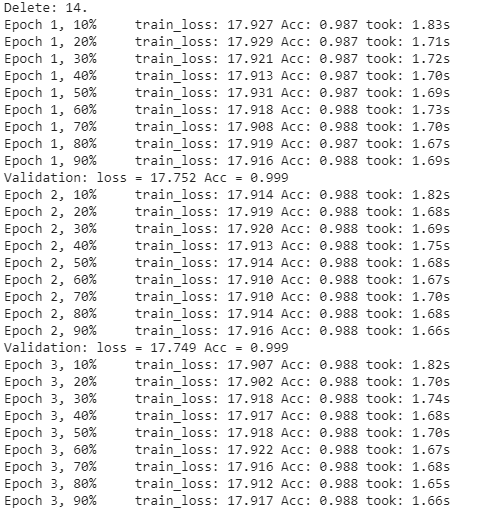
פרויקט סודוקו – תמצית

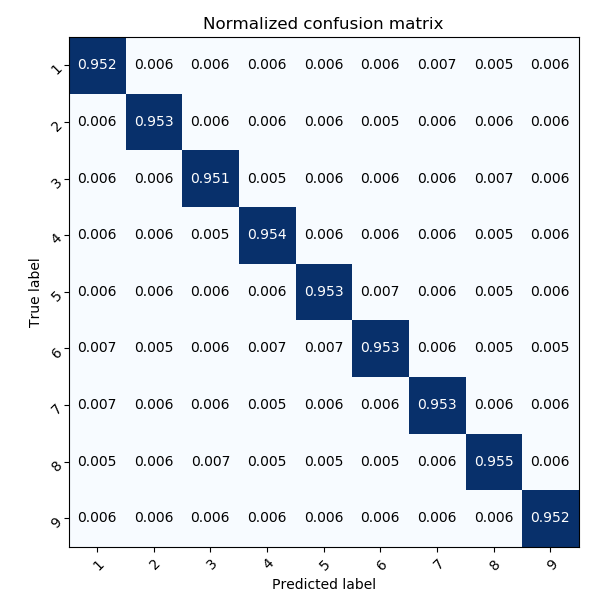
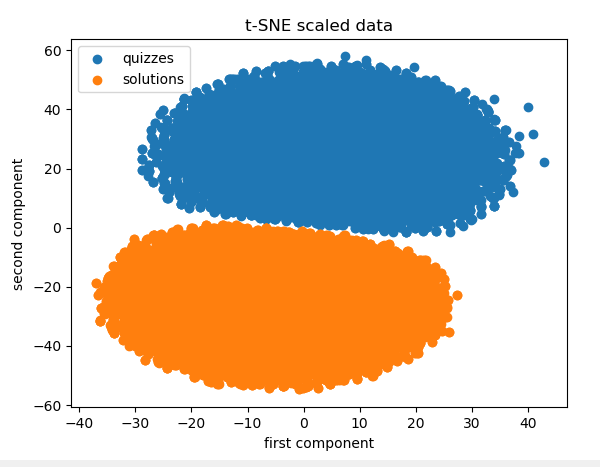
פתרונות תיאורטיים על סמך הכרת הבעיה:

1. יצירת גרעינים מותאמים לתבניות הסודוקו – שורה, עמודה, קוביה
2. עבודה עם וואן הוט וקטור מכיוון שאין ערך להפרש בין שני מספרים.
3. מילוי משבצות באופן איטרטיבי – בפתרון סודוקו בזמן אמת, נמלא כל פעם את המשבצת עם ההסתברות הגבוהה ביותר שהתקבלה מהרשת.

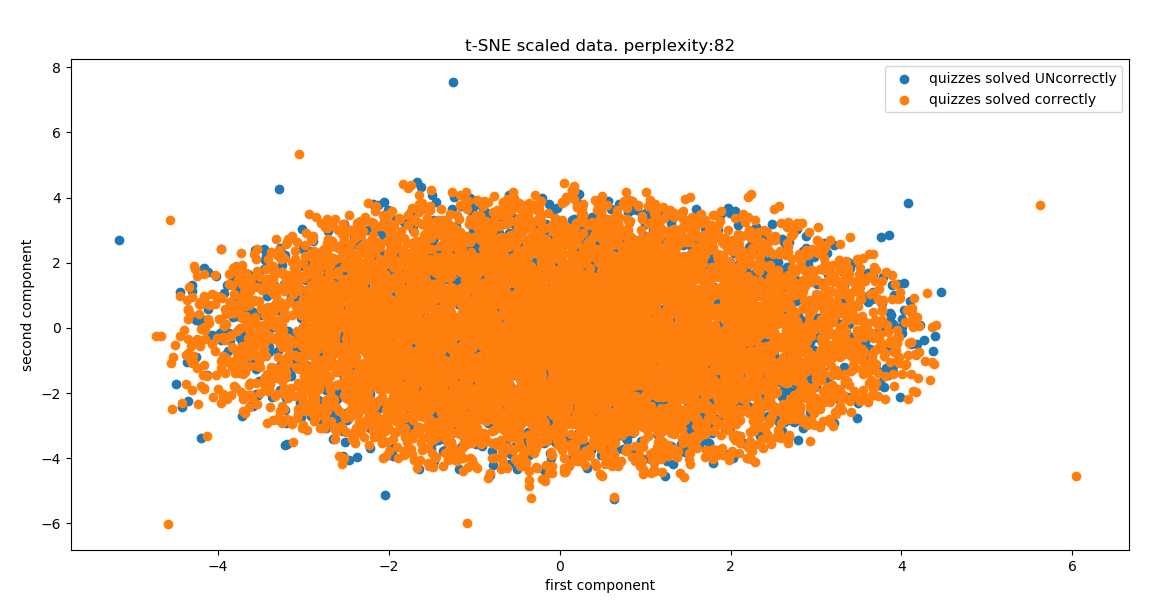
לאורך הדרך:

1. תחילה עבודה עם רשת FC פשוטה לדיבוגים כלליים ולנסות להצליח טוב על סט האימון.
2. עבודה על פונקצית לוס לטובת חישוב הלוס רק על האיברים שאותם אנחנו ממלאים. לצורך כך יש להכפיל את מטריצת הלוסים המתקבלת במסכה של האיברים שמילאנו ולסכום.
3. כדי לבדוק את הדיוק של הרשת שלנו, אנחנו צריכים לדמות מצב של טסט – כלומר מילוי איבר איבר. לכן בכל פעם שאנחנו בודקים את סט הוולידציה אנחנו למעשה לוקחים את הדיוק עליו לפי מילוי איבר איבר. לא עשינו זאת באימון מכיוון שזה לוקח הרבה זמן וזה חשוב בעיקר בשביל לראות את הדיוק במהלך האימון. מספיק בינתיים לראות את הדיוק רק בסוף כל איפוק על סט הוולידציה.
4. אחד הדברים שהיו בעייתיים זה הזמן ריצה הארוך של האימון והבדיקה את הדיוק האמיתי על הוולידציה. לכן השקענו זמן בלשפר את הזמנים בחישוב הדיוק לוולידציה. עשינו זאת על ידי זה שנפתרנו מלולאה (שהייתה בתוך לולאה אחרת) שרצה כל פעם על כל הבאטצ'. כך קיצרנו ביותר מפי 9 את הזמן של חישוב הדיוק האמיתי.
5. הגענו ככה למצב שהרשת לומדת יחסית לאט כאשר אנחנו מוחקים 14 מספרים ומעלה. כלומר הלוס משתנה ממש ממש לאט וזה מצריך הרבה איפוקים. אחת התיאוריות אומרת שזה קורה עקב כניסה של הרשת לאזור מינימום רחב כלשהו ששם הנגזרות קטנות יחסית. לכן ניסינו מדי פעם להגדיל באופן מלאכותי את קצב הלמידה. בנוסף, נשים לב שהבעיה שלנו משתנה כאשר אנחנו מוחקים מספר נוסף. כלומר, ייתכן שפונקצית המטרה משתנה ולכן מינימום שנקלענו אליו עד כה הוא פחות טוב ממינימום אחר שקיים אך לא נגיע אליו עקב קצב למידה קטן יחסית ומכך שאנחנו כלואים בתוך מינימום רחב. 

זאת אחת הבעיות העיקריות שלנו כרגע, אם כי עדיין הדיוק על סט הוולידציה משביע רצון יחסית (כשלא מוחקים יותר מדי מספרים). יוצא שבכל מחיקה של מספר נוסף אחרי ה 14 מחוקים, הלוס גדל בערך ב 5 ולא מצליח לקטון מספיק עד שמוחקים עוד מספר (וזה קורה אחרי 50 איפוקים עבור אותה מחיקה). להגדיל את מספר האיפוקים ייתכן שיעזור אך חשבנו שנוכל להגדיל את הקצב לימוד ולחסוך זמן ריצה.

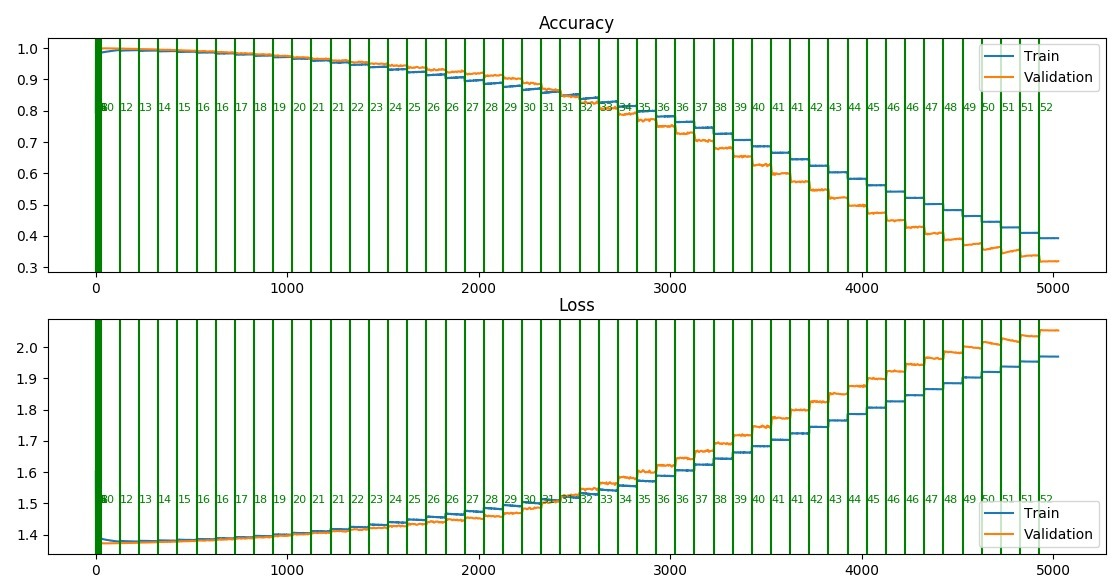
1. התחלנו להסתכל על ביצועי הרשת. ניתן לראות למשל שעבור 50 מחוקים (טסט דאטא) רשת FC שלנו הייתה עם 95% דיוק והניבה:  
   שמראה שהפילוג יחסית אחיד – כפי שציפינו. הרי אין הבדל משמעותי בין ספרה כזו או אחרת ולמעשה יש סימטריה. לכן גם ה confusion matrix יצאה סימטרית יחסית.
2. הסתכלנו על ההטלה למרחב דו מימדי של לוחות חסרים ושל לוחות פתורים מתאימים, על ידי TSNE וקיבלנו:  
   

כצפוי, ניתן לראות הבדלה מלאה בין המחלקות השונות.

זאת אבן ראשונה בהסתכלות על הטלת הלוחות. נמשיך וננסה הטלות מתוך מחלקות שונות.  
  
כעת ניסינו משהו אחר: לקחנו לוחות חסרים, והרצנו אותם ברשת. חלקם נפתרו נכון וחלקם לא. אז אנחנו מטילים את הלוחות החסרים לדו מימד בעזרת TSNE ומנסים לראות האם יש מאפיינים שדרכם ניתן לראות הבחנה בין הלוחות שהצלחנו לפתור לבין אלו שלא. התוצאות שקיבלנו לא מראות לנו הבדלים לצערנו:  


דבר נוסף שנוכל לבדוק הוא האם רשתות שונות שיצרנו טועות באותם לוחות.

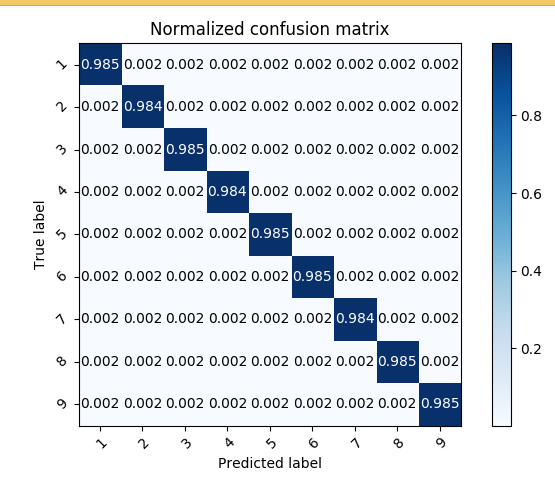
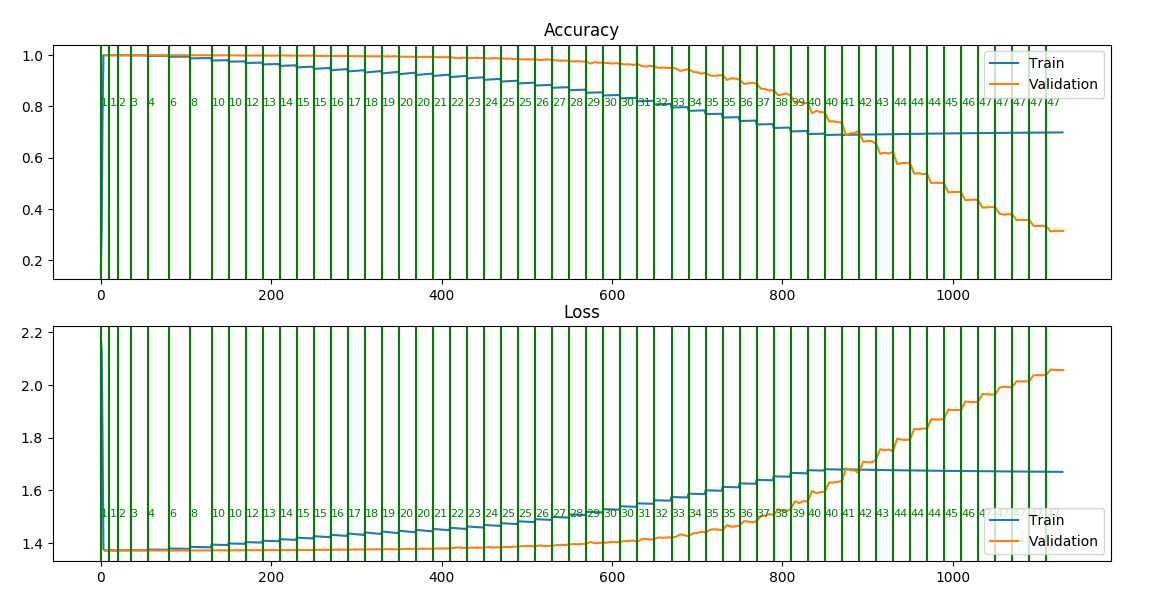
1. מסתבר שאנחנו מוחקים איברים, אז למשל כשמוחקים 60 איברים זה למעשה מוחק פחות כי זה מוחק את אותו איבר כמה פעמים.. אז לנסות לתקן את זה.. יוצא שאנחנו לא מוחקים 60 איברים אלא 38 בסוף.  
   לאחר שתיקנו את זה ופשוט מוחקים את אותם מקומות בכל הלוחות באותו הבאטץ' (תוצא לוואי בינתיים שאפשר לתקן), התוצאות נראות הרבה פחות טוב. למשל עבור רשת הקונבולוציה:



ננסה לעבור עם FC כי המידע אינו בהכרח לוקלי ואין לנו בעיה יותר מדי ממספר הפרמטרים של הרשת (האינפוט מאד קטן יחסית).

מה שמעניין הוא שרשת ה FC הביאה עדיין תוצאות טובות! כשאנחנו בודקים על טסט הדאטא כמו שהוא (ללא מחיקה מהפתרונות) אנחנו מקבלים דיוק **של 99.9% פר איבר ודיוק של 99.2% פר** **לוח**.

1. כשאנחנו מוחקים איברים, אנחנו למעשה עלולים ליצור לוחות שאינם לוחות סודוקו. מה שעלול להטעות את הרשת שלנו שלא לצורך (המטרה היא פתירת סודוקו). לכן, מה שננסה לעשות זה מתוך הלוחות המלאים, נמחק את האיברים שהיו מחוקים בלוחות החלקיים אחד אחרי השני. כלומר נניח שיש לנו לוח מלא ולוח חלקי. אז נמחק איבר אחד מהלוח המלא בהתחלה על ידי בחירת מיקום שבוא יש 0 בלוח החלקי (כלומר איבר מחוק) ואת המקום המתאים נמחק בלוח המלא. בשיטה זו בוודאות נשאר עם לוח שמתאים למשחק הסודוקו.
2. במצב הנוכחי, הרשת לומדת למלא את כל הלוח בבת אחת. בזמן מבחן אנחנו ממלאים את האיבר עם ההסברות הגבוהה ביותר. אולי נוכל לשנות את למידת הרשת כך שהיא תלמד למלא איבר אחד בלבד. אולי על ידי softmax על פני כל האיברים.
3. מה לגבי אוגמנטציות? אפשר לנסות למשל להחליף את כל המקומות בהם רשום 1 לספרה אחרת ולהפך. כך חוקיות המשחק נשמרת. או לסובב את הלוח סביב העמודה/השורה/האיבר האמצעי.
4. עבור רשת שטוחה יחסית (רק שתי שכבות FC יחד עם DROPOUT ) ועבור אימון בו בסט האימון אנחנו מוחקים באופן הולם ובולידציה לא, קיבלנו את התוצאות הבאות:

ניתן לראות שהמחיקה הלא הולמת משפיעה מאד על הולידציה. אנחנו נשווה את התוצאה הזאת מול מחיקה הולמת לבחינת כמות ההשפעה של אופן המחיקה על למידת הרשת.

1. בבואנו לשחזר את הרשת שעשינו בהתחלה ונתנה את התוצאות הטובות עד כה, רשת FC\_1 , עם אימון המכיל מחיקה לא לגיטימית עם חזרות, הפעם עשינו עד מחיקה של 55 איברים (ולא 62 כמו שהיה בפעם הקודמת). התוצאות שקיבלנו על הטסט המקורי החלקי, הוא **99.5% פר לוח ו 99.9% פר תא.** אלו התוצאות הטובות ביותר שקיבלנו עד כה. מה שמעניין הוא שהתוצאות אף השתפרו מפעם קודמת. האם ייתכן שהמחיקה של 55 רק ולא 62 זה מה שתרם לתוצאות? כלומר האם זה שאנחנו מקשים את הבעיה החל מסף מסוים פוגע בביצועי הרשת? יהיה מעניין לבדוק עם מחיקה מקסימאלית של 47 למשל ולבחון את התוצאות על הטסט. בנוסף, בדקנו את התוצאות על הטסט באופנים שונים: עבור מחיקה לגיטימית התוצאות היו טובות כמו הטסט המקורי. עבור מחיקה לא לגיטימית ללא חזרת התוצאות היו הרבה פחות טובות. עבור המחיקה לא לגיטימית עם חזרות, התוצאות היו גם פחות טובות מאשר מחיקה לגיטימית. כלומר אנחנו מבינים מכאן שהרשת למדה חלק ממאפייני לוחות הסודוקו ומחיקה לא לגיטימית יכול לפגוע בחוקיות לוח בסודוקו.

**הערות להמשך:**

לבדוק את התוצאות כאשר מוחקים בצורה לא לגיטימימת בtest כאשר מאמנים בצורה לגיטימית.

לעשות השוואה לעומת התוצאות בtest בצורה לגיטימית.

להתמקד בתוצאות עם לוחות לגיטימיים, ולעשות השוואה.

הגיוני להשאיר לוחות מחוקים ברמות שונות. לעשות את זה.

להסתכל על ההסתברות של המספרים שאנחנו לא משלימים. להסתכל מה קורה להסתברויות לאורך ההשלמה של הלוחות.

להסתכל האם ברגע שאני משלים אחד מהמספרים, אז במקום אחר באותה שורה ההסתברות של אותו מספר יורדת. כלומר האם הרשת מסיקה מסקנות באופן דומה למה שאדם עושה.

לסיכום:

לעשות שזה לא ירוץ 100 איפוקים אלא כמה איפוקים שצריך, אפשר להגביל ל-100 ולעצור לפני אם צריך.

למחוק לוחות בצורה תקינה, בלי חזרות באופן הגיוני.

האימון צריך להיות תמיד על לוחות תקינים.

להשאיר לוחות ברמות שונות של מחיקה.

לכתוב את הדו"ח

לראות את התוצאות של ה-CNN ולהשוות.

לנסות לראות אחרי שיש מודל מאומן, מה קורה מבחינת ההשלמה בtest - מה המספר שהיא נותנת על מיקום כלשהו, על ההסתברות, על המספרים שצריך להשלים - לאורך הריצה כולה.

אפשר אולי לחפור על coriculum בדו"ח כי זה נשמע מפוצץ.