SVM (Support Vector Machine)

אפשר לחשוב על SVM כאלגוריתם שממזער:

$$\min_{w \in \mathbb{R}^d} \quad \frac{1}{2}\|w\|^2$$
$$\forall_i y_i w^T x_i \geq 1$$

אם יש לנו נתונים שאינם ניתנים להפרדה לינארית, נשתמש בsoft-SVM - נוסיף soft margin:

$$\min_w \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i$$

ונשנה את התנאי על w ל:

$$\forall_i \qquad y_i\left(w^T x_i\right) \geq 1 - \xi_i \qquad \xi_i \geq 0$$

הרעיון הוא למקסם את השוליים תוך מתן אפשרות לחריגות.

Constrained Optimization

עד עכשיו, אם הייו להוסיף prior לבעיה היינו מוסיפים אותו כמרכיב לloss. למשל - אם הינו רוצים שהפרמטרים יהיו קטנים היינו מוסיפים $\mathcal{L} = \cdots + \lambda \cdot \frac{1}{2}\|w\|^2$. אבל יש גם מקרים עם אילוצים קשיחים כמו hard-SVM. במקרים כאלה אי אפשר להפעיל שיטות כמו gradient descent - צריך לעשות משהו אחר.

בד"כ נסתכל על בעיות מינימיזציה מהצורה[1]:

$$\min_x f(x) \qquad \forall_{i=1,\ldots,m} g_i(x) \leq 0$$

Lagrangian Funtion

אפשר לכתוב את הבעיה בצורה:

$$\max_{\lambda_i} \min_x f(x) + \lambda_i g_i(x)$$

ואז הלגראנג'יאן הוא:

$$\mathcal{L}(x, \lambda) = f(x) + \lambda_i g_i(x)$$

כאשר λ הוא הLagrangian multiplier.

אם הפתרון האופטימלי גם ככה נמצא באזור שעונה לאילוץ, אז אפשר לקבל מקסימום גם באמצעות $\lambda = 0$ - כלומר האילוץ לא פעיל.

כדי לפתור, נגזור את הלגראנג'יאן לפי כל המרכיבים של הווקטור x ולפי כל המרכיבים של הווקטור λ.

[1] גם המינימיזציה וגם הקביעה שהאילוצים יהיו מהצורה $g_i(x) \leq 0$ הן לצורך אחידות.

מקליד: עידן אריה
מתרגל: ניר בן ארי
תאריך: 11-05-2026

למידת מכונה
89-2511-05

מיקסום האנתרופיה

בד"כ יכולות להיות הרבה התפלגויות שיתאימו לנתונים, ונרצה איזה קריטריון בשביל לבחור איזו התפלגות לבחור. הנחה סבירה היא שאנחנו רוצים להניח כמה שפחות (מעבר למה שאנחנו כבר יודעים)
תזכורת: אנתרופיה של התפלגות P היא:

$$H(P) = -\sum_x P(x) \log P(x)$$

אפשר להסתכל על האנתרופיה כעל "חוסר מידע". ההתפלגות עם הכי הרבה אנתרופיה היא ההתפלגות האחידה[2] - אנחנו לא יודעים כלום. התפלגות עם אנתרופיה אפס היא כאשר יש הסתברות 1 לאחת מהאפשרויות.

אם נרצה למקסם אנתרופיה תחת אילוצים, נשתמש בלגרנג'יאן.

[2] במקרה הסופי - במקרה הרציף, אם יודעים ממוצע וסטיית תקן, זו תהיה ההתפלגות הנורמלית.

מקליד: עידן אריה
מתרגל: ניר בן ארי
תאריך: 11-05-2026

למידת מכונה
89-2511-05

מטריקות להערכת מודלים

להשתמש באחוזי דיוק זה לא תמיד מספיק טוב, כי גם מספרים שנראים גבוהים לא תמיד אומרים הרבה.

למשל: אם יש 9% חולים, והמודל שלנו אומר שכולם בריאים - אז הוא מדייק ב91%. מספר גבוה - אבל לא שווה הרבה.

כשמדברים על קלאסיפיקציה בינארית, יש שני סוגי טעויות:

- אם ניחשנו "חיובי" ובאמת זה שלילי - אז זה false positive.
- אם ניחשנו "שלילי" ובאמת זה חיובי - אז זה false negative.

Precision & Recall

נסמן:

True Positive	TP	
False Positive	FP	
True Negative	TN	
False Negative	FN	

במקום דיוק, נגדיר את הprecision בתור מספר החיוביים הנכונים מתוך כל אלו שאמרנו שהם חיוביים, ואת הrecall בתור כל אלו שאמרנו שהם חיוביים מתוך אלו שהם חיוביים בפועל:

$$Precision = \frac{TP}{TP + FP} \qquad Recall = \frac{TP}{TP + FN}$$

אידיאלית היינו רוצים להגיע ל$Precision = Recall = 1$ - אבל בד"כ אי אפשר, אז נרצה למצוא איזון. נגדיר F_1 (F-score) שמחבר אותם:

$$F_1 = \frac{2}{\text{recall}^{-1} + \text{precision}^{-1}} = 2\frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

ROC Curve

כשיש מודל סיווג עם threshold[3], הthreshold הזה שבוחרים משפיע על כמה המודל יהיה טוב. בשביל להעריך את איכות המודל ללא תלות בסף הנבחר, נרצה לצייר ROC (receiver operating characteristic) curve כאשר:

- ציר הy הוא TPR (True Positive Rates): $\frac{TP}{TP+FN}$.
- ציר הx הוא FPR (False Positive Rates): $\frac{FP}{FP+TN}$.

האלכסון של הגרף הוא "מודל אקראי" - אנחנו רוצים להתרחק ממנו כמה שיותר (בכיוון הנכון)

[3] בנוסף לפרמטרים האחרים של המודל. כלומר המודל נותן מספר, ואז רוצים להשתמש במספר הזה לצרכי סיווג.

Area Under Curve - AUC

מחשבים את השטח מתחת לROC Curve בין $(0,0)$ ל$(1,1)$. זו האיכות של המודל ללא תלות בthreshold - אם ניקח דוגמה שלילית אקראית ודוגמה חיובית אקראית, מה ההסתברות שהערך שהמודל נותן לדגימה החיובית יהיה גדול מהערך שהוא נותן לדגימה השלילית.