[7주차]연관규칙과 확률의 개념

확률(Probability)의 정의

$p :^\wedge f \to [0,1]$

상대도수의 극한

확률은 정의역에서 정의된 함수다.

• 확률(Probability)의 정의

$$P : \mathcal{F} \to [0,1],$$

where $\mathcal{F}$ is a set of events.

(i) $0 \leq P(A) \leq 1$, $A \in \mathcal{F}$

(ii) $P(\phi) = 0$ and $P(\Omega) = 1$,

Ω is a sample space.

(iii) $A_i \in \Omega$, $i = 1, 2, 3, \cdots, n$,

$A_i \cap A_j = \phi$; "disjoint"

$\Rightarrow P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$; "Countable Additivity"

$\Rightarrow A_i$ is a partition of Ω.

확률은 다음의 조건을 만족한다.

1) 전체집합에 속한다

2) 배반사건(즉, 독립이다.)

3) i는 1~n까지 union하면 그때의 확률이 합으로 나온다.

• 조건부 확률의 정의

For events $A, B \in \mathcal{F}$,

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) \neq 0$$

• 사건들의 독립성 (independent)

If $A \perp\!\!\!\perp B$ (독립) then only if $P(A \mid B) = P(A)$

implies

$$\frac{P(A \cap B)}{P(B)} = P(A)$$

$$\therefore P(A \cap B) = P(A) \cdot P(B)$$

P(A|B) = A와 B가 발생할 확률 / B가 발생할 확률,

단P(B) = 0

사건들의 독립성

A,B가 독립일때 P(A|B) = P(A)

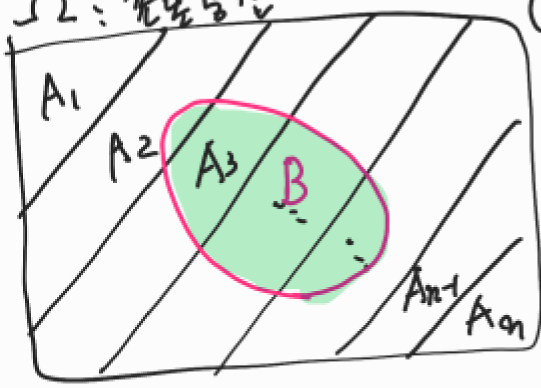
P(A n B) = P(A)P(B)

총확률의 법칙과 베이즈정리

총확률의 법칙과 베이즈정리

• 분할 (Partition) 의 개념.

Ω : 표본공간



(If)

① $A_i \cap A_j = \phi$

② $\bigcup_{i=1}^{n} A_i = \Omega$

for $A_i \subset \Omega$

then we call them (A_i's) a "partition"

즉 $P(B) = ?$

$$B = (A_1 \cap B) \cup (A_2 \cap B) \cup \cdots \cup (A_n \cap B)$$

take the Probability on both side,

$$P(B) = P\left(\bigcup_{i=1}^{n} (A_i \cap B)\right)$$

$$= \sum_{i=1}^{n} P(A_i \cap B), \quad (A_i \cap B) \cap (A_j \cap B) = \phi$$

$$= \sum_{i=1}^{n} P(B \mid A_i) \cdot P(A_i), \quad \text{conditional prob.}$$

$$P(B \mid A_i) = \frac{P(A_i \cap B)}{P(A_i)}$$

- 베이즈 정리

$$P(A_i \mid B) = \frac{P(A_i \cap B)}{P(B)}$$

$$= \frac{P(A_i \cap B)}{\sum_{i=1}^{n} P(A_i \cap B)},$$

by total of probability

확률과 조건부 확률에 근거한 연관규칙

지지도

• 확률과 조건부 확률에 근거한 연관규칙

① 지지도 (support)

$$A \to B \text{의 지지도} = \frac{\text{A와 B를 포함한 거래수}}{\text{전체 거래수}}$$

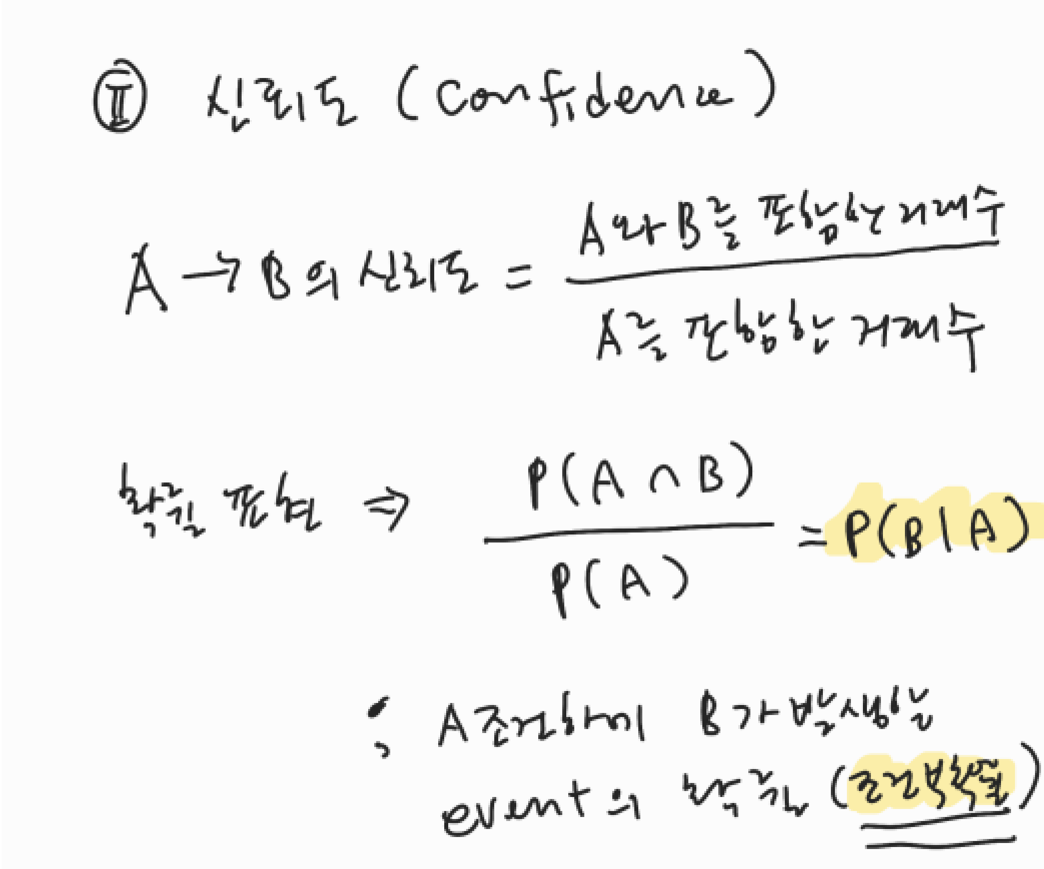
을 확률로 표기하면 $\Rightarrow \dfrac{P(A \cap B)}{P(\Omega)}$

$$= P(A \cap B)$$

$$B \to A \text{의 지지도} = \frac{\text{B와 A를 포함한 거래수}}{\text{전체거래수}}$$

을 확률로 표기하면 $\Rightarrow P(B \cap A)$

신뢰도



Ⅱ 신뢰도 (confidence)

$$A \rightarrow B\text{의 신뢰도} = \frac{\text{A와 B를 포함한 거래수}}{\text{A를 포함한 거래수}}$$

$$\text{확률 표현} \Rightarrow \frac{P(A \cap B)}{P(A)} = P(B|A)$$

: A조건하에 B가 발생할 event의 확률 (조건부확률)

신뢰도는 높지만 지지도가 낮은 경우에는 지지도를 감안해서 해석해야함 (대표성이 부족하기 때문에)

연관규칙을 할때, 세개의 지표를 고려하고, 표본의 대표성을 고려해야함 →(해석시 각 규칙별 정의을 고려해야함)

향상도

③ 향상도 (Lift)

$$A \rightarrow B \text{의 향상도} = \frac{A \rightarrow B \text{의 신뢰도}}{B\text{를 포함하는 거래 비율}}$$

를 확률로 표현하면,

$$\Rightarrow \frac{P(B|A)}{P(B)} = \frac{\frac{P(A \cap B)}{P(A)}}{P(B)}$$

$$= \frac{P(A \cap B)}{P(A) \cdot P(B)}$$

※ $A \perp B$ → $P(A \cap B) = P(A) \cdot P(B)$

$$= \frac{P(A \cap B) / P(B)}{P(A)}$$

$$= \frac{P(A|B)}{P(A)}$$

⇒ B→A 향상도

향상도 = 1 이라는것은

동시에 구매할수밖에 없게하던가

쿠폰을 발행했던가

향상도가 1이거나 1보다작으면 제외해야함

향상도가 1인데 신뢰도가 높으면 이벤트가있었나 생각해봐야함

신뢰도만 A→B, B→A가 다르게 나타날 수 있음

지지도는 필터링역할

향상도 B상품을 구매할때 A에 의존되어야한다.

B가 더 많이 팔리는게 삼성이야 애플이야 → A를 세분화시켜서 → 총확률의법칙, 베이즈정리를 활용