

למידת מכונה תרגיל 3

הכנות מקדימות טרם ריצת המודל:

- חלוקת קובץ ה-train - חילקנו את קובץ ה-train שקיבלנו לשני חלקים בגדלים של 20%, 80%. אופן החלוקה מתבצע רנדומלית ע"י shuffle על קובץ ה-train, כך שבזמן הרצת המודל ולפני תחילת תהליך ה-train קובץ ה-train_x עובר shuffle, שלאחריו נלקחים 80% ל-train ו-20% ל-cross validation. לאחר שהאלגוריתם למד את ה-80%, הרצנו אותו גם על ה-20% בכדי לוודא שאחוז השגיאה דומה בין שניהם (כלומר, ווידאו שהאלגוריתם שלנו לא למד "לשנן" את ה-training set -אלא באמת למד לסווג כל data מתאימה שהוא יקבל).
- איתחול מטריצת המשקולות וההיסט - איתחלנו את הפרמטרים w_1, b_1, w_2, b_2 בערכים רנדומיים. אלו הפרמטרים שהמודל שלנו לומד ומעדכן עד שהוא מגיע לתוצאת דיוק אופטימלית. שמנו לב שבגלל שפונקציית ה-softmax היא אקספוננציאלית היינו צריכים לנרמל את הערכים ההתחלתיים (הרנדומיים), אחרת חורגים מגודל אפשרי של משתנה מספרי, ומקבלים Nan. לכן, לאחר שנבחרו ערכים רנדומיים ביצענו את הנרמול הבא: כפלנו את הערך ב-0.2 והחסרנו ממנו 0.1, כך הבטחנו שתמיד נשאר בטווח מספרי ולא נקבל Nan.

בחירת היפר-פרמטרים:

- FIRST LAYER SIZE - גודל השכבה הניורונית. במודל שלנו ישנה שכבה חבויה אחת שגודלה הוא 150. כאשר בדקנו גדלים שונים ראינו כי עבור גודל שגדול מ-30 מקבלים תוצאות לא רעות. מתחת ל-30 נתן תוצאות פחות טובות. ניסינו גדלים שונים, ושמנו לב שככל שהגודל הנבחר גדל זמן הריצה מתארך משמעותית עקב כפל מטריצות גדולות זו בזו, ועל כן בחרנו ב-150 כך שהיה גדול מספיק בשביל לתת תוצאות דיוק טובות, אך לא גדול מדי מבחינת זמן הריצה של המודל.
- ETA - קצב הלמידה של המודל, ה-learning rate. תחילה ניסינו לשים את הערך 0.1 וראינו שהמודל מגיע לתוצאות לא טובות מספיק. לאחר מכן ע"י שיטת הניסוי והטעייה ניסינו להגדיל את קצת הלמידה ל-1 ולאחר מכן ל-10 וראינו שעבור ערכים אלו אין שיפור, ובהרצות מסויימות אפילו היו תוצאות רעות יותר, לכן ניסינו להקטין את ערך ה-ETA. כאשר הקטנו ל-0.001 קיבלנו שיפור משמעותי. ניסינו להקטין עוד ל-0.0001 ושם קיבלנו תוצאות פחות טובות מ-0.001, לבסוף ניסינו את קצב הלמידה שערכו 0.005 והוא זה שנתן את התוצאות הטובות ביותר למודל שלנו.
- EPOCH - שמנו לב שהאלגוריתם אינו מתכנס למספר סופי, אך ככל שמתקדמות האיטרציות הוא משתפר עוד ועוד, עד שבנקודה מסויימת מפסיק השיפור ובכל איטרציה מקבלים פחות או יותר תוצאות קרובות מאוד אחת לשנייה. שמנו לב שזה קורה לאחר כ-40 איטרציות, לכן קיבענו את ה-EPOCH

להיות 50, כך שהאלגוריתם גם יגיע לתוצאות הטובות ביותר שהוא מסוגל, וגם זמן הריצה יהיה סביר ולא ארוך מדי.

- פונקציות האקטיבציה פונקציית האקטיבציה שנדרשנו לממש היא ה-relu, ולכן במעבר בין השכבות כאשר נקבל פלט משכבה מסויימת, נפעיל את relu על פלט זה והתוצאה תכנס כקלט לשכבה הבאה ברשת הנורונים. בשכבה הסופית במקום להפעיל את relu אנו נפעיל את פונקציית-softmax בכדי לנרמל את הפרדיקציה שלנו, וכך קיבלנו את \hat{y} .