

למידת מכונה תרגיל 2

30 בנובמבר 2021

שיטות לדיוק הלא תלויות באלגוריתם

הוספת bias - הוספתי לכל feature של פרח המכיל 5 תכונות, תכונה נוספת. משמשת בתור bias כדי שיהיה יותר דיוק בתוצאות.

בהתאמה גם כל משקולת הורחבה כך שיהיה לה ערך שיש.

מימוש ניתן לראות בפונקציה add_bias בשורה 21

נרמול - השתמשתי בzscore כדי לנרמל את וקטורי featuresn על מנת שסדרי הגודל יאפשרו עבודה יותר מדויקת על

הדאטא

מימוש ניתן לראות בפונקציה zscore בשורה 28

שיטות לדיוק התלויות באלגוריתמים

- knn אלגוריתם

עברתי על k שונים מ 1 עד 20 וגיליתי ש $k = 5$ נותן לי תוצאות הכי טובות.

מעל 90 אחוז בכל ההרצות ומעל 95 אחוז ברובן (על 192 דוגמאות למידה ו481 מבחנים)

- perceptron אלגוריתם

עבור מספר האיטרציות (epoch) בחרתי ב10000. מצאתי שהדיוק במספר זה הוא גבוה (מעל 93 אחוז ברוב ההרצות) וגם מאפשר להריץ את כל הקוד בפחות מ5 דקות.

עבור משתנה הלמידה η בדקתי אפשרויות מ 10^{-5} עד 1 ומצאתי ש $\eta = 10^{-3}$ יוצא מדויק ביותר.

- SVM אלגוריתם

באופן דומה לאלגוריתם הקודם, בחרתי במספר איטרציות של 10000.

עבור משתנה הלמידה η בחרתי בערך 10^{-4} מכיוון שבאלגוריתם הקודם ראיתי ש 10^{-4} מתאים אבל רציתי שיהיה למשתנה

הלמידה פחות משקל מכיוון שקיים לנו עוד משתנה בחישוב. גם בבדיקות ערך זה הביא תוצאות מדויקות.

עבור המשתנה λ בדקתי אפשרויות מ1 עד 10 ומצאתי שעריך של 3 מביא תוצאות מדויקות מאוד. יותר מעריך זה ומשתנה

הלמידה נהיה זניח או שאיזון של משתנה הלמידה כך שיתאים ל λ מביא להבדלים גדולים באחוזי ההצלחה בין בדיקות

- Pa אלגוריתם

המימוש בעיקר מבוסס על SVM, כאשר τ מוגדר על פי פונקציית ההפסד והוקטור של featuresn כנלמד בתרגול.