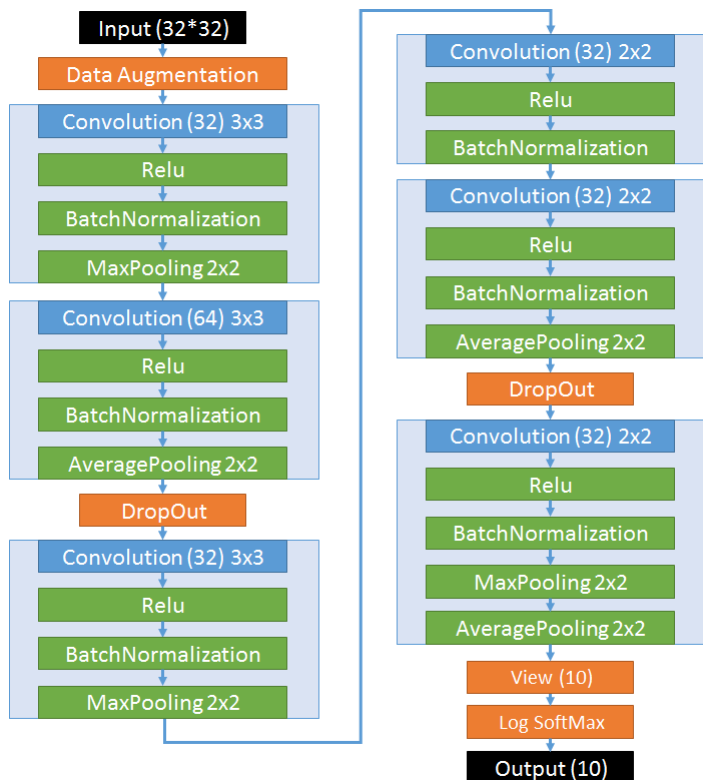


תרגיל בית 2 – נושאים מתקדמים במערכות מידע

037036209 שמעון ארזואן
 036565687 שגיא חכמון
 204097984 גיא רוזין

תיאור ארכיטקטורה

המודל שבחרנו מכיל 6 שכבות עיקריות, כמתואר באיור.



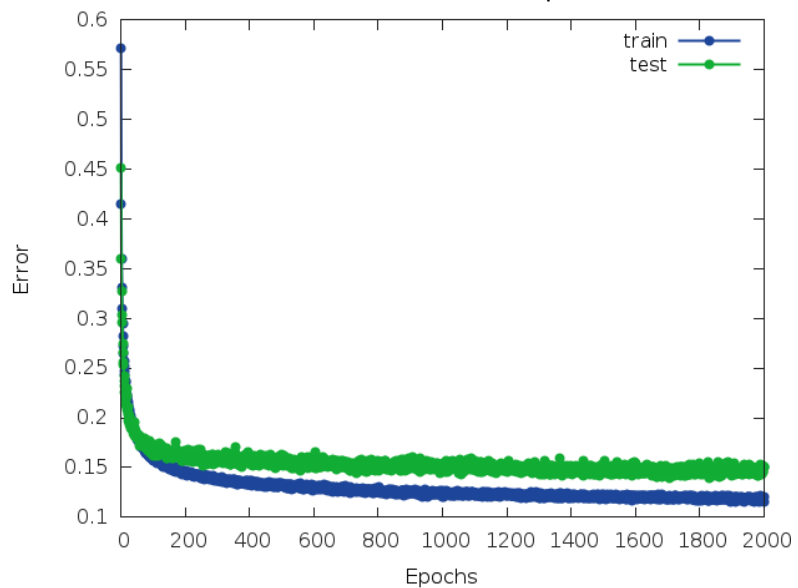
פרמטר	ערך
מספר Epochs	1704
גודל Batch	128
פונקציה אופטימיזציה	Adam
פרמטרי פונקציה האופטימיזציה	Learning Rate = 0.001 beta1 = 0.9 beta2 = 0.999 epsilon = 1e-8 weightDecay = 0.0001 momentum = 0.9
מספר פרמטרי המודל	47,806
View	כמספר המחלקות
הסתברויות לאוגמנטציה	1/3 – hflip 0 - rotation 2/3 – do nothing
DropOut	0.3

שיטת האימון

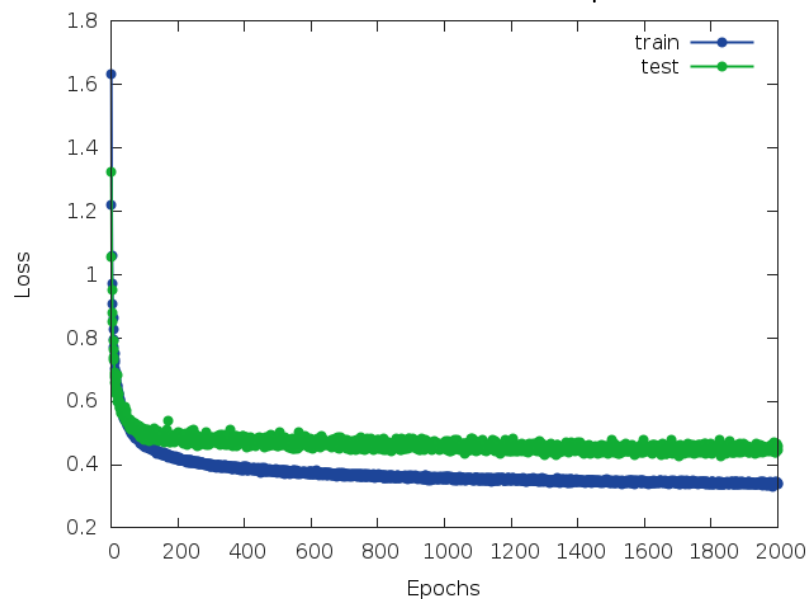
קראנו את סט האימון מספריית cifar-10, ונרמלנו אותו כך שהתוחלת שלו תהיה 0 וסטיית התקן 1. בנינו את הרשת לפי הארכיטקטורה המתוארת לעיל. לפני כל epoch ביצענו shuffle למידע. לאחר בניית המודל, ביצענו נרמול לסט המבחן לפי השונות והתוחלת של סט האימון, והרצנו את סט זה לאחר שביטלנו במודל את ה-dropout. ראינו כי לא כדאי להגדיר מראש את מספר ה-epochs מכיוון שלפעמים נתכנס במספר קטן ולפעמים במספר גדול של epochs. לכן בתהליך האימון בחרנו מספר גדול ממש של epochs, ובסיום ריצה הדפסנו גרף של השגיאה כתלות במספר ה-epochs, ושמרנו באיזה epoch קיבלנו את ערך הדיוק המקסימלי. כדי לחסוך בריצות מיותרות, הגדרנו שאם תוך 400 epochs לא שיפרנו אז נסיים את הריצה. בכך הקטנו באופן ניכר את זמן הריצה הכולל של האוטומציה שלנו שמריצה את המודל עם מספר פרמטרים שונים. המספר 400 נובע מכך שראינו שאם עדיין לא התכנסנו, נקבל ערך שגיאה קטן יותר בתחום של מעל 400 epochs (אפילו פחות, לקחנו יותר כטווח ביטחון). מספר ה-epochs הסופי להרצה נקבע לאחר ששמרנו באיזה epoch קיבלנו את ערך הדיוק המקסימלי. בסיום קיבלנו כי ה-Accuracy שלנו הוא **86.07**!

גרפים

שגיאה כתלות ב-epoch:



Loss כתלות ב-epoch:



סיכום הניסיונות שלנו

בכדי למצוא את השיטה הטובה ביותר הרצנו מספר ניסיונות על ארכיטקטורות שונות שבדקנו ע"י ניסוי וטעיה ותחושה מה עובד טוב יותר ומה פחות תורם. בנוסף בדקנו עומקים שונים של רשתות קונבולוציה. כמו כן ניסינו פונקציות אופטימום שונות לחישוב המשקולות של הרשת. ניסינו את השיטות : SGD ,ASGD ,AdaMax ,Adam ,AdaGrand ,AdaDelta כאשר לכל שיטה חיפשנו את הפרמטרים המתאימים לה (למשל קצב למידה, מומנטום ועוד). מצורפים קבצים של ניסיונות במודל עם פונקציית אופטימום:

1. SGD – התקבל דיוק 73.88 ב-epoch 1981, (גרפים מצורפים בנספחים).
 2. AdaGrad – התקבל דיוק 63.91 ב-epoch 1946, (גרפים מצורפים בנספחים).
- כמו כן מצורפים גרפים שהתקבלו מכל אחד מהם. לבסוף מצאנו את Adam כמתאימה ביותר עבורנו. בכל שיטה ניסינו לבדוק איך משפיע גודל ה-Batch של כל איטרציה, ולבסוף התקבענו על 128. ניסינו לראות אם שימוש בשכבות Dropout יכול לשפר את התוצאות והוא אכן שיפר. מצאנו לנכון להכניס 2

096260 - Deep learning

שכבות כאלה במודל. כדי לבדוק את ההשפעה של מספר ה-epochs על התוצאות, הרצנו עם מספר epochs גדול מאוד וניסינו למצוא באיזה epoch קיבלנו את התוצאה הטובה ביותר. זאת עשינו ע"י גרף של דיוק כתלות במספר ה-epoch ובנוסף שמרנו באיזה epoch התקבלה התוצאה הטובה ביותר. בנינו אוטומציה שמריצה את המודל עם מספר פרמטרים שונים בכל פעם כך בדקנו מה יהיו הפרמטרים המיטביים עבור המודל שלנו.

בהתחלה הייתה לנו בעיה עם ה-Data Augmentation, לכן הניסיונות הראשונים שלנו היו ללא האוגמנטציה עד שנתקענו על דיוק של 83%. בכדי לשפר את התוצאות הבנו שצריך להוסיף Data Augmentation. תחילה הסתכלנו איך יראו התמונות לאחר מספר פעולות אוגמנטציה שונות וראינו שלפעמים פעולות אלה משנות יותר מידי את התמונות כך שהן כבר כמעט נהרסות. לכן ראשית הפעלנו פונקציית הסתברות (מולטינומית) על האוגמנטציות השונות (או ללא שינוי תמונה). שנית, כיילנו כל אחת מהאוגמנטציות לשינויים קטנים בתמונה כדי שלא תהרס. הפעולה היחידה שגרמה לשיפור התוצאות הינה hflip. לאחר מספר ניסיונות הבנו שאכן לא הגיוני ש vflip יעבוד טוב, לכן החלטנו לוותר עליו ולהתמקד בסיבוב, חיתוך והשתקפות. כל אחת מהפונקציות לבדה פגעה בתוצאות: למשל, ביצוע סיבוב גרם ליצירת משולשים שחורים בפינות של התמונה, וביצוע חיתוך לפעמים חתך את האובייקט אותו אנחנו מנסים לזהות. לכן ניסינו לשלב בין פונקציות. כדי לעשות זאת נכון הסתכלנו על ה-Data וניסינו להבין איך הוא בנוי ואיך מגיב לפעולות אוגמנטציה שונות. לקחנו תמונות מסט האימון, יצרנו תמונה גדולה יותר בעזרת השתקפות, סובבנו את התמונה ואח"כ חתכנו את התמונה לגודל המקורי. כך למעשה פתרנו את הבעיה של המשולשים השחורים שנוצרו בפינות של התמונה בגלל סיבוב, בתקווה שישתפרו התוצאות. להלן תמונות שממחישות את הבעיה:



הוספנו את הפונקצייה החדשה ב-4 אופנים (סיבוב ימין/שמאל על תמונה מקורית/הפוכה), אך לבסוף זה לא שיפר לנו את התוצאות, כנראה בגלל שהסיבוב יוצר תמונה "מלוכלכת" יותר – כפי שניתן לראות בדוגמא. במהלך הניסיונות שלנו ניסינו שילובים שונים והסתברויות שונות לבחירת כל פונקציית אוגמנטציה. מצורפים קבצים וגרפים שהתקבלו מהניסיונות הללו:

1. הסתברות 0.5 ללא שינוי, 0.25 לסיבוב ימין, 0.25 לסיבוב שמאלה – דיוק 70.68.
2. הסתברות 0.5 ללא שינוי, 0.5 להיפוך הוריזונטלי (hflip) – דיוק 85.55.
3. הסתברות 2/3 ללא שינוי, 1/3 להיפוך (hflip) + סיבוב שמאלה – דיוק 73.96.

בכל התהליך בנינו קונפיגורציות שונות להרצה, הרצנו את כל הניסיונות שלנו באופן אוטומטי ולבסוף בחרנו כמה שיטות שעבורן קיבלנו את התוצאות הטובות ביותר ואותן ניסינו לשפר עוד קצת בעזרת שינוי הפרמטרים. לבסוף בחרנו את השיטה שהניבה את תוצאות הטובות ביותר על סט המבחן.

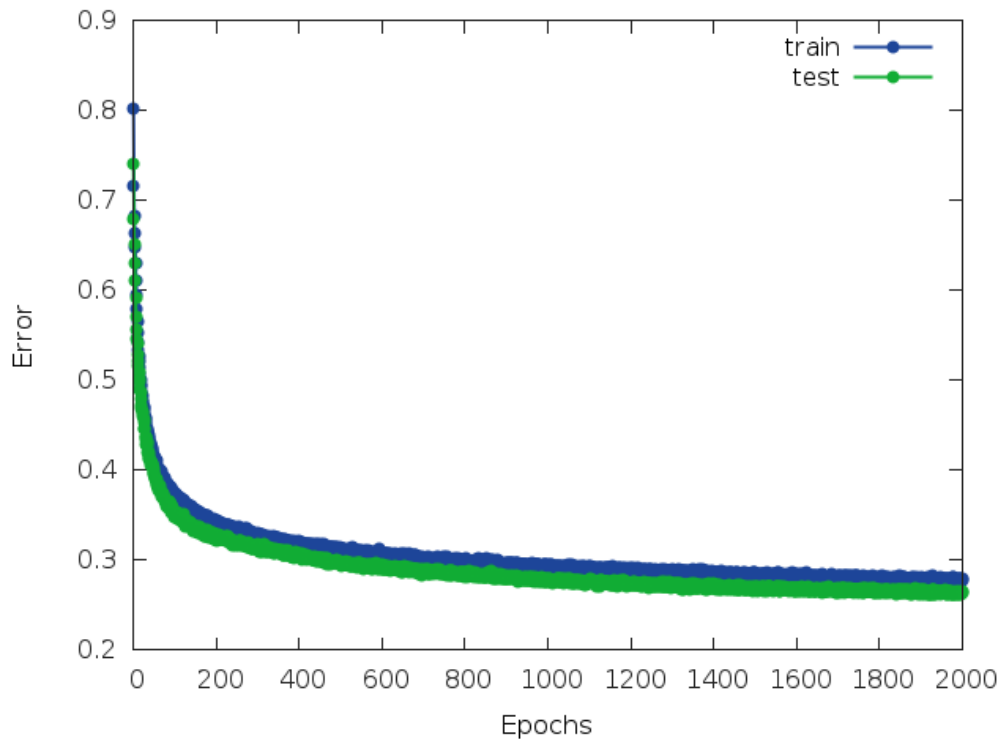
מסקנות

- פונקציות אופטימום שונות מתכנסות באופן שונה והן תלויות בבעיה, שווה לנסות כמה מהן בכדי לראות מי המתאימה ביותר לבעיה.
- גודל ה-Batch ומספר ה-Epochs משפיעים על התוצאות, איטרציות נוספות במקרים מסוימים משפרות את המודל, אך יש לשים לב כי ישנם מקרים שבהם מספר Epochs גדול מידי יכול לפגוע בתוצאות.
- ראינו כי Data Augmentation משפר את התוצאות בגלל שלמעשה מגדיל את כמות התמונות שעליהן אנחנו לומדים, רק יש להשתמש בו בזהירות כי לפעמים שינויים גדולים מידי בתמונה יכולים ליצור תמונות לא הגיוניות ולפגוע לבסוף בתוצאות.
- באוגמנטציה של הסיבוב המשופר שלנו, יתכן שעם תמונות גדולות וחדות יותר "הליכלוך/מריחה" שנוצרת בתמונה היה פחות מפריע והינו מקבלים שיפור בתוצאות.

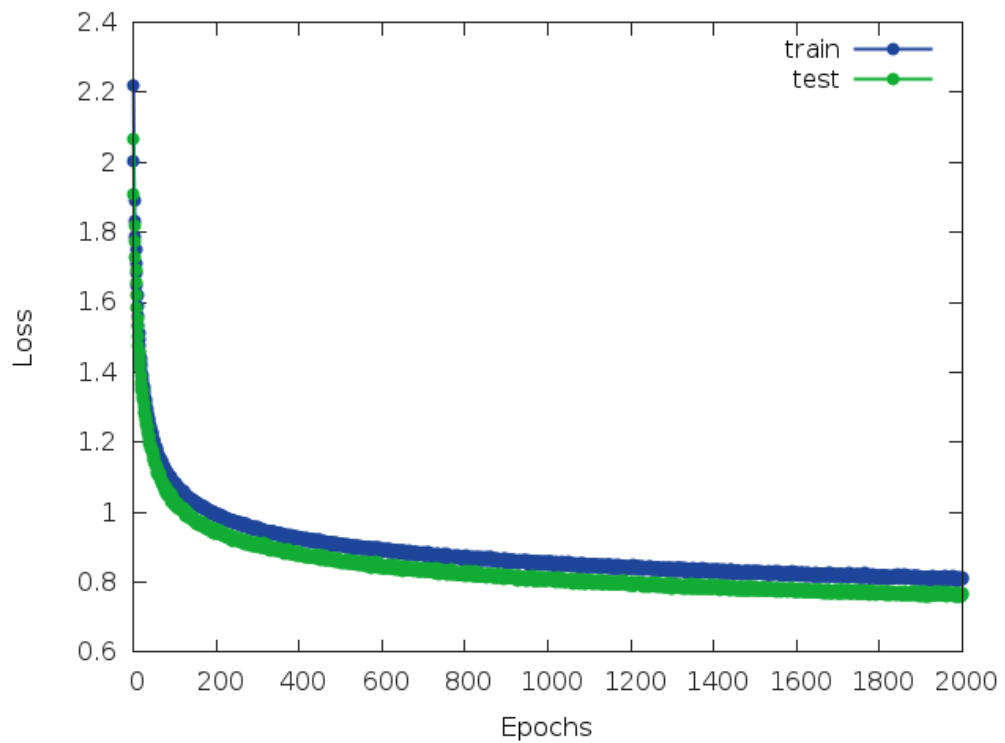
096260 - Deep learning

נספחים

— SGD - עם פרמטרים $\text{momentum} = 0$, $\text{weightDecay} = 0$, $\text{learningRateDecay} = 1e-4$, $\text{learningRate} = 1e-3$
התקבל דיוק 73.88, התקבל ב-epoch 1981.
שגיאה כתלות ב-epoch:

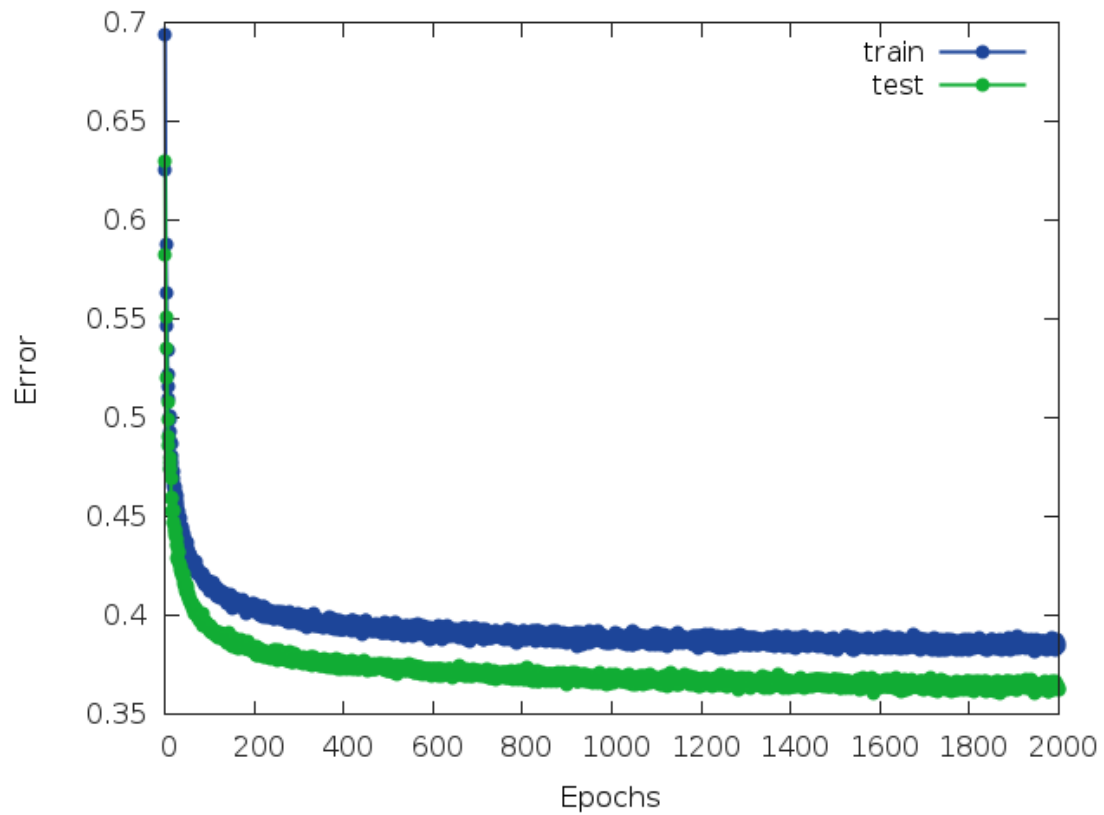


Loss כתלות ב-epoch:

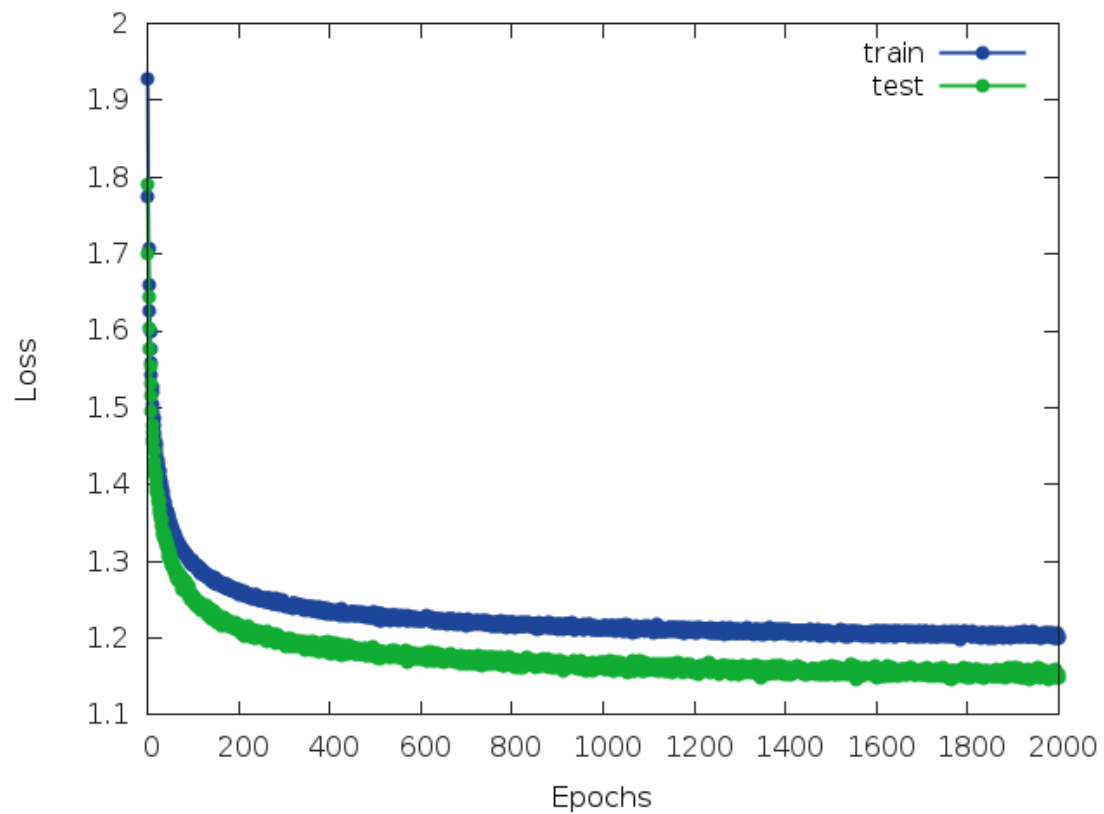


096260 - Deep learning

— AdaGrad - עם פרמטרים $\text{momentum} = 0$, $\text{weightDecay} = 0$, $\text{learningRateDecay} = 1e-4$, $\text{learningRate} = 1e-3$
התקבל דיוק 63.91 ב-1946 epoch.
שגיאה כתלות ב-epoch:

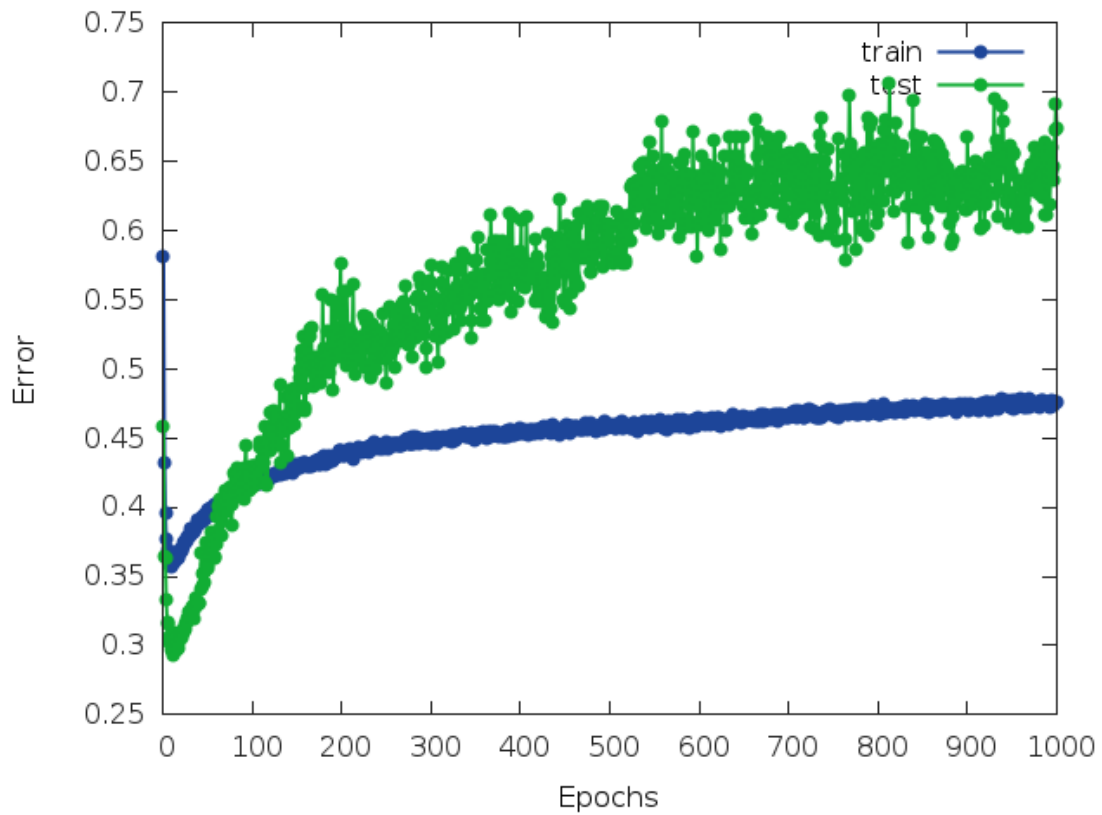


Loss כתלות ב-epoch:

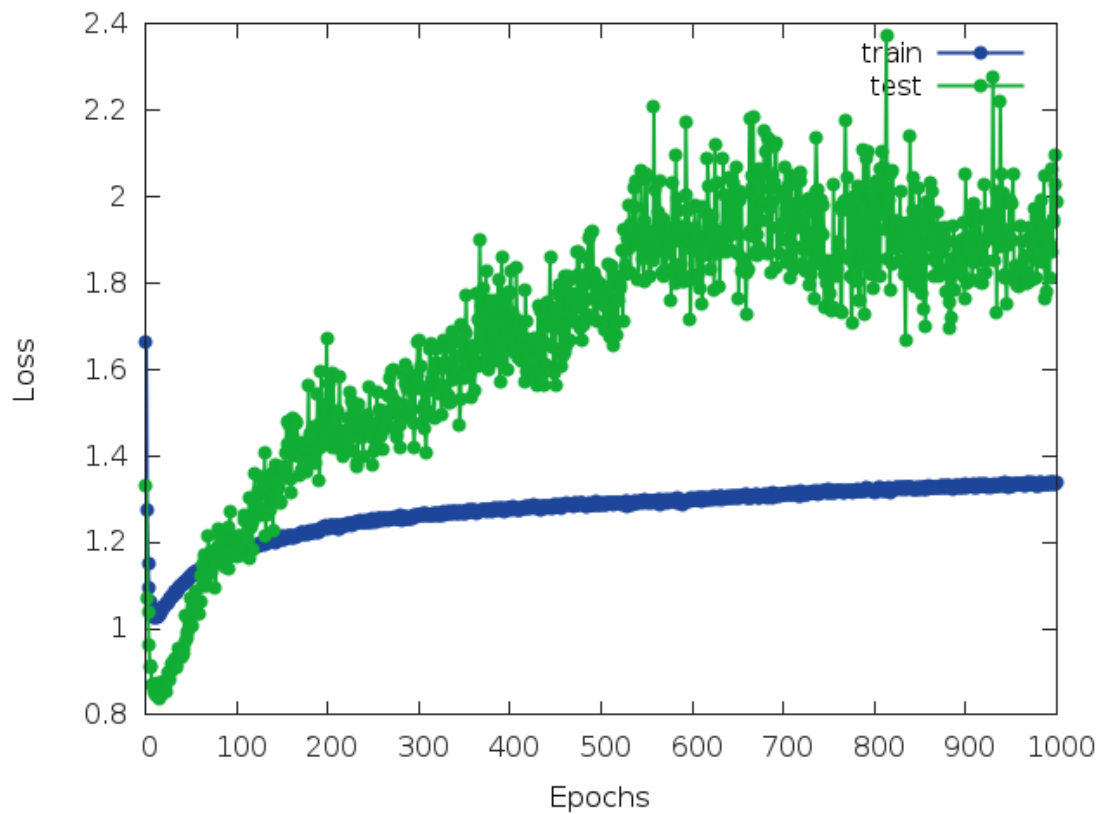


096260 - Deep learning

Data Augmentation 1 (A) עם הסתברויות 0.5 ללא שינוי, 0.25 לסיבוב ימינה, 0.25 לסיבוב שמאלה – דיוק 70.68 ב-epoch 11.
שגיאה כתלות ב-epoch:

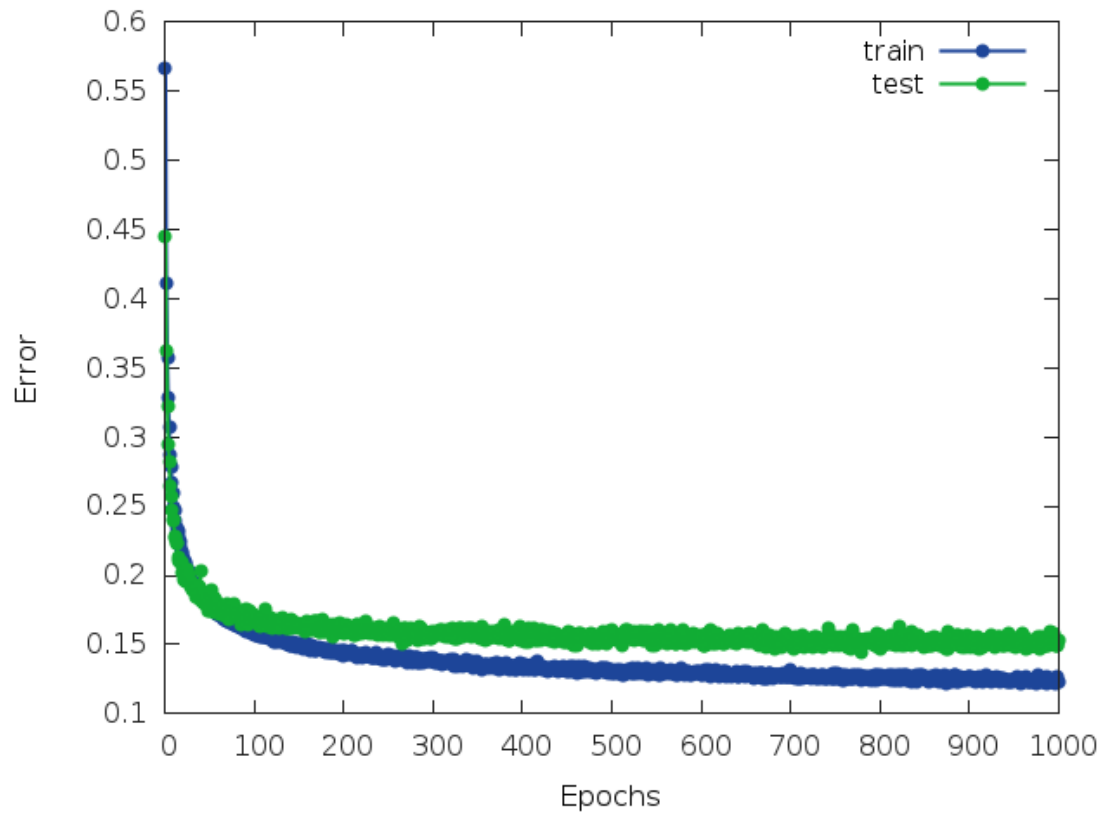


Loss כתלות ב-epoch:

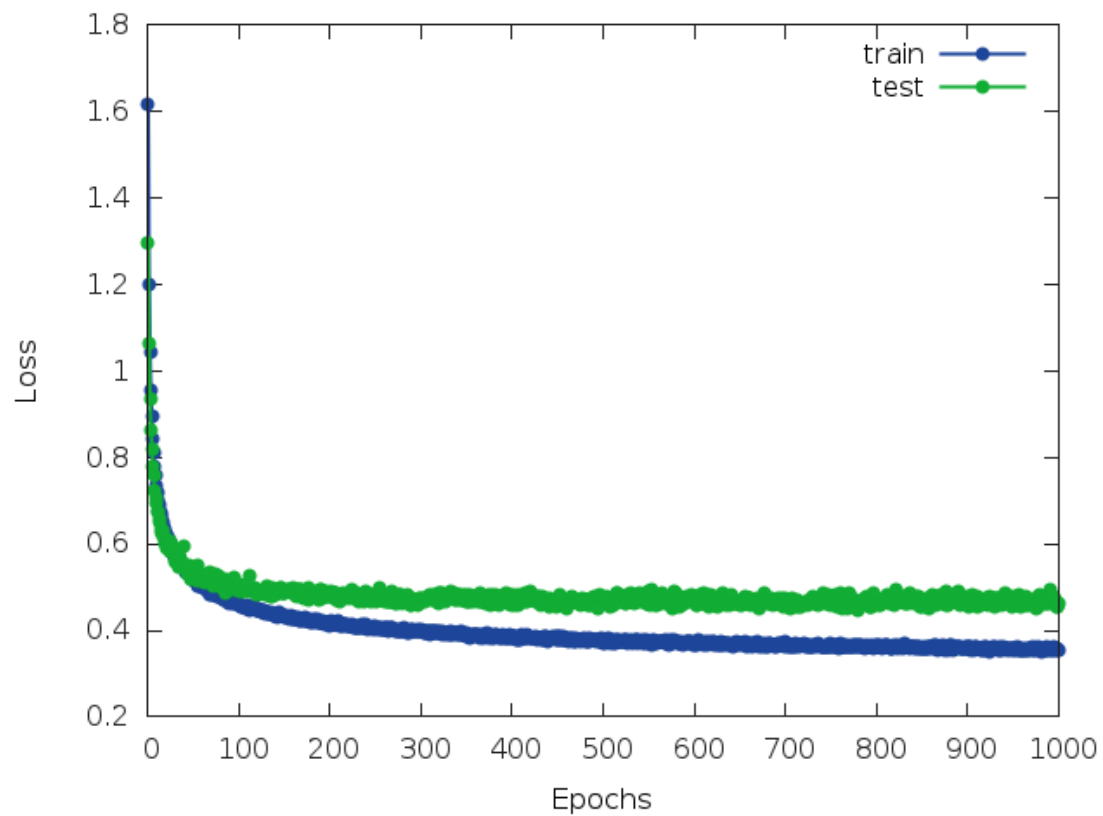


096260 - Deep learning

Data Augmentation 2 (B) עם הסתברויות 0.5 ללא שינוי, 0.5 להיפוך הוריונטלי (hflip) – דיוק
85.55 ב-779 epoch.
שגיאה כתלות ב-epoch:

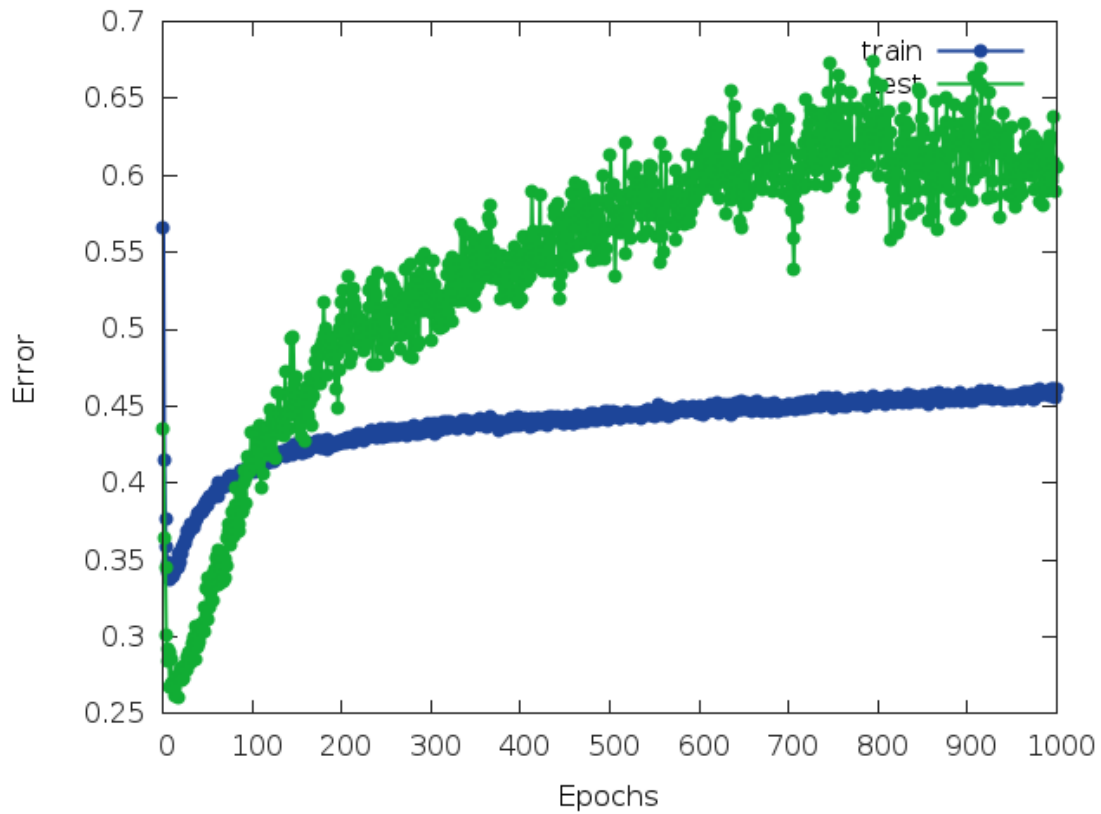


Loss כתלות ב-epoch:



096260 - Deep learning

– Data Augmentation 3 (C) עם הסתברויות 2/3 ללא שינוי, 1/3 להיפוך (hflip) + סיבוב שמאלה –
דיוק 73.96 ב-epoch 17.
שגיאה כתלות ב-epoch:



Loss כתלות ב-epoch:

