

מושגים בתורת האינפורמציה עבור שני משתנים

בהינתן שני משתנים מקריים X, Y , נסתכל על האנטרופיה של ההתפלגות המשותפת $p(X, Y)$ - Joint Entropy:

$$H(X, Y) = - \sum_x \sum_y p(x, y) \log p(x, y) = -E_{p(x, y)} \log p(x, y)$$

הגדרה - אנטרופיה מותנית Conditional Entropy

אנטרופיה מותנית היא ממוצע משוקלל של אנטרופיות $p(y|x)$:

$$\begin{aligned} H(Y|X) &= \sum_{x \in X} p(x) H(y|X=x) = - \sum_x p(x) \sum_y p(y|x) \log p(y|x) = \\ &= - \sum_x \sum_y p(x, y) \log p(y|x) = -E_{p(x, y)} \log p(y|x) \end{aligned}$$

שרשור אנטרופיה

Probability Chaining Rule: $p(x, y) = p(x)p(y|x)$
Entropy Chaining Rule: $H(X, Y) = H(X) + H(Y|X)$ מתקיים:

הוכחה: קלה - פיתוח המשוואות

אפשר להסתכל כהגדרה אלטרנטיבית:

$$H(Y|X) = H(X, Y) - H(X)$$

אינפורמציה הדדית - Mutual Information

מוגדרת כאובדן האינפורמציה במידה ונמדל את $p(x, y)$ ע"י $p(x)p(y)$:

$$I(X; Y) \triangleq D_{KL}[p(x, y) \| p(x)p(y)] = \sum_x \sum_y p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \geq 0$$

שוויון ל-0 כאשר X ו- Y בלתי תלויים.

• החלק $\log \frac{p(x, y)}{p(x)p(y)}$ נקרא point-wise MI (PMI)

• $I(X; Y)$ סימטרי, שכן:

$$\log \frac{p(x|y)}{p(x)} = \log \frac{p(y|x)}{p(y)}$$

מבט אחר על I :

$$\begin{aligned} I(X; Y) &= \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} = \sum_{x,y} \log \frac{p(x|y)}{p(x)} = \\ &= \underbrace{\sum_x \sum_y p(x,y) \log p(x)}_{p(x)} - \left(- \sum_{x,y} p(x,y) \log p(x|y) \right) = \\ &= H(X) - H(x|y) = H(Y) - H(Y - X) \end{aligned}$$

שימוש באינפורמציה החדדית

תזכורת: X נושאים/קטגוריות

Y מסמכים(תצפיות)

$p(x|y)$ הסתברות הסיווג של $y \in Y$ ל $x \in X$, מוגדר לכל $x \in X$

$p(w_k|x_i)$ הסתברות למילה $w \in V$ במסמכים מסוג x_i

המטרה: feature selection - לסנן חלק מהמאפיינים שמייצגים כל דוגמא לסיווג, כדי לאפשר למסווג להתמקד במאפיינים האינפורמטיביים ביותר.

מדד Information Gain

מיועד למדוד את האינפורמציה החדדית בין מאפיין מסויים f לקטגוריה מסויימת x .
נגדיר שני משתנים בינריים:

$X : \{x^+, x^-\}$ מצייין שיוך לקטגוריה של הדוגמה לסיגויות

$F : \{f^+, f^-\}$ מצייין את הופעת/אי-הופעת f

את ההסתברויות נאמוד ממדגם אימון מתוייג.

$$I(X; F) = \sum_{f \in F} \sum_{x \in X} p(f, x) \log \frac{p(f, x)}{p(f)p(x)}$$

השימוש: נחשב $I(F, X)$ לכל המאפיינים במדגם האימון. נשאר רק $\text{top-}k$ או $\text{top-}k\%$ מאפיינים לפי ערך $I(F, X)$.