

# (67103) רשתות נוירונים לתמונות | תרגיל 1

שם: רונאל חרדים | ת"ז: 208917641

חלק I

## חלק פרקטי:

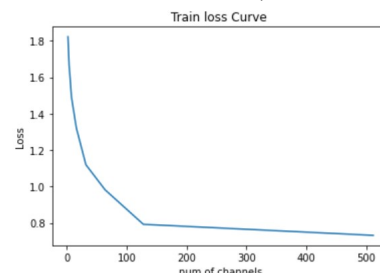
### שאלה 1

נשים לב כי עד 128 ערוצים מתרחש *underfitting* ולכן ככל שאנחנו מעלים את מספר הערוצים ה  $Loss$  יורד. (מספר ה  $epochs$  שהרצתי הוא 10).

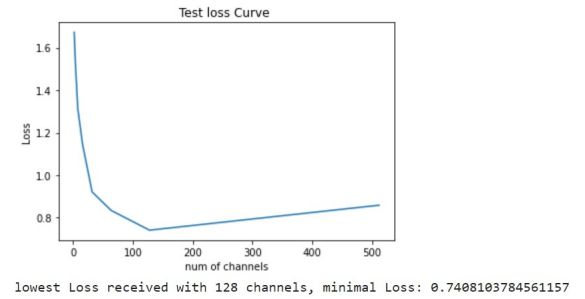
אך לאחר מכן מתרחש *overfitting* ומספר הערוצים הגבוה גורם ל  $Loss$  לעלות. מספר הערוצים שהשיג את ה  $Loss$  הנמוך ביותר הוא : 128

הגרפים שהדפסתי הם גרפים שהמראים את השינוי ב  $Loss$  הממוצע, כתלות במספר הערוצים. כשאר מספר הערוצים היה: [2, 4, 8, 16, 32, 64, 128, 512]

גרף ה  $Loss$  על ה  $Train$ :



גרף ה  $Loss$  על ה  $Test$ :



כאמור, ניתן לראות כי הגרפים של סט האימון וסט הבוחן היו קרובים עד 128, אך לאחר מכן ה  $Loss$  על סט האימון יורד, בעוד על סט הבוחן הוא עולה משום שמתרחש  $overfit$ .

### מספר הפרמטרים הנלמדים של הרשת הם:

יש לנו שכבת קונבולוציה, שכבת  $ReLU$  ושכבת  $pooling$  כך שלש פעמים, כאשר גודל הפילטר = 3. לכן סה"כ הפרמטרים של שכבות הקונבולוציה הם:

$first\ conv\ layer : (32 \cdot 32 \cdot 3 \cdot \text{num of out channels})$

$2\ other\ layers : (32 \cdot 32 \cdot \text{num of out channels} \cdot \text{num of out channels})$

כאשר

$$\text{num of out channels} \in [2, 4, 8, 16, 32, 64, 128, 512]$$

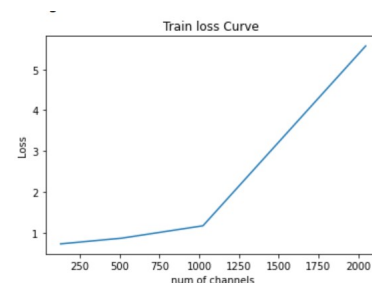
## שאלה 2

לאחר שהורדנו את שכבת ה  $ReLU$  וגרמנו לרשת להיות לינארית. גם עם מספר הערוצים שמקודם הבאנו את ה  $Loss$  למינימום קיבלנו את ה  $Loss$  הבא:

Average loss on the test set: 1.139

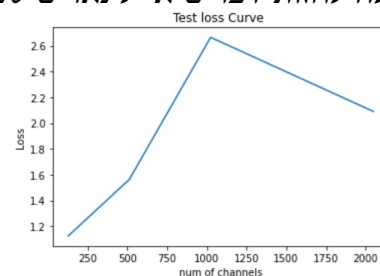
כאשר ה  $Loss$  עם השכבות האי לינאריות היה נמוך יותר. בחרתי את מספר ערוצי הפלט להיות כל פעם אחד מהבאים: [128, 512, 1024, 2048]

**כעת נראה את הגרף לאחר שהגדלנו את הרשת:**  
**הגרף על סט האימון:**



## הגרף על הטסט:

ניתן לראות כי הגדלת הרשת לא תרמה לירידת ה  $Loss$ . משום שהורדת השכבות אי הלינאריות מהרשת פוגעת בביצועים שלה וביכולות שלה לחזות דברים אי לינארים שמתרחשים בעולם האמיתי, וגם הגדלת הרשת לא תעזור להתמודד עם זה.



lowest Loss received with 128 channels, minimal Loss: 1.124134093952179

## שאלה 3

לאחר ששינינו את הרשת לעומק רדוד יותר, קיבלנו את התוצאות הבאות:

חיברתי את שכבת ה  $FC$  לפונקציית  $ReLU$  של שכבת הקונבולוציה הראשונה. כאשר מספר הפילטרים עמד על 512, כך שמספר המשקולות הנלמדים ברשת היה גבוה, אך ה  $Loss$  שהתקבל הוא:

Average loss on the test set: 1.507

ניתן לראות כי ביצועי הרשת פחות טובים, משום שהיא פחות יכולה לזהות פונקציות מורכבות הקיימות בעולם האמיתי.

## חלק II

## חלק תיאורתי:

### שאלה 1

נראה כי התנאי  $L[x(i+k)](j) = L[x(i)](j+k)$  שקול להגדרת הקונבולוציה:

נגדיר את פונקציית הדלתא להיות הפונקציה הבאה:

$$\delta(i) = \begin{cases} 1 & i = 0 \\ 0 & \text{else} \end{cases}$$

ונוכל לכתוב את  $x(t)$  באופן הבא:

$$x(t) = \sum_i x(i) \delta(t - i)$$

כי אם מתקיים ש  $i = t$  אז נקבל את  $x(t)$ , אחרת נקבל 0.

כעת נגדיר את  $g = L[\delta(t)]$ .

מהנתון מתקיים:

$$L[\delta(t - i)](s) = L[\delta(t)](s - i) = h(s - i)$$

ומהלינאריות של  $L$  מתקיים:

$$L[x(t)] = L\left[\sum_i x(i) \delta(t - i)\right] = \sum_i x(i) L[\delta(t - i)] = \sum_i x(i) h(s - i)$$

וזאת הגדרת הקונבולוציה כנדרש.

## שאלה 2

כאשר אנחנו עושים *flattening* לפלט  $2D$  ומכניסים אותו לשכבת  $FC$  הסדר לא משנה.

**הסיבה היא:**

לכל נירון בשכבה הקודמת יש צלע לכל נירון בשכבה הבאה. כלומר לא משנה כיצד נסדר את הקודקודים בשכבה הקודמת כולם יחוברו לשכבה הבאה והרשת תלמד את המשקולות ללא שינוי בסדר הקודקודים.

## שאלה 3

1.  $ReLU$ : פונקציה זו אינה לינארית כמו שראינו בהרצאה ולכן אינה  $LTI$ .

2.  $strafe$ : אינה  $LTI$ . משום שאם נבצע הזזה של תמונה ימינה, אזי - אם קודם נעשה פולינג ולאחר מכן נזיז אנחנו נקבל את כל הפיקסלים שממוקמים בצד ימין למעלה בכל חלון. מצד שני אם קודם נזיז ולאחר מכן נבצע פולינג נקבל את הפיקסלים מצד שמאל של החלון.

3.  $bias$ : אינו  $LTI$  משום שהוא לא לינארי כי הוא לא מקיים את הכלל של אדטיביות.

4. Multiplication with a fully-connected matrix: אם משקולות המטריצה קבועים, אז כפל המטריצה הינו לינארי. הוא מקיים את תכונות הלינאריות ולכן  $LTI$ . אך אם משקולות המטריצה ישתנו כתלות בזמן, אזי הפעולה לא  $LTI$ .