

למידת מכונה – תרגיל 2

ההיפר פרמטרים:

Hyper-Parameter	Value
Epoch	60
Preceptron – Eta	0.1
SVM - Eta	0.1
SVM - lamda	0.075

בחירת הערכים בוצעה באופן אוטומטי, יצרנו רשימות עבור ערכים אפשריים לכל אחד מהיפר פרמטרים, לימדנו את המכונה, כך ש-80% היוו את training_set ו-20% היוו את validation_set. לאחר הריצה (שלקחה זמן רב) מצאנו את הפרמטרים שנתנו לנו את אחוז loss הנמוך ביותר ואותם בחרנו. במהלך הריצה שמנו לב כי מכיוון שה-data אינו ספרבילי לא מתקיימת התכנסות לפני ההנחות שהנחנו במקרה של data ספרבילי, ולכן הגדלת מספר ה-epoch לא בהכרח עזר (מה שכן יעזור עד מספר מסויים ב-data ספרבילי).

בנוסף על מנת לקבל יציבות ומהימנות של התיוגים, ביצענו 6 חזרות על תהליך הלמידה, ולקחנו עבור כל index לקחנו את התיוג שיצא הכי הרבה פעמים.

נרמול הנתונים:

ניסנו לבצע 2 צורות של נרמול:

- MinMax
- Zscore

שמנו לב שמכיוון שהמספרים כבר בין 0 ל-1, נרמול ה-MinMax אינו תרם בצורה משמעותית ולכן בחרנו להשתמש ב-Zscore שראינו כי מביא לתוצאות יותר טובות.

המרת נתונים לא מספריים:

אחד מהפיצ'רים שקיבלנו הינו "מין" שיכול להיות אחד מהשלושה הבאים:

- M
- F
- I

שלושת האלגוריתמים שלנו יודעים לעבור רק עם ערכים מספריים ולכן היינו צריכים להמיר ערכים אלו למספרים. בדקנו לכך 2 אופציות:

- המרה למספרים קבועים
- One-hot

One-hot מאפשר המרה אם קיימים מספר סופי של ערכים עבור הפיצ'ר (במקרה שלנו 3) ובכך המרנו את הפיצ'ר ל3 פיצ'רים שונים וסימנו 1 במקום המתאים לפי הערך המקורי של הפיצ'ר, דבר שהוביל לביצועים יותר טובים עבור data שלנו מאשר עבור המרה למספרים קבועים.