**תרגיל בית 2**

מגישים:

רם יזדי 305246308

גיא לורברבום 302773338

ספי לנגמן 305183527

קישור לקובץ 7t:

<https://drive.google.com/open?id=0B4kqDb__VN_TMTViSkxRaXY3aUE>

קישור לgithub:

<https://github.com/sefilang/machine-learning>

**ארכיטקטורת המודל:**

Conv: 5x5, 60  
Batch\_norm  
ReLU

Log-Soft-Max

Max-Pooling  
(8x8,1,1)

Conv: 1x1, 10  
Batch\_norm  
ReLU

Conv: 3x3, 30  
Batch\_norm  
ReLU

Conv: 3x3, 26  
Batch\_norm  
ReLU

Dropout  
0.3

Average-Pooling  
(3x3,2,2)

Conv: 3x3, 46  
Batch\_norm  
ReLU

Conv: 3x3, 32  
Batch\_norm  
ReLU

Conv: 1x1, 32  
Batch\_norm  
ReLU

Dropout  
0.3

Max-Pooling  
(3x3,2,2)

Conv: 3x3, 30  
Batch\_norm  
ReLU

Conv: 1x1, 42  
Batch\_norm  
ReLU

**תיאור המודל:**

הרשת הנ"ל מכילה 8 שכבות קונבולוציה כאשר השכבה הרביעית מורכבת משרשור של שתי קונבולוציות בעלות פילטרים בגדלים שונים. הורדנו את ממדי התמונות במהלך הרשת באמצעות שכבות Max/Average pooling בלבד כאשר בשכבות הקונבולוציה ריפדנו כך שממדי הקלט של השכבה יהיו זהים לממדי הפלט. לצורך הימנעות מהתאמת-יתר השתמשנו במהלך האימון בשכבות Dropout עם הסתברות של 0.3 לכבות נוירון. כמו כן, לצורך אוגמנטציה השתמשנו בהיפוך התמונה בצורה אופקית (hflip) כך שלכל תמונה במהלך האימון הייתה הסתברות 1/3 להתהפך. בנוסף, בטרם תחילת האימון אתחלנו את הקשתות לפי התפלגות נורמלית עם תוחלת 0 ושונות 0.25 ואת הbias ל-0.

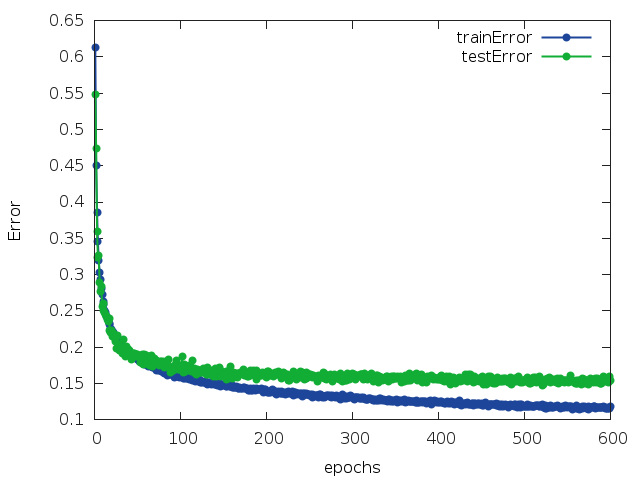
כמו כן, נרמלנו את מדגם האימון לפי כל ערוץ בנפרד.

פונקציית המטרה (קריטריון) היא Log-loss (ClassNLLCriterion) והאופטימייזר הוא adam.

מספר הepochs שבחרנו עבור המודל הוא 600, וכן בחרנו שהמודל יתאמן על המדגם ב-batchים בגודל 128. מספר הפרמטרים שהתקבלו הינו 49912**.**

**ניתוח המודל:**

להלן גרף המציג את טעות האימון והמבחן:



ניתן לראות כי השגיאה על מדגם האימון יורדת ככל שכמות ה-epochים עולה.

בנוסף, מ-epoch מספר 400 השגיאה על מדגם המבחן מתנהגת בצורה מונוטונית.

להלן גרף המציג את ה-loss על מדגם האימון ומדגם המבחן: (גרף זה נכתב ידנית מהפלט שקיבלנו לקובץ out):

**מסקנות וניסיונות:**

בתחילת התהליך, בדקנו כיצד לאתחל את קשתות המודל בצורה היעילה ביותר- ניסינו לאתחל עם ערך אפס, וכן עם ערכים מהתפלגות נורמלית. גילינו כי האתחול השני תורם לקצב התכנסות המודל.

בהמשך, ניסינו לממש ארכיטקטורות שונות ביניהן: שינוי במספר הקונבולציות כטרייד אוף בהוספת שכבה לינארית בסוף הרשת ( הוספת שכבה לינארית העלתה את מספר הפרמטרים באופן ניכר).

מסקנתנו הייתה שהורדת השכבות הלינאריות מהמודל ועל ידי כך אפשרות להוסיף קונבולציות נוספות (תוך הורדת ממדי התמונות ל1X1 ) הגדילה את דיוק המודל מסדר גודל של 77% ל-84.5%.

בנוסף, בארכיטקטורה דומה לזו שהגשנו אך ללא שימוש בשכבה מקבילית התקבלו תוצאות פחות טובות (83%).

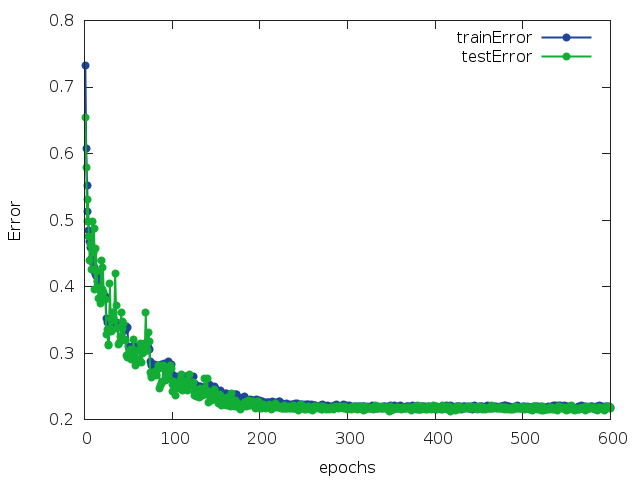
במהלך ניסיונותינו, הבחנו כי המודל מרבה לטעות על תמונות של חיות לעומת שאר הלייבלים (ככל הנראה בגלל שחיות יכולות להופיע בצורות מגוונות יותר). בעקבות כך, ניסינו בשלב האוגמנטצייה לבצע הטייה (rotate) לתמונות בעלות התיוג הנ"ל. ניסיון זה גרע מהדיוק ולכן מסקנותינו שלא כל אוגמנטצייה יכול לתרום לתהליך האימון.

במהלך ניסיונותינו השתמשנו בשני פונקציות אופטימיזציה שונות- sgd וadam:

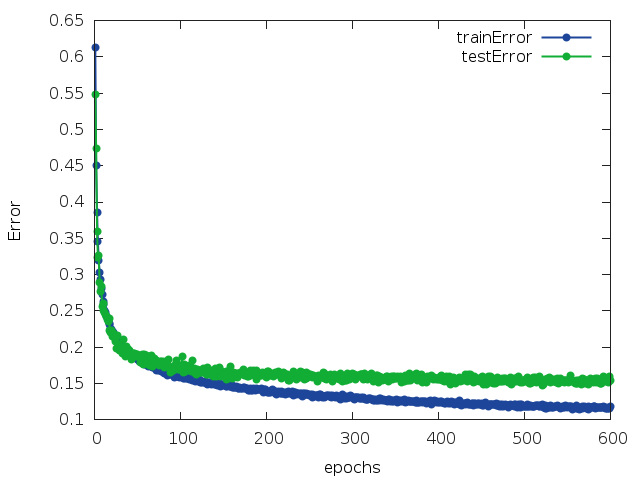
1. עבור sgd הדיוק הטוב ביותר שהתקבל הוא 78.5 לאחר אימון של 200 epoch – ים.

2. עבור adam הדיוק הטוב ביותר שהתקבל הוא 84.5 לאחר אימון של 500 epoch – ים.

3. להלן גרף המציג את טעות מדגם האימון והמבחן בעבור פונקציית sgd:



להלן גרף המציג את טעות מדגם האימון והמבחן בעבור פונקציית adam:



למרות שבאמצעות sgd מתקבל מודל יציב יותר (שגיאות המבחן והאימון דומות) התוצאות של פונקציית adam הינן טובות יותר וכן ניתן לראות כי למרות ההבדל בין שגיאות האימון לשגיאות המבחן אין הדרדרות בתוצאות שגיאת המבחן ולכן ניתן להסיק שאין overfitting.

**שיטות אוגמנטציה:**

1. אומגנטציה המכילה קומבינציה של hflip , crop עם ריפוד באפסים, crop עם ריפוד בהשתקפות, rotate (רק על מחלקות החיות) בהסתברויות שונות.

הדיוק ב- epoch החמישי היה מסדר גודל של 60, והדיוק הלך וירד עד שלאחר 300 epoch- ים הגיע לתוצאה של 19.441.

ניתן לראות כי שגיאת האימון גבוהה יותר משגיאת המבחן לאורך כל ההרצה. דבר זה עומד בניגוד לעיקרון הלמידה וככל הנראה נגרם עקב ריבוי אוגמנטצייה (האוגמנטצייה נעשית על מדגם האימון בלבד).

2. אוגמנטציה המכילה קומבינציה של hflip , rotate (רק על מחלקות החיות) בהסתברויות שונות.

הדיוק הטוב ביותר שהתקבל הינו 62.64% לאחר כ-580 epoch –ים.

ניתן להסיק כי אוגמנטציית הcrop אכן פגמה בתצורת המודל.

3. אוגמנטציית המודל שבחרנו- מכילה אך ורק hflip בהסתברות 1/3 עבור כל תמונה.

הדיוק הטוב ביותר שהתקבל הינו 84.5% לאחר 600 epoch- ים. (הגרף הרלוונטי מוצג למעלה).

מסקנתנו היא שיש להשתמש באוגמנטצייה במידה נכונה תוך כדי הכרות עם הdata אתו אנו עובדים.