתם SGD עם Momentum = 0.9, Nesterov = true, learning rate = 0.02 ומספר epochs = 30 - חלה ירידה מובהקת במדד ה Accuracy.
* הרצת אלגוריתם SGD עם Momentum = 0.9, Nesterov = true ומספר epochs = 65 - חלה ירידה קלה במדד ה Accuracy.
* בוצעו כמה ניסיונות נוספים להרצת האלגוריתם עם learning rate גדול מ 0.05, אך התוצאות היו קטסטרופליות, והביאו אותנו למסקנה שהגדלת קצב הלמידה לא תביא אותנו לתוצאות טובות.

1. אלגוריתם ADAM:

* הרצת אלגוריתם ADAM עם פרמטרים דיפולטיביים - הביא לשיפור קל במדד ה Accuracy שעומד על כ 80%.



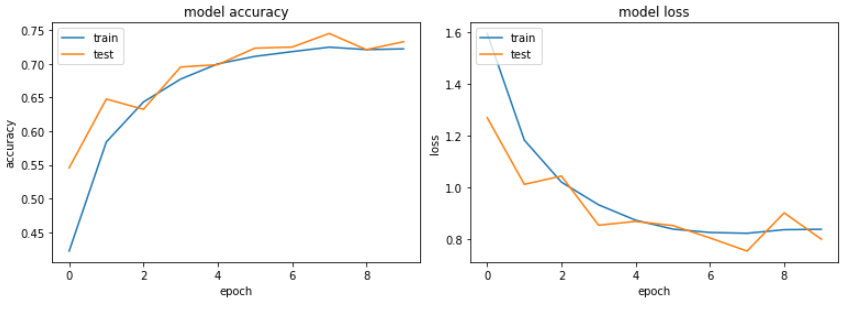
* הרצת אלגוריתם ADAM עם learning rate 0.0005 - ככל שהקטנו את קצב הלמידה, שיפרנו את מדד ה Accuracy שעומד על כ 81%.



* הרצת אלגוריתם ADAM עם מספר epochs = 75 - חל שיפור קל במדד ה Accuracy.
* הרצת אלגוריתם ADAM עם מספר epochs = 35 - חל שיפור קל מאוד במדד ה Accuracy, כמעט זהה לגרף המקורי.
* בנוסף, ניסינו להגדיל את קצב הלמידה למספרים שונים כמו 0.01 ו 0.005, אך הניסיונות לשיפור כשלו והביאו לתוצאות של כ 10% Accuracy, ומכך הסקנו שהגדלת קצב הלמידה לא תביא אותנו לתוצאות טובות.

1. אלגוריתם :RMSPROP

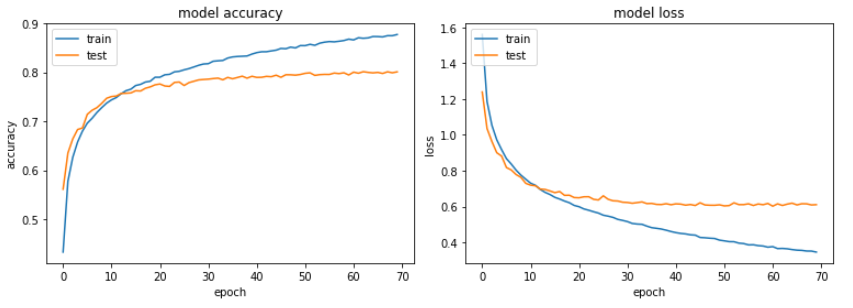
* הרצת אלגוריתם RMSPROP עם learning rate 0.001 - גרם לירידה משמעותית במדד ה Accuracy שעומד על כ 10%, למרות שעד לפני האפוק העשירי עמד על דיוק של כ 70%
* הרצת אלגוריתם RMSPROP עם learning rate 0.0005 - גרם לירידה במדד ה Accuracy שעומד על כ 50%, ניתן להבחין בירידה משמעותית בסביבות האפוק ה 15.
* הרצת אלגוריתם RMSPROP עם learning rate 0.003 ומספר epochs = 25 - גרם לירידה במדד ה Accuracy שהיה מתחת ל 50%.
* הרצת אלגוריתם RMSPROP עם learning rate 0.0005 ומספר epochs = 10 - הביא את התוצאה הטובה ביותר עבור אלגוריתם האופטימיזציה הנ"ל, עם מדד Accuracy שעומד על כ 73%. אבל זוהי תוצאה שאינה יותר טובה מהקונפיגורציה המקורית.



* בוצעו ניסיונות נוספים להפעלת האלגוריתם עם פרמטרים דיפולטיביים ומספר אפוקים גדול יותר אך התוצאות לא היו טובות. נראה כי כאשר מריצים אלגוריתם RMSPROP ככל שמעלים את כמות האפוקים, כך נראה ירידה חדה יותר ויותר בביצועי הרשת.

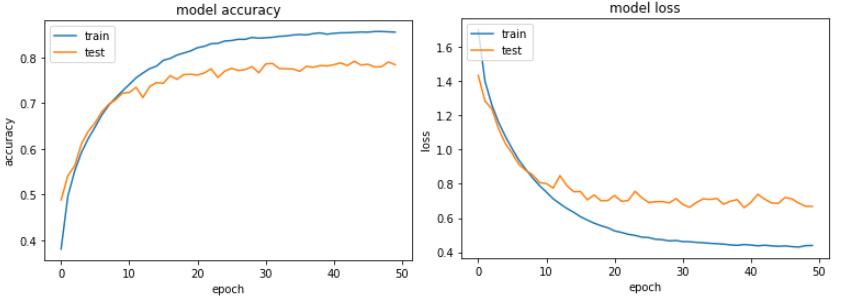
1. אלגוריתם ADAGRAD:

* הרצת אלגוריתם ADAGRAD עם פרמטרים דיפולטיביים - הביא לשיפור קל במדד ה Accuracy שעומד על כ 80% (כמו כן שיפר את תוצאות סט האימון).
* הרצת אלגוריתם ADAGRAD עם learning rate 0.005 - הביא לתוצאות כמעט זהות עם עלייה קלה במדד ה Accuracy שעומד על יותר מ 78%.
* הרצת אלגוריתם ADAGRAD עם מספר epochs = 30 - התוצאות כמעט זהות לסעיף הקודם עם מדד Accuracy שעומד על יותר מ 78%.
* הרצת אלגוריתם ADAGRAD עם מספר epochs = 70 - הביא לשיפור הגדול ביותר ב Accuracy שעומד על יותר מ 80% (כמו כן שיפר את תוצאות סט האימון).



1. Dropout:

* שינוי כל ה Dropout ברשת ל 0.5 - כנראה שלא אימנו חלק גדול מדיי מהרשת ולכן אנו רואים ירידה קלה בביצועים.
* שינוי כל ה Dropout ברשת ל 0.65 – כעת חלה ירידה משמעותית בביצועים, שכן לא אימנו חלק גדול אף יותר מהרשת. מדד ה Accuracy עומד על כ 78%.



* שינוי ה Dropout בשתי שכבות הקונבולוציה הראשונות ל 0.1 ובשכבה השלישית ל 0.25 – חל שיפור קל במדד ה Accuracy.
* בסך הכל, נראה כי Dropout קטן יותר, ובשל כך אימון של חלקים גדולים יותר ברשת, גורם לביצועים טובים יותר. יש לשים לב לא להתעלם לחלוטין מה Dropout, כיוון שהוא קיים במטרה למנוע overfitting.

\*\* כאשר אנו מביטים בעיקר על הגרפים של RMSPROP ו SGD עם MOMENTUM, ניתן לראות בברור שלאחר מספר מסויים של אפוקים ביצועי הרשת מדרדרים באופן משמעותי.

**רשת מספר 2: IMDB CNN**

**שאלה מספר 1:**

**פירוט שכבות הרשת** (עם ובלי פרמטרים) **:**

1. **שכבת Embedding -** עבור 10,000 מילים, נכפול בווקטור בגודל 300 (עם 1600 ביקורות). סך כל הפרמטרים בשכבה זו : 3000000.
2. **שכבת קונבולוציה -** בעלת 64 קרנלים, נכפול בגודל כל קרנל 3 (ממד 1), נכפול בעומק 300 (ל 1600 הווקטורים) שהם קלטים משכבת הEMBEDDING בתוספת 64 bias אחד לכל קרנל. סך כל הפרמטרים בשכבה זו : 57664.
3. **שכבת קונבולוציה -** בעלת ממד כניסה בגודל 64, נכפול ב 32 קרנלים, נכפול בגודל כל קרנל 3 (בממד 1) בתוספת 32 bias אחד בעבור כל קרנל. סך כל הפרמטרים בשכבה זו : 6176.
4. **שכבת קונבולוציה -** בעלת ממד כניסה בגודל 64, נכפול ב 16 קרנלים, נכפול בגודל כל קרנל 3 (בממד 1) בתוספת 16 bias אחד בעבור כל קרנל . סך כל הפרמטרים בשכבה זו : 1552.
5. **שכבת Flatten -** שתפקידה להעביר לווקטור (חד ממדי) את הפרמטרים,גודל הווקטור 25600 (16X1600) ובנוסף bias אחד.
6. **שכבת Dropout -** למניעת overfitting נבחרים רנדומלית פרמטרים שייעלמו ונתעלם מהם, בשכבה זו נעלים 0.2 אחוז ברשת שלנו.
7. **שכבת dense -** הקלט לשכבה זו הוא 25600 פרמטרים, נכפול ב 180 נוירונים של השכבה הנוכחית ונוסיף 180 bias. סך כל הפרמטרים בשכבה זו : 4608180.
8. **שכבת Dropout -** בשכבה זו נעלים 0.2 אחוז ברשת שלנו.
9. **שכבת dense** הקלט לשכבה זו הוא 180, נכפול בנוירון אחד בממד היציאה ונוסיף 1 עבור ה-bias. סך כל הפרמטרים בשכבה זו : 181.

**סך כל הפרמטרים ברשת זו :**

7,673,753

**פירוט נוסף :**

Batch size = 64

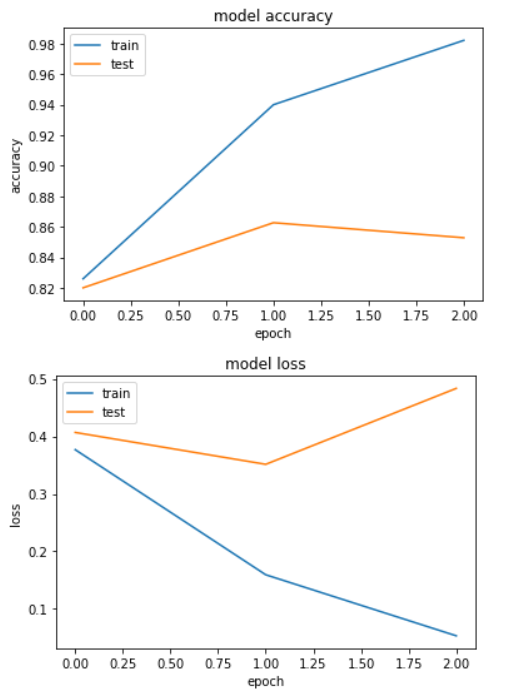
Epochs = 3

Loss function = Binary Cross Entropy

Optimizer = Adam

**תוצאות הביצועים המושגים :**

test loss, test acc: [0.48332326229095457, 0.85288]



**שאלה מספר 2:**

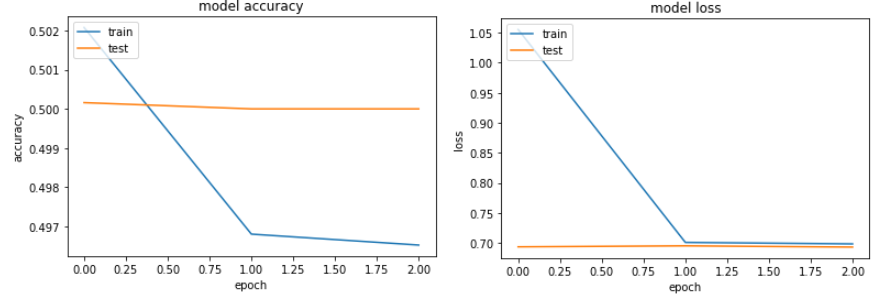
**סעיף מספר 1:**

1. הוספת שכבה אחת לרשת:

ניסיון ראשון - הוספנו שכבת קונבולוציה חד מימדית עם 100 קרנלים בגודל 3.

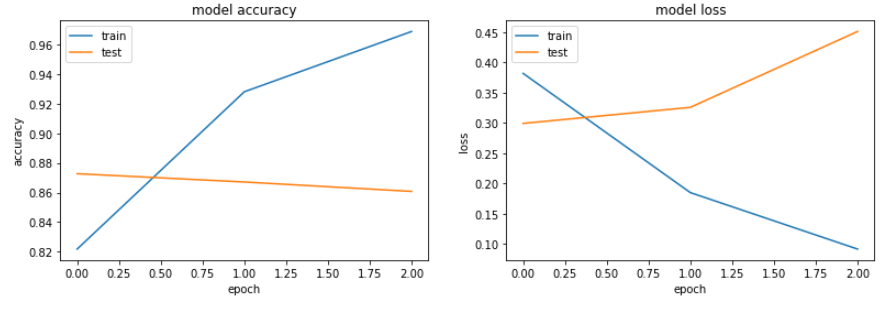
ניסיון שני - הוספנו שכבת Max Pooling עם pool size = 20 ו strides = 10.

הוספת שכבה לרשת יכולה לעזור במשיכת פיצ'רים רבים ומובהקים יותר, שיקלו על אימון הרשת. אך בשני הניסיונות חידוד הפיצ'רים לבסוף גרם אצלנו ככל הנראה להיווצרות של Over-Fitting: במהלך האימון פספסנו פרמטרים רלוונטיים והתמקדנו בפיקסלים ספציפיים מדיי.



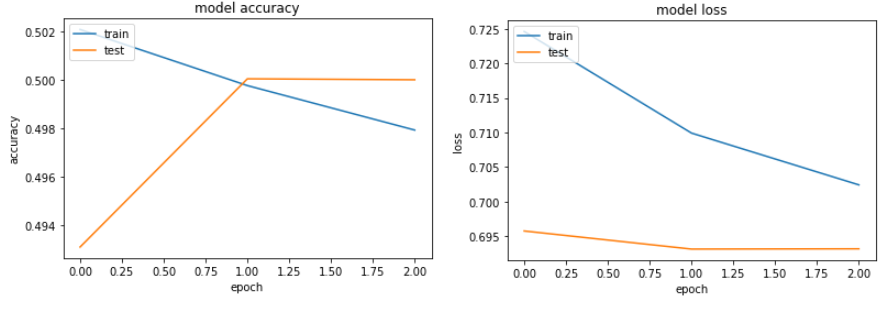
1. הוספת שתי שכבות לרשת: הוספנו שכבת קונבולוציה חד מימדית עם 8 קרנלים בגודל 3, ואחריה שכבת קונבולוציה חד מימדית נוספת עם 4 קרנלים בגודל 3.

הרעיון בהוספת שתי שכבות לרשת זהה לזה של הוספת שכבה אחת. כאן נקבל פיצ'רים אף יותר מובהקים וזה יכול להועיל אך גם להזיק ולגרום ל Over-Fitting. אצלנו הוספת שכבה שניה גרמה לתוצאות טובות יותר ונראה כי הגדלת הרשת והסיבוכיות שלה (יותר משקולות) תרמה לחידוד הפיצ'רים וללמידה טובה יותר.

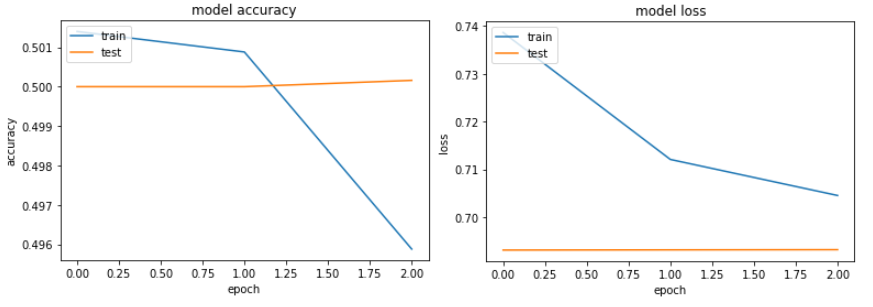


1. השינויים שביצענו:

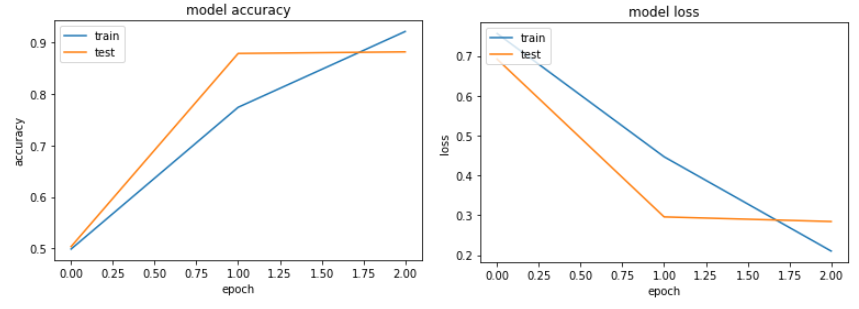
* החלפת מספר שכבות ברשת: החלפנו את כל שכבות הקונבולוציה הקיימות ב 3 שכבות חדשות, כל אחת בעלת 100 קרנלים, בגדלים 3, 4 ו5 לפי הסדר. הרעיון העומד מאחורי חישוב זה טמון במאמר שקראנו. שינוי זה פגם בביצועים: נראה לפי מדד ה Accuracy של קבוצת המבחן שעומד על 50% (כאשר הרצנו את אותו הקוד בפלטפורמה אחרת קיבלנו אחוזי דיוק גדולים מאוד שעמדו על 89%).



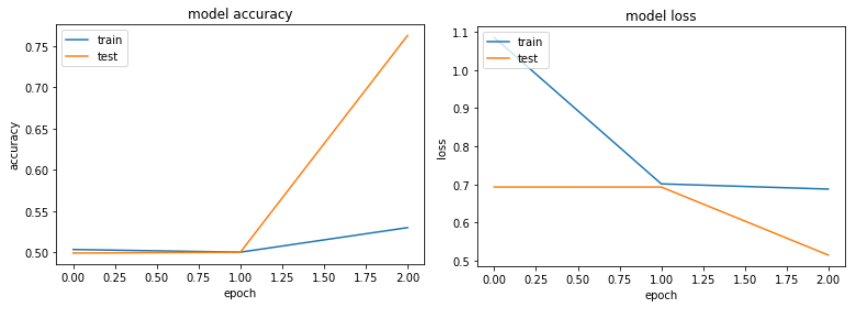
* החלפת מספר שכבות ברשת: כמו בסעיף הקודם, כאשר גדלי הקרנלים בסדר הפוך (5, 4, 3). גם שינוי זה לא תרם לביצועים ואף פגם בהם.



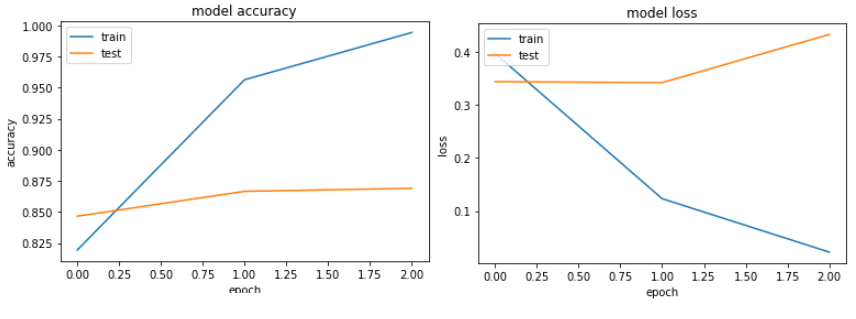
* החלפת מספר שכבות ברשת: החלפנו את כל שכבות הקונבולוציה הקיימות בשכבת קונבולוציה חד מימדית חדשה, בעלת 100 קרנלים בגודל 3. הרעיון העומד מאחורי שינוי זה הוא להוריד ולצמצם את רמת סיבוכיות הרשת (כמות המשקולות), ובכך לדייק יותר. שינוי זה גרם לביצועים טובים יותר משאר הניסיונות: נראה לפי מדד ה Accuracy של קבוצת המבחן שהשתפר מעט מהגרף המקורי, אך עדיין לא טובים יותר מאלו של הרשת המקורית.



* החלפת מספר שכבות ברשת: החלפנו את כל שכבות הקונבולוציה הקיימות בשכבת קונבולוציה חד מימדית חדשה, בעלת 200 קרנלים בגודל 3. הכפלנו את מספר הקרנלים מהניסיון הקודם. שינוי זה פגם בביצועים: למרות שאנו רואים שיפור חד מהאפוק הראשון לשני, ואולי אם היינו מבצעים מספר הרצות נוספות של כל הדוגמאות היינו מקבלים תוצאות טובות יותר.



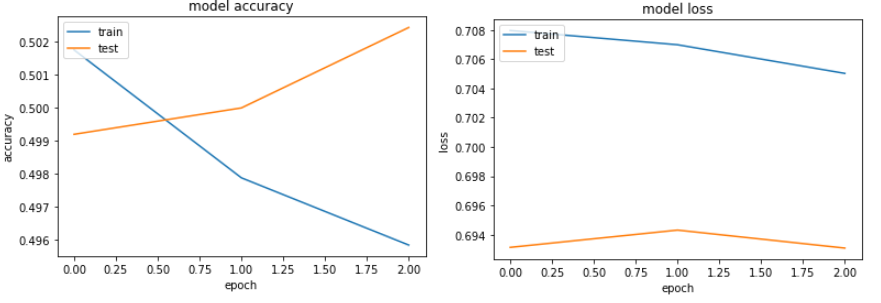
* החלפת מספר שכבות ברשת: החלפנו את כל שכבות הקונבולוציה הקיימות בשכבת קונבולוציה חד מימדית חדשה, בעלת 63 קרנלים בגודל 3. כל זאת עדיין תחת ההנחה שהורדת סיבוכיות הרשת תגרום לביצועים מדויקים יותר. שינוי זה פגם בביצועים: אמנם חל שיפור במדד ה Accuracy של סט האימון בהשוואה לרשת המקורית, אך מדד ה Accuracy של קבוצת המבחן ירד מעט וזה מרמז על Over-Fitting.



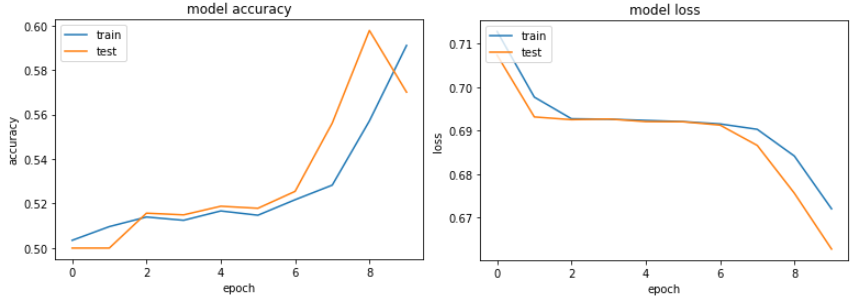
**סעיף מספר 2:**

1. אלגוריתם SGD:

* הרצת האלגוריתם SGD - גרם לירידה חדה במדד ה Accuracy בהשוואה ל ADAM וכן גם לתצאות Loss גבוהות יותר.

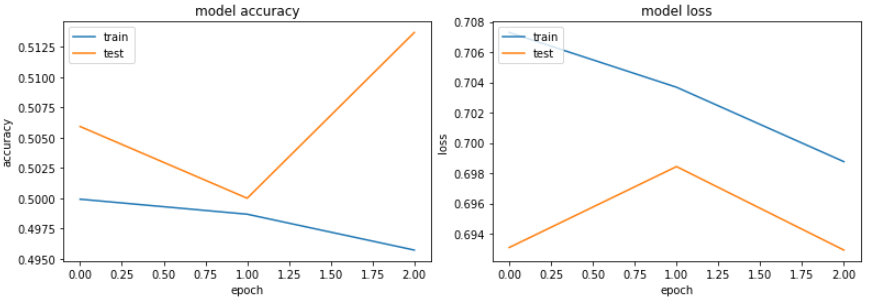


* הרצת האלגוריתם SGD עם מספר epochs = 15 – גם כאן נבחין בירידה חדה במדד ה Accuracy בהשוואה ל ADAM.
* הרצת האלגוריתם SGD עם מספר epochs = 15 ו learning rate = 0.1 - גם כאן נבחין בירידה חדה במדד ה Accuracy בהשוואה ל ADAM, ובאופן כללי נראה כי אחרי האפוק העשירי לא חל שיפור בתוצאות.
* הרצת האלגוריתם SGD עם מספר epochs = 10 ו learning rate = 0.1 - התוצאות הטובות ביותר שקיבלנו עבור אלגוריתם SGD, אך אלו גרועות בהרבה מזה של אלגוריתם ADAM.

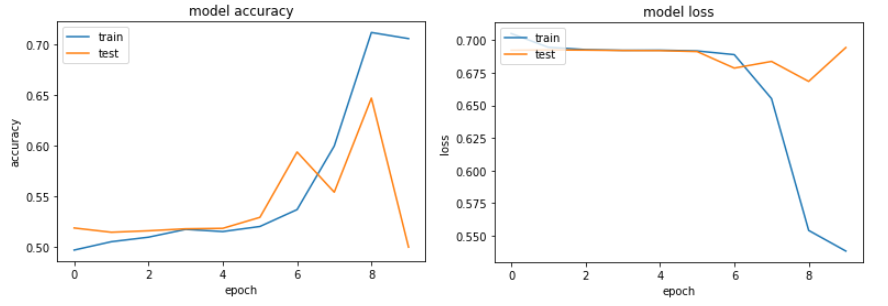


1. אלגוריתם SGD עם MOMENTUM:

* הרצת אלגוריתם SGD עם Momentum = 0.5 - תחילה בחרנו מומנטום של 0.5 כדי לראות כיצד ישפיע על הרשת. חלה ירידה מובהקת במדד ה Accuracy שעומד על כ 50%.

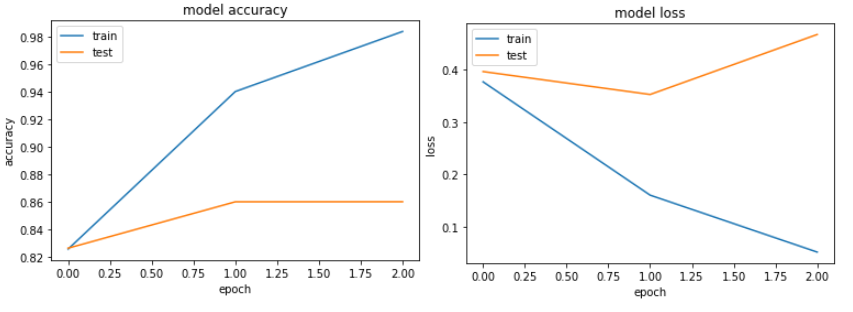


* הרצת אלגוריתם SGD עם Momentum = 0.5 ומספר epochs = 10 - חל שינוי קל ולא משמעותי בהשוואה לתוצאות הסעיף הראשון.
* הרצת אלגוריתם SGD עם Momentum = 0.5, מספר epochs = 10 ו learning rate = 0.1 – התוצאות גם כאן ללא כל שינוי משמעותי.
* הרצת אלגוריתם SGD עם Momentum = 0.5, מספר epochs = 10 ו Nesterov = true - התוצאות הטובות ביותר שקיבלנו עבור אלגוריתם זה, אך אלו גרועות בהרבה מזה של אלגוריתם ADAM.



1. אלגוריתם ADAM:

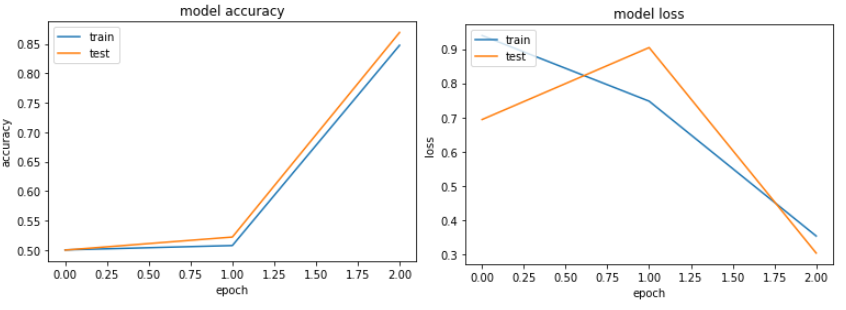
* הרצת האלגוריתם ADAM - שימוש באלגוריתם זה בוצע ברשת המקורית והביא לכ 85 אחוזי דיוק. הרצנו את האלגוריתם עם פרמטרים שהוצעו ב KerasDoc והגענו לתוצאות כמעט זהות של 86 אחוזי דיוק.



* הרצת האלגוריתם ADAM עם מספר epochs = 10 - מספר רב יותר של חזרות הביא לתוצאות זהות לסעיף הקודם במדד ה Accuracy וה Loss.
* הרצת האלגוריתם ADAM עם מספר epochs = 15 - התוצאות שוב זהות לסעיף הקודם, נראה כי לאחר 3 חזרות על תהליך הלמידה, מספר האפוקים פחות משפיע.
* הרצת האלגוריתם ADAM עם מספר epochs = 3, ו learning rate = 0.1, 0.01, 0.05 - הביא לתוצאות פחות טובות עם מדד Accuracy שעומד על 50%
* ניסינו בנוסף לשנות את גודל ה batch אך גם ניסיון זה עלה בתוהו.

1. אלגוריתם :RMSPROP

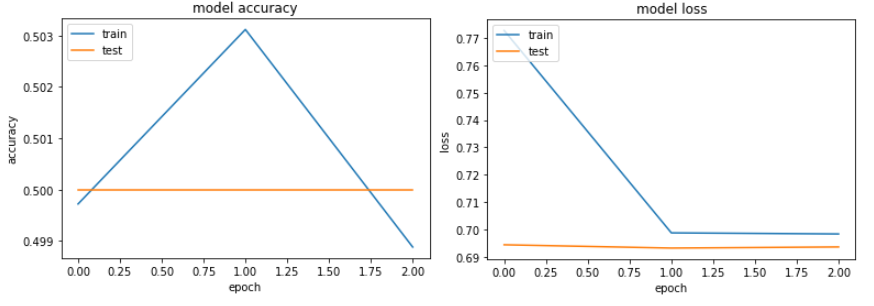
* הרצת האלגוריתם RMSPROP - עם פרמטרים דיפולטיביים שהוצעו ב KerasDoc, התוצאה שקיבלנו כאן הייתה טובה, מדד ה Accuracy עומד על כ 87% וה Loss נמוך מ 0.3.



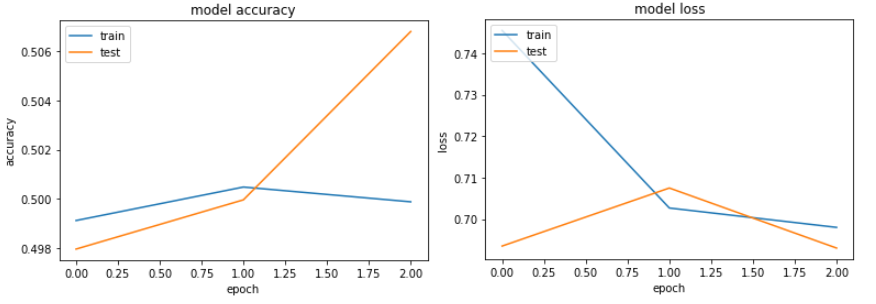
* הרצת האלגוריתם RMSPROP עם מספר epochs = 10 - מספר רב יותר של חזרות הביא לתוצאות זהות לסעיף הקודם במדד ה Accuracy וה Loss, נראה שקיים כיוון שהרשת צדקה כמעט בכל סט האימון.
* הרצת האלגוריתם RMSPROP עם מספר epochs = 15 - כמו בסעיף הקודם, מספר רב יותר של חזרות הביא לתוצאות זהות לסעיף הקודם במדד ה Accuracy וה Loss.
* הרצת האלגוריתם RMSPROP עם מספר epochs = 5 - כאן חלה אף ירידה קלה בביצועים ונראה כי באופן כללי כאשר עוברים את סביבות ה 3 אפוקים, לא חל שינוי משמעותי בביצועים.

1. אלגוריתם ADAGRAD:

* הרצת האלגוריתם ADAGRAD - עם פרמטרים דיפולטיביים שהוצעו ב KerasDoc, התוצאה כאן הייתה לא טובה עם מדד Accuracy שעומד על פחות מ 50%.



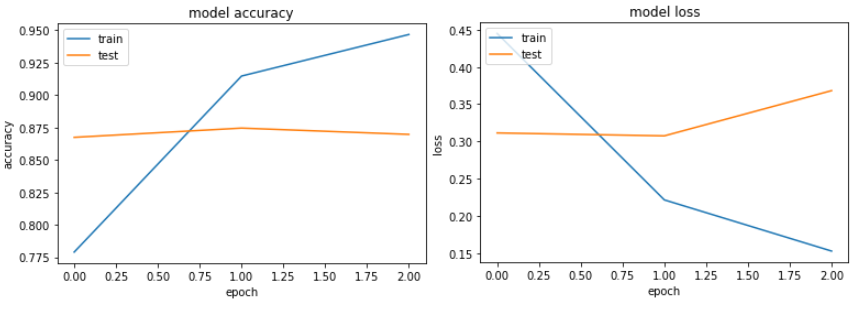
* הרצת האלגוריתם ADAGRAD עם מספר epochs = 10 - הגדלת כמות האפוקים הביאה לתוצאות זהות לסעיף הקודם.
* הרצת האלגוריתם ADAGRAD עם learning rate 0.1 – הביא לשיפור זניח במדד ה Accuracy של קבוצת המבחן.



* הרצת האלגוריתם ADAGRAD עם learning rate 0.1 – שוב אותן התוצאות של סעיפים 1 ו 2.

1. Dropout:

* שינוי ה Dropout ל 0.5 - הוביל לעלייה מינורית במדד ה Accuracy וירידה במדד ה Loss.
* שינוי ה Dropout ל 0.1 - הוביל לתוצאות מעט פחות טובות ולירידה קטנה במדד ה Accuracy.
* שינוי ה Dropout ל 0.8 - הוביל לתוצאה הכי פחות טובה בסעיף הזה, ולירידה במדד ה Accuracy שעומד על כ 50%.
* שינוי ה Dropout ל 0.6 - הביא לתוצאה זהה לסעיף הראשון, ולמדד Accuracy העומד על כמעט 87%.



\*\* ניתן לראות בברור באלגוריתם SGD בסעיף השלישי כי חלה ירידה בביצועים מהאפוק העשירי. ובאופן כללי ניתן לראות שאחרי כמות אפוקים מסוימת מפסיקים לחול שיפורים בכלל האלגוריתמים הנ"ל.

**רשת מספר 3: IMDB LSTM**

**שאלה מספר 1:**

**פירוט שכבות הרשת** (עם ובלי פרמטרים) **:**

1. **שכבת Embedding -** עבור 20,000 מילים, נכפול בוקטור בגודל 128. סך כל הפרמטרים בשכבה זו : 2560000.
2. **שכבת LSTM-** עבור 128 פרמטרים מהשכבה החבויה נוסיף 128 פרמטרים משכבת הכניסה, ואליהם נוסיף 1 bias  נכפול ב- 128 פרמטרים מה- cell state שאותם נכפול ב 4 רשתות הנוירונים אליהם הפרמטרים נכנסים. סך כל הפרמטרים בשכבה זו : 131584.
3. **שכבת Dense** - וקטור יחיד המורכב  ממד אחד, נכפול ב 128 פרמטרים משכבת היציאה  של הLSTM המועברת לשכבה זו ונוסיף 1 bias. סך כל הפרמטרים בשכבה זו :ג129.

**סך כל הפרמטרים ברשת זו:**

2,691,713

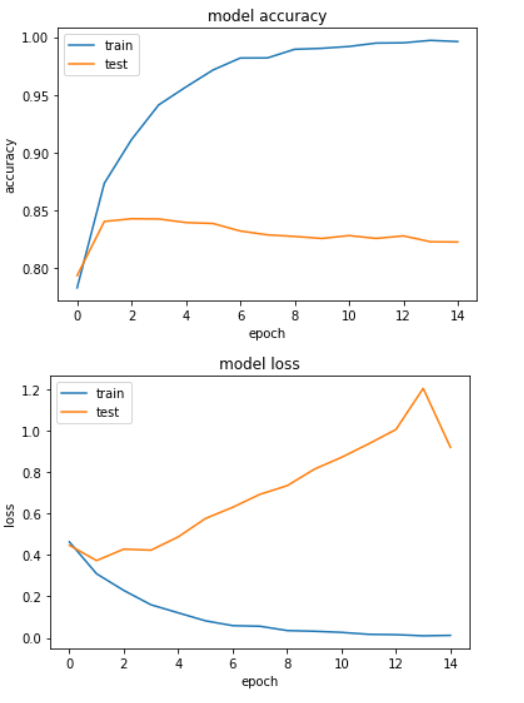
**פירוט נוסף:**

לרשת שלנו פונקציית loss מסוג binary cross entropy, האופטימיזציה בה אנו משתמשים הינה ADAM אשר מחשבת שיעורי למידה אדפטיבית עבור כל פרמטר בנוסף לאחסון ממוצע של גרדיאנטים בריבוע , batch 32 חלוקת המידע לסט קבוצות של 32 דוגמאות בכל קבוצה , מספר epochs 15 בעצם במשך 15 הרצות אנחנו לומדים את הדוגמאות .

**תוצאות הביצועים המושגים :**

Test score: 0.919022608088255

Test accuracy: 0.82264



**שאלה מספר 3:**

**סעיף ט' :**

LSTM -

הרצת האלגוריתם לראשונה עם הפרמטרים הנתונים הביאה לתוצאות הטובות ביותר מבין כל שאר האלגוריתמים, עם מדד Accuracy שעומד על כ 83%. אמנם, נראה שככל שהלמידה נמשכת ומספר האפוקים גדל כך יורד מדד ה Accuracy (דיוק הרשת במגמת ירידה)

סיכום הפרמטרים + גרף : מוצג בשאלה מס' 1.

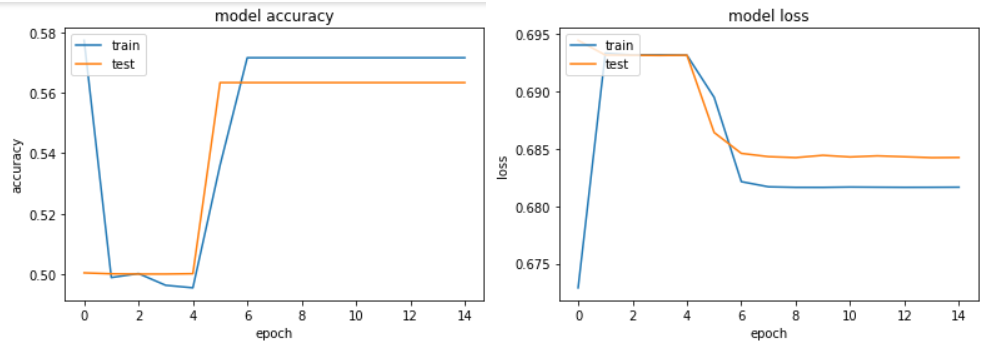
GRU -

השינוי משכבת LSTM לשכבת GRU עם הפרמטרים הנתונים גרם לירידה במדד ה Accuracy, שירד מסביבות ה 83% לסביבות ה 81%.

סיכום הפרמטרים :

1. שכבת embedding - נכפול את סך כל אוצר המילים שלנו בממד ה embedding (20000X128). סך כל הפרמטרים בשכבה זו : 2560000.
2. שכבת GRU - עבור שער היציאה בגודל 128 נכפול בעצמו, נוסיף לכך מכפלה של גודל היציאה בגודל הכניסה שגם הוא 128, אליהם נוסיף את גודל היציאה ואת כל זאת נכפיל בשלושה שערים שקיימים ב GRU 3\*(128^2+128\*128+128). סך כל הפרמטרים בשכבה זו : 98688.
3. שכבת GRU - עבור שער היציאה בגודל 1 נכפול בעצמו, נוסיף לכך מכפלה של גודל היציאה בגודל הכניסה שהוא 128, אליהם נוסיף את גודל היציאה ואת כל זאת נכפיל בשלושה שערים שקיימים ב GRU 3\*(1^2+1\*128+1). סך כל הפרמטרים בשכבה זו : 390.
4. שכבת Dense - וקטור יחיד המורכב  ממד אחד, נכפול בפרמטר אחד משכבת היציאה של ה GRU המועבר לשכבה זו ונוסיף 1 bias. סך כל הפרמטרים בשכבה זו : 2.

סך כל הפרמטרים ברשת זו : 2659080.



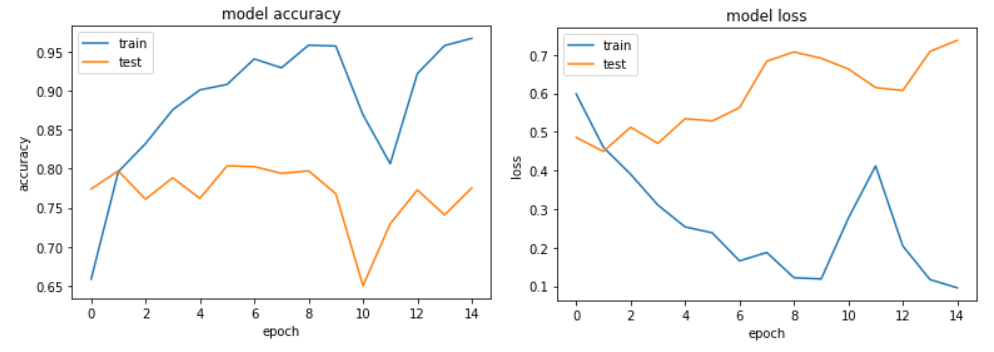
SimpleRnn -

השינוי משכבת LSTM לשכבת SimpleRnn עם הפרמטרים הנתונים הביא לירידה קצת פחות משמעותית במדד ה Accuracy, שירד מסביבות ה 83% (כמוזכר למעלה) לסביבות ה 77%.

סיכום הפרמטרים :

1. שכבת embedding - נכפול את סך כל אוצר המילים שלנו בממד ה embedding (20000X128). סך כל הפרמטרים בשכבה זו : 2560000.
2. שכבת SimpleRnn – נחבר את ממד שכבת הכניסה שהוא 128 לממד של המטריצה הפנימית שגם הוא 128, את כל זה נכפול בממד היציאה שגם הוא 128 ונוסיף בסוף 128 ביאס. סך כל הפרמטרים בשכבה זו : 32896.
3. **שכבת Dense** - וקטור יחיד המורכב  ממד אחד, נכפול ב 128 פרמטרים משכבת היציאה  של הLSTM המועברת לשכבה זו ונוסיף 1 bias. סך כל הפרמטרים בשכבה זו :ג129.

סך כל הפרמטרים ברשת זו : 2593025.

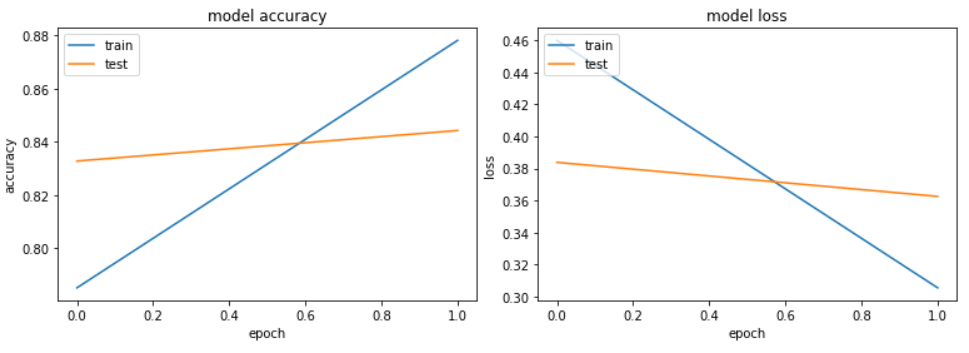


**סעיף י' :**

עבור כל אחת ואחת מהרשתות ניסינו לשנות את כמות האפוקים (להגדיל ולהקטין), על מנת לראות האם ביצועי הרשת ישתפרו. עבור כל אחת מהרשתות נראה את השיפור הטוב ביותר שהשגנו.

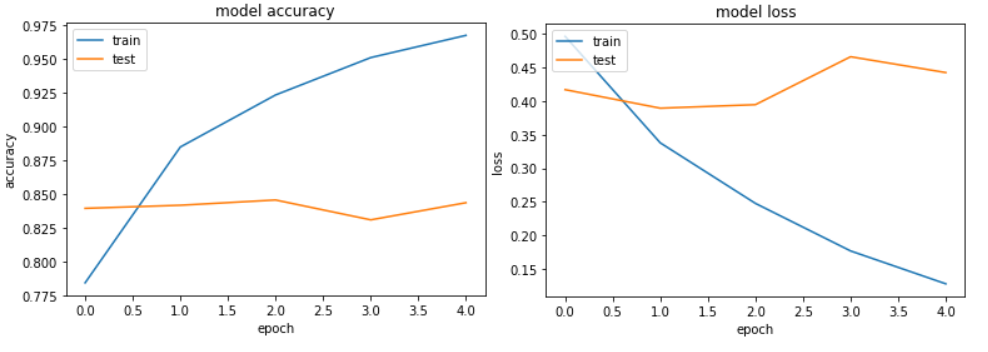
LSTM -

השינוי הטוב ביותר שהשגנו כאן היה כאשר כאשר הקטנו את כמות האפוקים ל 2, מדד ה Accuracy עלה ב 0.02%, לעומת זאת כאשר הגדלנו את הכמות הרשת חוותה ירידה קלה בביצועים



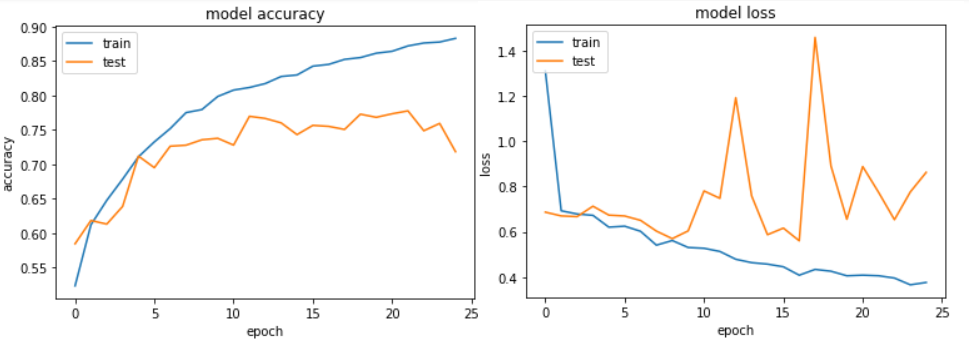
GRU -

השינוי הטוב ביותר שהשגנו כאן היה כאשר הקטנו את כמות האפוקים ל 5, מדד ה Accuracy עלה ל 84%. הושג שיפור גם כאשר העלנו את כמות האפוקים אם כי לא משמעותי, (בנוסף גם ראינו כי תהליך הלמידה בניסיון זה הלך והתדרדר ככל שעלינו בכמות האפוקים).



SimpleRnn -

לא הצלחנו להביא לשיפור במודל זה. השינוי שהביא לשינוי הכי פחות גרוע היה כאשר העלנו את מספר האפוקים ל25



מסקנות-

ניתן לראות כי מודל זה נמצא over fitting, לכן ניחשנו תחילה שככל שמספר האפוקים ילך ויגדל המודל יתאים את עצמו יותר ויותר לקבוצת האימון ויתקשה להכליל את יכולתו לפתור את בעיה זו על המידע.

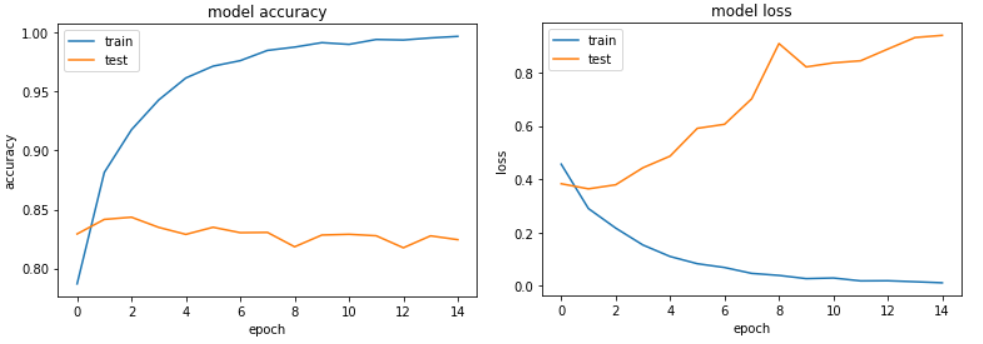
בפועל נראה כי התוצאות היו מעט שונות, אך המגמה הכללית מראה שאכן המערכת מפסיקה להשתפר ואף להתדרדר אחרי כמות מסוימת של אפוקים.

**סעיף י"א :**

המשמעות של שינוי וקטור המצב היא לשנות את מספר המכונות הפועלות, ובכך אנו קובעים כמה זיכרון אנחנו צריכים בתוך הרשת. במקרה שלנו כמות המכונות הדיפולטיבית בקוד שקיבלנו הייתה 128, עבור כל אחת מהמכונות ניסינו לשנות את כמות ה units. את כל אחת מהתוצאות השוונו להרצה המקורית, ולהלן התוצאות:

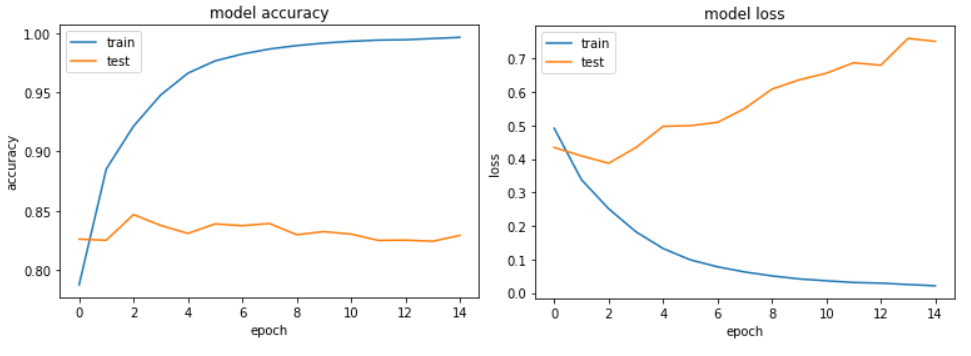
LSTM -

* כאשר הקטנו את כמות ה units ל 50 ולאחר מכן כאשר הגדלנו את כמות ה units ל 256, נראה כי הייתה ירידה קלה בביצועים, וכן הלמידה הפכה פחות יעילה לאורך האיטרציות. באופן כללי, עלייה או ירידה קיצונית בכמות ה units כנראה תביא לתוצאות פחות טובות של הרשת.
* כאשר הגדלנו את כמות ה units ל 180 וגם כאשר הקטנו את הכמות ל 70 הייתה עלייה מינורית וכמעט לא משמעותית בביצועים.



GRU -

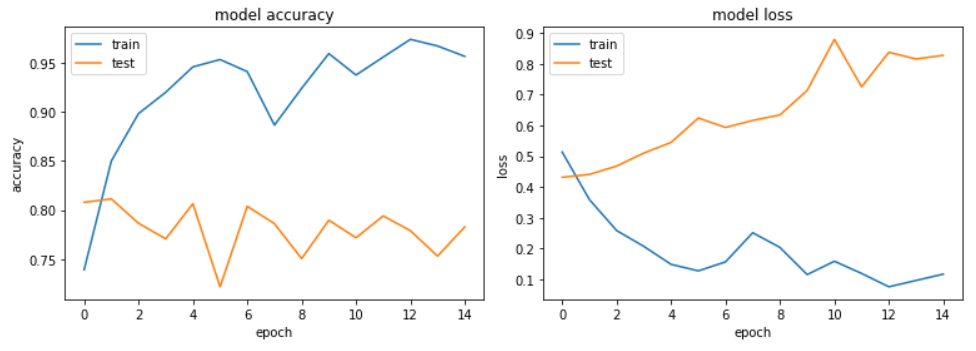
* כאשר הקטנו את כמות ה units ל 50 רואים שיפור קל בביצועים, מדד ה Accuracy עלה מ 81% ל 83%. ואלו התוצאות הטובות ביותר שקיבלנו בסעיף זה עבור ה GRU



* כאשר הקטנו את כמות ה units ל 70 נראה גם שיפור קל בביצועים, אך לא כמו בסעיף הקודם.
* ככל שהוספנו יותר units הייתה ירידה יותר ויותר משמעותית בביצועים, כנראה יותר מדיי זיכרון אחורה עשוי לגרום ל over-fitting. במצב זה נראה שבחירה בפחות מכונות תביא לתוצאות טובות יותר.

SimpleRnn -

* כאשר הורדנו את הווקטור ל 50 units ראינו ירידה מינורית וכמעט לא משמעותית בביצועי הרשת.
* כאשר הקטנו את כמות ה units ל 70, ראינו שיפור קל בביצועי הרשת, מדד ה Accuracy עלה מ 77% ל 78%. ובסך הכל אלו היו תוצאות ההרצה הטובות ביותר עבור SimpleRnn.



* ככל שהוספנו יותר units הייתה ירידה יותר ויותר משמעותית בביצועים

מסקנות -

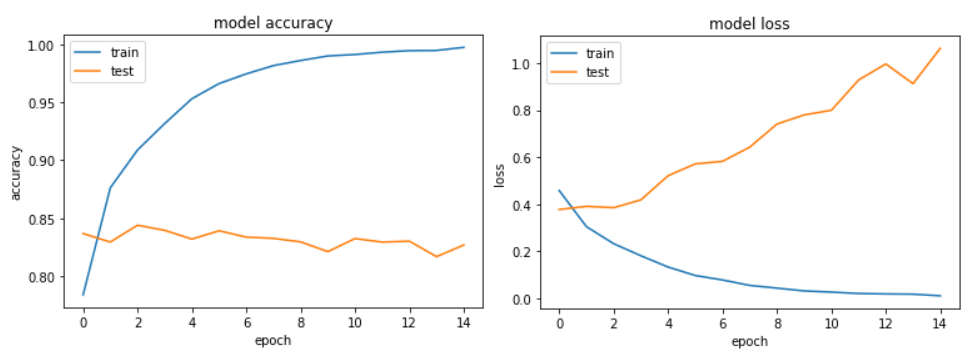
ניתן לראות כמעט בכל אחת מההרצות כי המודל נמצא במצב של overfitting וכי שינוי גודלו של וקטור המצב כמעט ואינו משפיע במצב זה. אבל בכל זאת הבחנו במגמת התדרדרות קלה כאשר העלנו את כמות היחידות, כנראה יותר מדיי זיכרון אחורה עשוי לגרום ולהגביר את ה over-fitting.

**סעיף י"ב - שינוי ממדי הכניסה:**

כאן אנו בודקים את השפעת שינוי גודל וקטור הייצוג מצורת one-hot של כל מילה בטקסט. את כל אחת מהתוצאות השוונו להרצה המקורית, ולהלן התוצאות:

LSTM -

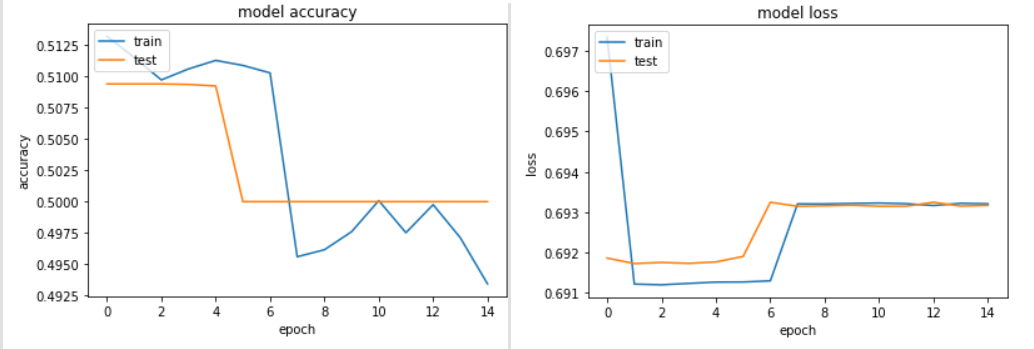
* שינינו את ממד הכניסה ל 15,000 הקטנת ממד הכניסה גרמה לשיפור מינורי וכמעט לא מורגש של 0.05% במדד ה Accuracy.



* שינינו את ממד הכניסה ל 30,000 הגדלת ממד הכניסה גרמה לירידה מינורית בביצועים.

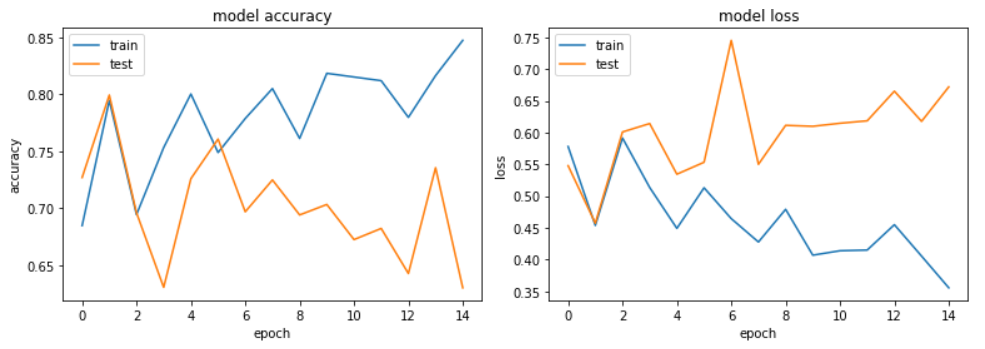
GRU -

* שינינו את ממד הכניסה ל 10,000 הקטנת ממד הכניסה גרמה לירידה חדה בשל ביצועי הרשת, מדד ה Accuracy עומד על 58% לעומת 81% בהרצה הראשונית.
* שינינו את ממד הכניסה ל 30,000 גם הגדלת ממד הכניסה גרמה לירידה משמעותית מאוד בביצועים (ואף גדולה יותר), לאחר הקטנת ממד הכניסה תהליך הלמידה הלך והתדרדר ככל שעבר הזמן.



SimpleRnn -

* שינינו את ממד הכניסה ל 10,000 הקטנת ממד הכניסה גרמה לירידה חדה בביצועי המערכת, מדד ה Accuracy ירד מ 77% ל 63%



* שינינו את ממד הכניסה ל 30,000 הגדלת ממד הכניסה גרמה לירידה קלה בביצועים וכמעט ולא מורגשת.

מסקנות -

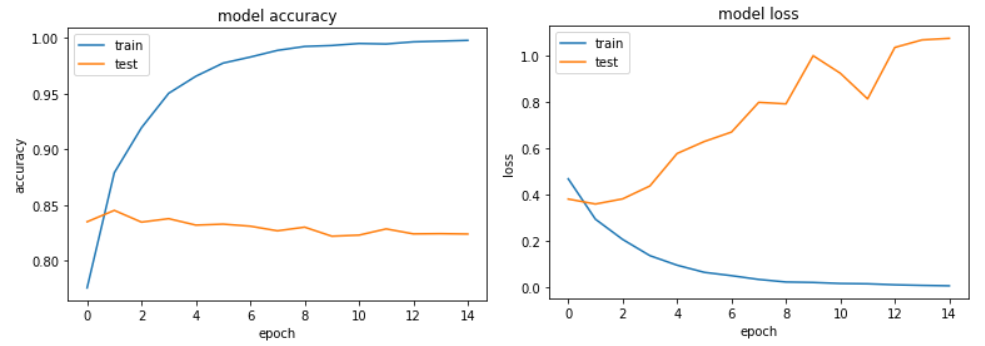
נראה כי ברוב המערכות הורדת ממד הכניסה גורם למודל לעבוד הרבה פחות טוב. ניתן לחשוב שזה קורה בגלל שהורדה בממד הכניסה נותן לנו אוצר מילים קטן מדיי שהמערכת מתקשה להסיק ממנו מסקנות.

**סעיף י"ב - שינוי ממדי היציאה:**

זהו בעצם שינוי מרחב ה embedding. את כל אחת מהתוצאות השוונו להרצה המקורית, ולהלן התוצאות:

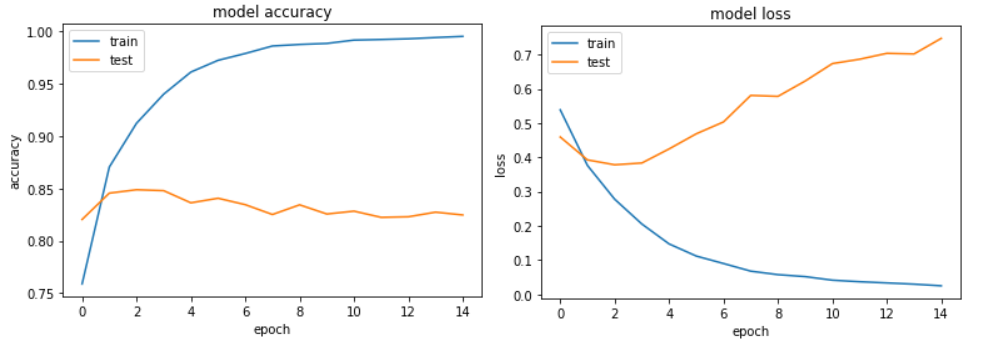
LSTM -

* שינינו את ממד היציאה ל 50 הקטנת ממד היציאה גרמה לירידה באחוזי הדיוק של 0.02%, זוהי ירידה לא משמעותית
* שינינו את ממד היציאה ל 180 הגדלת ממד היציאה גרמה לעלייה מינורית בביצועי הרשת ובסופו של דבר התוצאות זהות מאוד.



GRU -

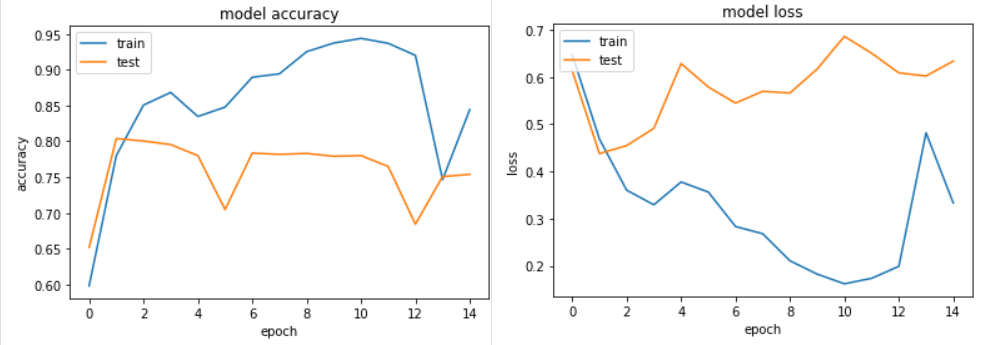
* שינינו את ממד היציאה ל 50 הקטנת ממד היציאה גרמה לשיפור קל במדד ה Accuracy שעלה מ 81% ל 82%.



* שינינו את ממד היציאה ל 180 הגדלת ממד היציאה גרמה לירידה חדה בביצועי המערכת ומדד ה Accuracy עומד על 58%

SimpleRnn -

* שינינו את ממד היציאה ל 50 הקטנת ממד היציאה גרמה לירידה קלה מאוד בביצועי המערכת.



* שינינו את ממד היציאה ל 180 הגדלת ממד היציאה גרמה גם היא לירידה קלה בביצועים.

מסקנות -

ציפינו שככל שממד היציאה יגדל וגם משום שה-data חסר ואינו מאוזן המודל יתקשה יותר לפתור את הבעיה מאחר והוא משתמש ביותר מילים מכל ביקורת ונאלצים לחשב "מרחקים" יותר גדולים בין מילים (ראינו בברור בתוצאות ה GRU). כלומר, כל שהממד גדל כך המודל פותר בעיה יותר "קשה"והמידע שנתון לא מספיק למודל ולכן מספק בעיה קשה מלכתחילה.

בפועל נראה כי אכן הורדת הממד הביאה לביצועים פחות טובים של המערכות, אך לא באופן חד משמעי.

**סעיף י"ד :**

עבור כל מודל ניתן 3 דוגמאות שהמערכת טעתה והביאה false positive או false negative, כלומר 3 ביקורות טובות עליהן המערכת אמרה שהן אינן טובות ולהפך:

LSTM -

* have long since become camp this one is filled with dubbed songs anachronistic slang and slapstick it's a truly crop of corn and pretty near intolerable today it was nominated for its imaginative special effects which are almost <UNK> in this day and age consisting mainly of trick photography the only outstanding positive feature which survives is its beautiful color and clarity sad to say of the many films made in this genre few of them come up to alexander <UNK> original thief of baghdad almost any other arabian nights film is superior to this one though it's a loser
* weird the box got opened boom they came was the only one that could bargain her way out of it first because of uncle frank then because she had information about the this movie at least attempts to stick to all that even though it was a bad story it was still somewhat hellraiser no i'm pretty sure part 5 was the first part to completely and utterly destroy the hellraiser series now they are remaking 1 and i don't even think i will watch it oh who am i kidding i probably will and probably will be disappointed again
* the mother in this movie is reckless with her children to the point of neglect i wish i wasn't so angry about her and her actions because i would have otherwise enjoyed the flick what a number she was take my advise and fast forward through everything you see her do until the end also is anyone else getting sick of watching movies that are filmed so dark anymore one can hardly see what is being filmed as an audience we are impossibly involved with the actions on the screen so then why the hell can't we have night vision

GRU -

* please give this one a miss br br kristy swanson and the rest of the cast rendered terrible performances the show is flat flat flat br br i don't know how michael madison could have allowed this one on his plate he almost seemed to know this wasn't going to work out and his performance was quite lacklustre so all you madison fans give this a miss
* the mother in this movie is reckless with her children to the point of neglect i wish i wasn't so angry about her and her actions because i would have otherwise enjoyed the flick what a number she was take my advise and fast forward through everything you see her do until the end also is anyone else getting sick of watching movies that are filmed so dark anymore one can hardly see what is being filmed as an audience we are impossibly involved with the actions on the screen so then why the hell can't we have night vision
* have long since become camp this one is filled with dubbed songs anachronistic slang and slapstick it's a truly crop of corn and pretty near intolerable today it was nominated for its imaginative special effects which are almost <UNK> in this day and age consisting mainly of trick photography the only outstanding positive feature which survives is its beautiful color and clarity sad to say of the many films made in this genre few of them come up to alexander <UNK> original thief of baghdad almost any other arabian nights film is superior to this one though it's a loser

SimpleRnn -

* please give this one a miss br br kristy swanson and the rest of the cast rendered terrible performances the show is flat flat flat br br i don't know how michael madison could have allowed this one on his plate he almost seemed to know this wasn't going to work out and his performance was quite lacklustre so all you madison fans give this a miss
* the height of that beleaguered country's civil war it would be easy to see this as a parable about those events <UNK> may or may not have had <UNK> turmoil in mind when he made <UNK> but whatever prompted his choice of material the film stands as a cautionary tale of universal application <UNK> could be the soviet union italy germany or japan in the 1930s or any country of any era that lets its guard down and is overwhelmed by tyranny it's a fascinating film even a charming one in its macabre way but its message is no joke
* copies for everybody i know everything and everybody in this movie did a good job and i haven't figured out why disney hasn't put this movie on dvd or on vhs in rental stores at least i haven't seen any copies this is a wicked good movie and should be seen by all the kids in the new generation don't get to see it and i think they should it should at least be put back on the channel this movie doesn't deserve a cheap download it deserves the real thing i'm them now this movie will be on dvd