

דו"ח תרגיל 2 בלמידת מכונה – אורי דאבוש 212945760:

בתרגיל מימשתי 4 אלגוריתמים – KNN, Perceptron, SVM ו-PA. כל אחד מהאלגוריתמים (פרט ל-KNN) מחזיק מטריצת w שמייצגת n וקטורים שונים לפי מספר המחלקות (במקרה שלנו 3). בנוסף הוא מחזיק וקטור bias. לכל אלגוריתם כלל עדכון ופונקציית loss משלו.

האלגוריתם Perceptron משתמש בפונקציית loss של 0-1 (כלומר 0 במקרה של תיוג נכון ו-1 אחרת). בנוסף כלל העדכון שלו הוא הבא:

$$w_{t+1}^y = w_t^y + \eta * x$$

$$w_{t+1}^{\hat{y}} = w_t^{\hat{y}} - \eta * x$$

$$w_{t+1}^{i \neq \hat{y}, y} = w_{t+1}^{i \neq \hat{y}, y}$$

האלגוריתמים SVM ו-PA משתמשים באותה פונקציית loss, שהיא:

$$\ell(x^t, y^t, w) = \max(0, 1 - w_{y^t} x^t + w_{\hat{y}^t} x^t)$$

$$\hat{y}^t = \arg \max_{y^i \neq y^t} w_{y^i} \cdot x^t \quad \text{כאשר}$$

כלל העדכון של SVM הוא:

$$w_{t+1}^y = (1 - \eta \lambda) w_t^y + \eta * x$$

$$w_{t+1}^{\hat{y}} = (1 - \eta \lambda) w_t^{\hat{y}} - \eta * x$$

$$w_{t+1}^{i \neq \hat{y}, y} = (1 - \eta \lambda) w_{t+1}^{i \neq \hat{y}, y}$$

$$\hat{y}^t = \arg \max_{y^i \neq y^t} w_{y^i} \cdot x^t \quad \text{כאשר}$$

וכלל העדכון של PA הוא:

$$w_{t+1}^y = w_t^y + \tau * x$$

$$w_{t+1}^{\hat{y}} = w_t^{\hat{y}} - \tau * x$$

$$w_{t+1}^{i \neq \hat{y}, y} = w_{t+1}^{i \neq \hat{y}, y}$$

$$\hat{y}^t = \arg \max_{y^i \neq y^t} w_{y^i} \cdot x^t \quad \text{כאשר} \quad \tau = \frac{\ell(w, x, y)}{2 \cdot \|x\|^2} \quad \text{ו-}$$

את ה-bias של כל אלגוריתם מעדכנים בהתאם לכלל העדכון, כאשר מתייחסים ל-x כאילו הוא 1.

לכל אלגוריתם יש כמה היפר-פרמטרים שקבעתי באופן שרירותי לפני ההגשה. כדי למצוא את ההיפר-פרמטרים האופטימליים יצרתי סקריפט שיריץ את האלגוריתמים עם פרמטרים שונים, עד שמצאתי את ההיפר-פרמטרים שמוציאים את התוצאות הטובות ביותר. כדי למדוד את התוצאות ביצעתי כל אלגוריתם 16 פעמים, ולקחתי את הממוצע של התוצאות השגויות על סט האימונים (train_x) ועל סט המבחן (test_x).

בנוסף להיפר-פרמטרים של כל אלגוריתם, ישנם 3 שיטות נרמול – z score (נרמול לפי סטיית תקן וממוצע), min max (נרמול לטווח [0,1]) והשארת הנתונים כפי שהם. לכל הפרמטרים בחנתי את 3 שיטות הנרמול.

בשביל KNN הרצתי את האלגוריתם עם 3 שיטות הנרמול ועם ערכי K שונים, וקיבלתי שהקומבינציה הכי טובה היא k=5 עם נרמול של min max. (ב-KNN לא הרצתי 16 פעמים לכל קומבינציה של פרמטרים כי הפלט יהיה אותו פלט (אין חלק רנדומלי)).

בשביל Perceptron הרצתי את האלגוריתם עם 3 שיטות הנרמול ועם learning rate של 0.1 עד ל-1 בקפיצות של 0.1. קיבלתי שהתוצאות הטובות ביותר היו עם נרמול z score ועם learning rate של 0.8. בנוסף השתמשתי כאן ב-k fold cross validation עם k=17 (בחנתי מספר אפשרויות עבור k). האלגוריתם רץ עם 200 איפוקים (epochs) כדי למצוא את ה-w הטוב ביותר לכל סט אימונים וסט ולידציה.

בשביל SVM הרצתי את האלגוריתם עם 3 שיטות הנרמול ועם learning rate ו-lambda של (פרמטר רגולרציה) של 0.1 עד ל-1 בקפיצות של 0.1. מצאתי שהתוצאות הטובות ביותר היו עם נרמול z score ועם learning rate של 0.1 ו-lambda של 0.001. בנוסף השתמשתי כאן ב-k fold cross validation עם k=10 (בחנתי מספר אפשרויות עבור k). האלגוריתם רץ עם 200 איפוקים (epochs) כדי למצוא את ה-w הטוב ביותר לכל סט אימונים וסט ולידציה.

בשביל PA הרצתי את האלגוריתם עם 3 שיטות. מצאתי שהתוצאות הטובות ביותר היו עם נרמול z score. בנוסף השתמשתי כאן ב-k fold cross validation עם k=10 (בחנתי מספר אפשרויות עבור k). האלגוריתם רץ עם 200 איפוקים (epochs) כדי למצוא את ה-w הטוב ביותר לכל סט אימונים וסט ולידציה.