



تحلیل داده‌های حجیم

مدرس: دکتر ایمان غلامی

[پاییز ۱۴۰۰]

نگارنده: امیرمحمد شعبانی

تمرین سری ۳: سوال ۱

الف

ابتدا معادله را ساده‌تر می‌کنیم:

$$\text{conv}(A \rightarrow B) = \frac{1 - S(B)}{1 - \text{conf}(A \rightarrow B)} = \frac{1 - Pr(B)}{1 - Pr(B|A)} = \frac{Pr(B')}{Pr(B'|A)} = \frac{Pr(B')Pr(A)}{Pr(B' \cap A)}$$

احتمال اینکه B در سبد نباشد به احتمال اینکه وقتی A در سبد باشد، B در سبد نباشد. به عبارتی نرخ رخداد A وقتی B رخ ندهد. به عبارتی در مورد مقداری که ممکن است یک قانون اشتباه باشد حرف می‌زند.

ب

conf

$$\text{conf}(A \rightarrow B) = Pr(B|A) = \frac{Pr(A \cap B)}{Pr(A)}$$

$$\text{conf}(B \rightarrow A) = Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}$$

$$\text{if } \text{conf}(A \rightarrow B) = \text{conf}(B \rightarrow A) \Rightarrow \frac{Pr(A \cap B)}{Pr(A)} = \frac{Pr(A \cap B)}{Pr(B)} \Rightarrow Pr(A) = Pr(B)$$

می‌دانیم به وضوح می‌شود مثالی زد که $Pr(A) \neq Pr(B)$ ؛ پس متقارن نیست.

lift

$$\text{lift}(A \rightarrow B) = \frac{\text{conf}(A \rightarrow B)}{S(B)} = \frac{Pr(A \cap B)}{Pr(A)Pr(B)}$$

$$\text{lift}(B \rightarrow A) = \frac{\text{conf}(B \rightarrow A)}{S(A)} = \frac{Pr(A \cap B)}{Pr(B)Pr(A)}$$

بنابراین متقارن است.

conviction

$$\text{conv}(A \rightarrow B) = \frac{Pr(B')Pr(A)}{Pr(B' \cap A)}$$

$$\text{conv}(B \rightarrow A) = \frac{Pr(A')Pr(B)}{Pr(A' \cap B)}$$

برای نقض آن مثال نقضی می‌آوریم. سبد زیر را در نظر بگیرید:

$$\{B\}, \{A\}, \{A, C\}, \{A, B\}, \{A, B, C\}, \{A, B, D\}, \{C\}$$

$$Pr(B) = \frac{4}{7}, \quad Pr(B') = \frac{3}{7}, \quad Pr(A) = \frac{5}{7}, \quad Pr(A') = \frac{2}{7}$$

$$Pr(B' \cap A) = \frac{2}{7}, \quad Pr(A' \cap B) = \frac{1}{7}$$

$$\text{conv}(A \rightarrow B) = \frac{15}{14}, \quad \text{conv}(B \rightarrow A) = \frac{8}{7}$$

بنابراین متقارن نیست.