

연관성 분석

가. 다음의 구매이력을 보고 문제 1~4번에 답하시오.

거래번호	제품명
1	넥타이, 셔츠, 양말
2	양말, 벨트, 장갑, 셔츠
3	지갑, 넥타이, 셔츠
4	양말, 벨트, 장갑, 바지

문제 1. 셔츠를 구매한 고객이 넥타이도 함께 구매할 지지도는 얼마인가?

문제 2. 셔츠를 구매한 고객이 양말도 함께 구매할 지지도는 얼마인가?

문제 3. 셔츠를 구매한 고객이 넥타이도 함께 구매할 신뢰도는 얼마인가?

문제 4. 셔츠를 구매한 고객이 넥타이도 함께 구매할 향상도는 얼마인가?

나. 다음의 거래 영수증을 보고 문제 5~6번에 답하시오.

거래번호	제품명
1	달걀, 라면, 참치캔
2	라면, 핫반
3	라면, 소주, 핫반
4	달걀, 라면, 파
5	달걀, 양파, 파

문제 5. 라면을 구매한 고객이 달걀도 함께 구매할 신뢰도를 구하시오

문제 6. 라면을 구매한 고객이 달걀도 함께 구매할 향상도를 구하시오

다. 다음의 거래내역을 보고 문제 7~9번 물음에 답하시오.

거래번호	제품명
1	빵
1	우유
1	콜라
2	빵
2	콜라
3	우유
3	콜라

문제 7. 고객이 빵을 사면 우유도 함께 구매할 지지도를 구하시오.

문제 8. 고객이 빵을 사면 우유도 함께 구매할 신뢰도를 구하시오.

문제 9. 고객이 빵을 사면 우유도 함께 구매할 향상도를 구하시오.

문제 10. 한 상점의 연관분석을 위해 다음과 같은 매트릭스 데이터를 만들었다. 빵을 산 손님이 버터도 함께 구매할 향상도를 구하시오.

거래번호	상품			
	우유	빵	버터	맥주
1	1	1	0	0
2	0	1	1	0
3	0	0	0	1
4	1	1	1	0
5	0	1	1	0

문제 11. 연관성분석에서 유의미한 규칙을 찾아내기 위해 사용되는 측도 중 아래의 설명이 가르키는 것으로 가장 적절한 것은?

항목 A를 포함한 거래 중에서 항목 A와 항목 B가 같이 포함될 확률

- ① 신뢰도(confidence)
- ② 지지도(support)
- ③ 향상도(Lift)
- ④ 빈도(Count)

문제 12. 다음 사례 중에서 지도학습 기법을 사용하여 분석을 수행하기에 적절한 것을 고른 것은?

(a) 고객의 거래 구매 패턴을 분석하여 고객이 구매하지 않은 상품을 추천
(b) 고객의 은행거래 내역을 분석하여 대출 가능 고객인지 아닌지를 분류
(c) 전체 고객의 거래 구매 내역을 분석하여 고객집단을 세분화
(d) 상품을 구매할 때 그와 유사한 상품을 구매한 고객들의 구매 데이터를 분석하여 제시

- ① (a)
- ② (b)
- ③ (c)
- ④ (d)

문제 13. 아래 거래 전표에서 신뢰도가 가장 큰 규칙은?

물품	거래건수
{A}	10
{B}	5
{C}	25
{A, B}	5
{B, C}	20
{A, C}	20
{A, B, C}	15

① $A \Rightarrow C$

② $C \Rightarrow A$

③ $B \Rightarrow C$

④ $C \Rightarrow B$

문제 14. 다음중 아래 거래전표에서 연관규칙 "우유 \Rightarrow 맥주"의 향상도는?

물품	거래건수
빵	100
우유	100
맥주	100
빵, 우유, 맥주	50
우유, 맥주	200
빵, 우유	250
빵, 맥주	200

문제 15. 어떤 항목집합이 빈발하다면, 그 항목집합의 모든 부분집합도 빈발하다는 원리로 연관 규칙 알고리즘 중에서 가장 먼저, 많이 사용되고 있는 알고리즘은?

답안

1. 0.5
2. 0.5
3. 2/3
4. 4/3
5. 0.5
6. 5/6
7. 1/3
8. 1/2
9. 3/4
10. 5/4
11. 1
12. 2
13. 3
14. 25/33
15. 아프리오리(Apriori)