תרגיל בית 1

גיא לורברבום 302773338

רם יזדי 305246308

ספי לנגמן 305183527

הרשת בנויה סה"כ משמונה שכבות (7 שכבות חבויות). ארכיטקטורת הרשת מורכבת משכבות ליניאריות I-RELU לסירוגין כאשר השכבה האחרונה הינה log softMAX. וכל שכבה מחוברת באופן מלא לשכבה שקודמת לה. (סה"כ פרמטרים: 64660)

ניתן לתאר את הרשת באופן וויזואלי כך:

28\*28

10

254

linear

linear

RELU

linear

linear

log

Soft max

RELU

RELU

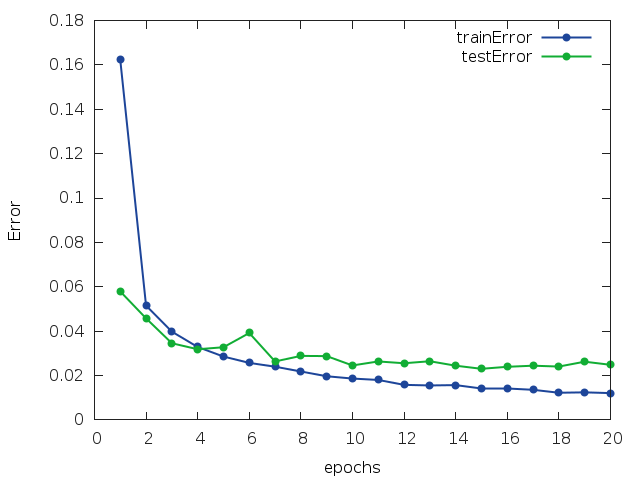
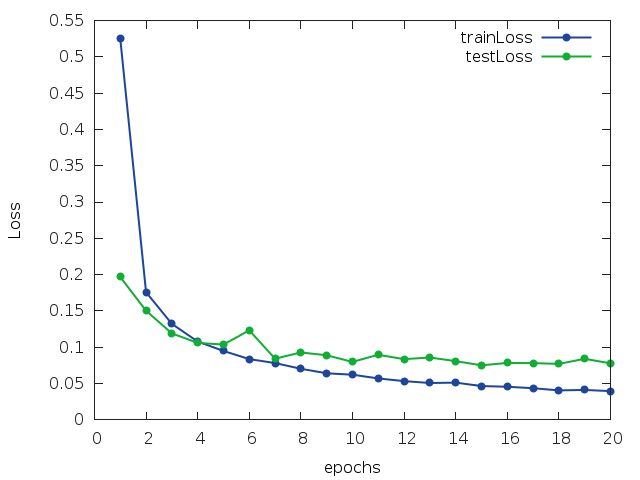
128

32

Input

הנתונים מנורמלים לפני תהליך האימון, גם סט-האימון וגם סט-המבחן מנורמלים לפי סט-האימון. כלומר, כל אחד מהפיקסלים מנורמל ע"י חיסור ממוצע הפיקסלים בסט-הנתונים וחלוקה בסטיית התקן. לפני כל epoch ביצענו shuffle של סט האימון.

 התוצאות שהתקבלו:

פונקציית האופטימיזציה שהשתמשנו בה היא Stochastic Gradient Descent

נסיונות:

* שימוש בפונקציית אופטימיזציה adagrad (ללא שינוי מספר ה-epochs) אך ללא שיפור בשגיאת המבחן (0.0291) ובשגיאת האימון (0.0214).
* שימוש בפונקציית אופטימיזציה SGD ושינוי מספר הepochs ל30. לא חל שיפור בשגיאת המבחן (0.0252) וחל שיפור בשגיאת האימון (0.0096).
* שימוש בפונקציית אופטימיזציה SGD ושינוי מספר הepochs ל5. לא חל שיפור בשגיאת המבחן (0.0354) וגם לא בשגיאת האימון (0.0309).
* שימוש בפונקציית אופטימיזציה SGD ושינוי מספר הbatchים ל16. לא חל שיפור בשגיאת המבחן (0.0147) ובשגיאת האימון (0.0327).
* שימוש בפונקציית אופטימיזציה SGD ושינוי מקדם הלמידה מ0.1 ל0.2. שגיאת האימון נשארה כפי שהייתה (0.0129) ולא חל שיפור בשגיאת המבחן (0.0293).

מסקנות:

1. על פי ניסיונותינו, פונקציית הadagrad נותנת תוצאות פחות טובות מפונקציות הsgd עבור מסד הנתונים עליו עבדנו.
2. כאשר מגדילים את מספר הepochs ישנו חשש לover fitting – כפי שהתקבל בניסיונותינו, קיבלנו כי שגיאת האימון קטנה בצורה משמעותית, ואילו שגיאת המבחן עלתה במעט מהשגיאה המקורית.
3. כפי שצפינו, כאשר מקטינים את מספר הepochs שגיאת המבחן וכן שגיאת האימון גדלות (מכיוון שמבצעים פחות איטרציות של האימון- המערכת מתקשה ללמוד בצורה טובה את המידע).
4. על פי ניסיונותינו, לא מצאנו קשר ישיר בין שינוי מספר הbatchים/ מקדם הלמידה לבין שיפור בשגיאות המבחן והאימון, על כן בחרנו אותם באמצעות tuning.

לבסוף כתבנו את הפונקציה predict שמקבלת את שם המודל כפרמטר ומחזירה את השגיאה. נציין, כי פונקציה זו מנרמלת את הנתונים לפי ערכי סט-האימון (כפי שהתבצע בשלב אימון הרשת)