מקליד: עידן אריה
מרצה: פרופ' עידו דגן
תאריך: 2016-01-20

מודלים הסתברותיים יישומיים במדעי המחשב
89-919-01

מושגים בתורת האינפורמציה עבור שני משתנים

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

$$H\left(X,Y\right)=-\sum_{x}\sum_{y}p\left(x,y\right)\log p\left(x,y\right)=-E_{p(x,y)}\log p\left(x,y\right)$$

הגדרה - אנטרופיה מותנית Conditional Entropy

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

$$H\left(Y|X\right)=\sum_{x\in X}p\left(x\right)H\left(y|X=x\right)=-\sum_{x}p\left(x\right)\sum_{y}p\left(y|x\right)\log p\left(y|x\right)=$$

$$=-\sum_{x}\sum_{y}p\left(x,y\right)\log p\left(y|x\right)=-E_{p(x,y)}\log p\left(y|x\right)$$

שרשור אנטרופיה

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

הוכחה: קלה - פיתוח המשוואות

אפשר להסתכל כהגדרה אלטרנטיבית:

$$H\left(Y|X\right)=H\left(X,Y\right)-H\left(X\right)$$

אינפורמציה הדדית - Mutual Information

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

$$I\left(X;Y\right)\triangleq D_{KL}\left[p\left(x,y\right)\|p\left(x\right)p\left(y\right)\right]=\sum_{x}\sum_{y}p\left(x,y\right)\log\frac{p\left(x,y\right)}{p\left(x\right)p\left(y\right)}\geq0$$

שוויון ל0 כאשר X וY בלתי תלויים.

- החלק $\log\frac{p(x,y)}{p(x)p(y)}$ נקרא point-wise MI(PMI)
- $I\left(X;Y\right)$ סימטרי, שכן:

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

מבט אחר על I:

$$I(X;Y)=\sum_{x,y}p(x,y)\log\frac{p(x,y)}{p(x)\,p(y)}=\sum_{x,y}\log\frac{p(x|y)}{p(x)}=$$

$$=\sum_{x}\underbrace{\sum_{y}p(x,y)}_{p(x)}\log p(x)-\left(-\sum_{x,y}p(x,y)\log p(x|y)\right)=$$

$$=H(X)-H(x|y)=H(Y)-H(Y|X)$$

שימוש באינפורמציה הדדית

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 ההדדית בין מאפיין מסויים f לקטגוריה מסויימת x.
נגדיר שני משתנים בינריים:

$X:\{x^+,x^-\}$ מציין שיוך לקטגוריה של הדוגמה לסיגווית

$F:\{f^+,f^-\}$ מציין 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. נשאיר רק top-k או top-$k\%$ מאפיינים לפי ערך $I(F,X)$.