

3과목 - 정형데이터마이닝 인쇄용(PART3)

 tbcase73@gmail.com (공유되지 않음) [계정 전환](#)



* 필수항목

1. 기업이 보유하고 있는 일일 거래 데이터, 고객 데이터, 상품 데이터 혹은 각종 * 1점
마케팅 활동에 있어서의 고객 반응 데이터 등과 이외의 외부 데이터를 포함하는
모든 사용가능한 원천 데이터를 기반으로 감춰진 지식, 기대하지 못했던 경향
또는 새로운 규칙 등을 발견하고 이를 실제 비즈니스 의사결정 등에 유용한 정
보로 활용하는 일련의 작업을 무엇이라 하는가?

- ☐ 분류 분석
- ☐ 데이터 마이닝
- ☐ 기술분석
- ☐ 군집분석

2. 데이터 마이닝 단계 중 목적 변수를 정의하고 필요한 데이터를 데이터 마이닝 * 1점
소프트웨어에 적용할 수 있게 데이터를 준비하는 단계는?

- ☐ 데이터 가공
- ☐ 데이터 준비
- ☐ 검증
- ☐ 데이터 마이닝 기법의 적용



3. 교차판매/물건배치 등에 이용되는 분석 기법은? *

1점

- ☐ 연관분석
- ☐ 주성분분석
- ☐ 회귀분석
- ☐ SOM

4. 다음 중 데이터마이닝 기법 사례와 분석방법이 옳은 것은? *

1점

- ☐ 카탈로그 배열 및 교차판매, 공격적 판촉행사 등의 마케팅 계획 - 연관분석
- ☐ 부모가 있는 어린이의 수를 추정, 가족 구성원의 총수입 추정 - 의사결정나무
- ☐ 생물을 종, 속, 과로 나누는 것, 물질을 요소별로 나누는 것 - 장바구니 분석
- ☐ 시장 세분화 첫 단계로 판촉 활동에 가장 반응률이 높은 고객선별 - 회귀분석

5. 데이터마이닝의 목적 중 사람, 상품에 관한 이해를 증가시키기 위한 것으로 데이터의 특징 및 의미를 표현 및 설명하는 기능을 무엇이라 하는가? * 1점

- ☐ 분류분석
- ☐ 추정
- ☐ 예측
- ☐ 기술



6. 이질적인 모집단을 세분화시키기 위한 방법은? *

1점

- ☐ 연관분석
- ☐ 군집분석
- ☐ 분류분석
- ☐ 주성분분석

1. 로지스틱 회귀모형에서 $\exp(x_1)$ 의 의미는 나머지 변수가 주어질 때 x_1 이 한 단위 증가할 때마다 성공($Y=1$)의 ()가 몇 배 증가하는지를 나타낸다 * 1점

내 답변

2. 종속변수가 성공 또는 실패인 이항변수로 되어 있을 때 종속변수와 독립변수 간의 관계식을 이용하여 두 집단 또는 그 이상의 집단을 분류하고자 할 때 사용되는 분석기법을 무엇이라 하는가? * 1점

- ☐ 로지스틱 회귀분석
- ☐ 다중 회귀분석
- ☐ 의사결정나무
- ☐ 앙상블 모형



3. 반응변수가 범주 형인 경우에 적용되는 회귀분석 모형을 무엇이라 하는가? * 1점

- ☐ 다중 회귀분석
- ☐ 로지스틱 회귀분석
- ☐ 더미 회귀분석
- ☐ 교차분석

1. 아래 표는 불순도 측정 결과이다. 지니지수는 얼마인가? *

1점

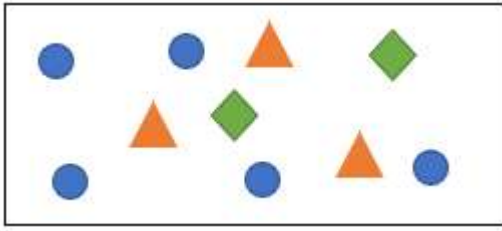


- ☐ 3
- ☐ 2
- ☐ 1
- ☐ 1/2



2. 다음 중 의사결정나무에서 지니지수를 계산한 결과는? *

1점



- ☐ 0.62
- ☐ 0.56
- ☐ 0.44
- ☐ 0.68

3. 다음 중 의사결정 나무의 분류기준에 대한 설명 중 가장 적절하지 않은 것은? * 1점

- ☐ 상위노드로부터 하위노드로 나무 구조를 형성하는 매 단계 분류변수와 분류기준 값의 선택이 중요하다.
- ☐ 지니값(Gini index)이 작을수록 이질적이며, 순수도가 떨어진다.
- ☐ 카이제곱 통계량의 p-값은 그 값이 작을수록 자식 노드 내의 불확실성(이질성)이 큼을 의미한다.
- ☐ 엔트로피 지수는 $p=0.5$ 일 때 이질성이 가장 크다



4. 다음 중 의사결정나무(Decision Tree)에 대한 설명 중 틀린 것은? *

1점

- ☐ 정지규칙이란 더 이상 분리가 일어나지 않고 현재의 마디가 최종마디가 되도록 하는 여러 가지 규칙으로 지니 지수, 엔트로피 지수, 카이제곱통계량 등이 있다
- ☐ 최종마디가 너무 많으면 모형이 과대적합된 상태로 현실 문제에 적용할 수 있는 적절한 규칙이 나오지 않게 되며, 이를 해결하기 위해 가지치기를 한다.
- ☐ 의사결정나무를 위한 알고리즘은 CHAID, CART, ID2, C5.0, C4.5가 있으며 상향식 접근 방법을 이용한다
- ☐ 의사결정나무는 목표변수가 이산형인 경우의 분류나무(classification tree)와 목표변수가 연속형인 경우의 회귀나무(regression tree)로 구분된다.

5. 의사결정나무모형에 관한 내용이다 적절하지 않은 것은? *

1점

- ☐ 의사결정나무의 목적은 새로운 데이터를 분류(classification)하거나 해당 범주의 값을 예측(Prediction)하는 것이다
- ☐ 목표변수 유형에 따라 범주형 분류나무(Classification Tree)와 연속형 회귀나무(Regression Tree)로 분류된다.
- ☐ 분리변수의 P차원 공간에 대한 현재 분할은 이전 분할에 영향을 받지 않는다.
- ☐ 부모마디보다 자식마디의 순수도가 증가하도록 분류나무를 형성해 나간다.

6. 의사결정나무에서 더 이상 분리가 일어나지 않고 현재의 마디가 끝마디가 되도록 하는 규칙을 무엇이라 하는가? *

1점

내 답변



7. 의사결정나무의 형성 과정 중 최종마디가 너무 많으면 모형이 과대적합 상태 * 1점으로 현실문제에 적용할 수 없는 규칙이 나오게 된다. 이러한 과대적합(overfitting) 문제를 해결하기 위해 필요한 것은 무엇인가?

내 답변

8. 의사결정나무 모형에서 가지치기나 정지규칙을 하지 않고 가지를 모두 사용 * 1점했을 때 실제 데이터에 적용할 수 없게 되는 문제점을 무엇이라 하는가?

내 답변

9. 의사결정나무 알고리즘의 설명 중 가장 적절하지 않은 것은? * 1점

- ☐ CHAID는 목표변수가 이산형 일 때, Pearson의 카이 제곱 통계량을 분리기준으로 사용한다.
- ☐ CART는 가장 성취도가 좋은 변수 및 수준을 찾는 것에 중점 둔 알고리즘이다.
- ☐ Information Gain(정보이득)이란 어떤 속성을 선택함으로써 인해서 데이터를 더 잘 구분하게 되는 것을 의미한다.
- ☐ 의사결정나무는 가지치기의 비용함수(Cost Function)를 최대로 하는 분기를 찾아내도록 학습한다.

10. 의사결정나무에서 이산형 목표변수는 지니지수, 연속형 목표변수는 분산 감 * 1점소량을 사용하는 알고리즘은 무엇인가?

- ☐ CHAID
- ☐ CART
- ☐ C4.5
- ☐ C5.0



11. 목표변수가 연속형인 경우 회귀나무의 경우 사용하는 분류기준은 무엇인가? * 1점

- ☐ 카이제곱통계량, 지니지수
- ☐ 지니지수, 엔트로피지수
- ☐ 엔트로피지수, 분산감소량
- ☐ 분산감소량, F-통계량의 p-값

12. 의사결정 나무 결과이다. 해석으로 부적절한 것은? * 1점

- ☐ 끝 노드로 갈수록 불순도가 상승한다
- ☐ 구조가 단순하여 해석이 용이하다
- ☐ 수치형 또는 범주형 변수를 모두 사용할 수 있다
- ☐ 선형성, 정규성, 등분산성 등의 수학적 가정이 불필요한 비모수적 모형이다

1. 다음 중 앙상블 모형이 아닌 것은? * 1점

- ☐ 시그모이드(sigmoid)
- ☐ 배깅(bagging)
- ☐ 랜덤 포레스트(random forest)
- ☐ 부스팅



2. 다음 중 앙상블의 특징으로 가장 적절하지 않은 것은? *

1점

- ☐ 앙상블 학습은 하나의 모델만을 학습시켜 사용하지 않고 여러 모델을 학습시켜 결합하는 방식으로 문제를 처리한다.
- ☐ 약하게 학습 된 여러 모델들을 결합하여 사용하는 것을 앙상블 학습이라 할 수 있다.
- ☐ 성능을 분산시키기 때문에 과적합(overfitting) 감소 효과가 있다
- ☐ 상호 연관성이 높을수록 정확도는 향상이 된다.

3. 보험사에서 해지할 예상 고객을 예측시 사용할 수 있는 적절한 기법은? *

1점

- ☐ 랜덤포레스트
- ☐ 주성분분석
- ☐ 군집분석
- ☐ 연관분석

4. 다음 앙상블 모형 중 매번 분할을 수행할 때마다 설명 변수의 일부분만을 고 * 1점
려함으로 성능을 높이는 방법을 무엇이라 하는가?

- ☐ 배깅
- ☐ 부스팅
- ☐ 랜덤포레스트
- ☐ 의사결정나무



5. 재표본 과정에서 각 자료에 동일한 확률을 부여하지 않고, 분류가 잘못된 데이터에 더 가중을 주어 표본을 추출하는 분석 기법은? * 1점

내 답변

6. 원 데이터부터 집합 크기가 같은 표본의 중복을 허용하고, 복원추출하여 각 표본에 대해 분류기(classifiers)를 생성하는 기법은? * 1점

- ☐ 배깅
- ☐ 부스팅
- ☐ 랜덤포레스트
- ☐ 퍼셉트론

7. 데이터 집합으로부터 크기가 같은 표본을 여러 번 단순 임의 복원 추출하여 각 표본에 대해 분류기를 생성한 후 그 결과를 앙상블하는 방법을 무엇이라 하는가? * 1점

내 답변

8. bootstrap aggregating의 줄임말로 원 데이터 집합으로부터 크기가 같은 표본을 여러 번 단순 임의 복원 추출하여 각 표본에 대해 분류기를 생성한 후 그 결과를 앙상블 하는 방법을 무엇이라 하는가? * 1점

내 답변



9. 부스팅(boosting) 알고리즘 중 Leaf-wise-node 방법을 사용하는 알고리즘을 * 1점
무엇이라 하는가?

- ☐ AdaBoost
- ☐ GBM
- ☐ Xgboost
- ☐ Light GBM

10. lazy learning(게으른 학습)이 사용되는 지도학습 알고리즘을 무엇이라 하는 * 1점
가?

- ☐ kNN
- ☐ SVM
- ☐ 로지스틱회귀
- ☐ 의사결정나무

1. 인공신경망의 특징으로 부적절한 것은? *

1점

- ☐ 분석가의 주관과 경험에 따른다.
- ☐ 입력변수의 속성에 따라 활성화 함수의 선택이 달라진다.
- ☐ 역전파 알고리즘이 동일 입력층에 대해 원하는 값이 출력되도록 개개의 weight를 조정하는 방법으로 사용된다.
- ☐ 이상치 잡음에 민감하지 않다.



2. 시그모이드 함수의 일반화된 형태로 목표치가 다범주인 경우 각 범주에 속할 * 1점
사후 확률을(posterior probability) 제공하는 활성화 함수를 무엇이라 하는가?

내 답변

3. 시그모이드 함수의 범위는? *

1점

- ☐ 0~1
- ☐ -1~1
- ☐ -1~0
- ☐ 0.5~1

4. Logistic 함수라 불리기도 한다. 선형인 멀티퍼셉트론에서 비선형 값을 얻기 * 1점
위해 사용하기 시작했다. 이 함수의 범위는 0~1 값을 가진다. 활성화 함수를 무엇이라 하는가?

- ☐ 시그모이드 함수
- ☐ 계단 함수
- ☐ 소프트맥스 함수
- ☐ 부호 함수



5. 신경망에서 가중치 초기화는 $-1.0 \sim 1.0$ 사이의 임의 값으로 설정하지만, 가중치를 지나치게 큰 값으로 초기화하면 활성화 함수를 편향시키게 되며 활성화 함수가 과적합 되는 상태를 무엇이라 하는가? * 1점

- ☐ 포화상태
- ☐ 불포화 상태
- ☐ 입력가중치
- ☐ 활성화 함수

6. 인공신경망(ANN)의 은닉 노드의 수가 적으면 발생하는 문제점은? * 1점

- ☐ 네트워크가 복잡한 의사결정 경계를 만들 수 없다.
- ☐ 네트워크의 일반화가 어렵다.
- ☐ 레이어(layer)가 적을 때 기울기 소실 문제가 발생한다.
- ☐ 노드의 수가 적으면 과대적합 문제가 발생한다.

7. 다층 신경망에서 은닉층이 많아 인공신경망 기울기 값을 베이스로 하는 역전파 알고리즘으로 학습시키려고 할 때 발생하는 어려움을 무엇이라 하는가? * 1점

- ☐ 기울기 소실문제(Vanishing Gradient Problem)
- ☐ 활성화 함수
- ☐ parameter value 문제
- ☐ 신경망 레이어 소실



8. 인공신경망에서 역전파 알고리즘은 출력층으로부터 하나씩 앞으로 되돌아가 * 1점
며 각 층의 가중치를 수정하는 방법이다. 은닉층이 늘어나면서 기울기가 중간에
0이 되어 버리는 문제를 무엇이라 하는가?

내 답변

9. 다음 중 신용카드 고객 파산여부를 예측하는 모형이 아닌 것은? * 1점

- ☐ 로지스틱회귀분석
- ☐ 선형회귀분석
- ☐ 의사결정나무
- ☐ 앙상블 모형

1. 과적합(overfitting) 발생 여부를 확인하기 위해서는 주어진 데이터의 일정 부 * 1점
분을 모델을 만드는 훈련 데이터로 사용하고, 나머지 데이터를 사용해 모델을
평가한다. 이렇게 데이터를 훈련, 테스트 데이터로 분리하여 검증하는 방법을
무엇이라 하는가?

- ☐ 홀드아웃(Hold-Out)
- ☐ 신경망 모형
- ☐ 향상도 곡선
- ☐ 오분류표



2. 검증용 자료로 사용하는 다음의 방법을 무엇이라 하는가?

1점

주어진 원천 데이터를 랜덤하게 두 분류로 분리하여 교차검정을 실시하는 방법으로 하나는 모형 학습 및 구축을 위한 훈련용 자료로 다른 하나는 성과평가를 위한 검증용 자료로 사용하는 방법

내 답변

3. 붓스트랩 방식을 이용하였을 때 일반적인 훈련 데이터의 양은? *

1점

- ☐ 63.20%
- ☐ 10.20%
- ☐ 23.80%
- ☐ 36.80%



4. 다음의 내장 데이터 Hitters data 분할에 관한 설명 중 적절하지 않은 것은? * 1점

```
Hitters = read.csv('../data/Hitters.csv', fileEncoding='UTF-8-BOM')
data("Hitters")
hitters <- Hitters[, -1]
set.seed(1234)
idx <- sample(2, nrow(hitters), replace=TRUE, prob=c(0.7, 0.3))
traindata <- hitters[idx==1,]
testdata <- hitters[idx==2,]
```

- ☐ sample()은 데이터의 무작위 재배열보다는 데이터에서 일부를 train/test data로 분리하는데 유용하게 사용된다
- ☐ Hitters dataset 70%는 train data로 분할된다
- ☐ 반복 수행 시 매 회 다른 데이터분할이 샘플링 된다
- ☐ 데이터 분할을 하는 이유는 주어진 데이터에서만 높은 성과를 보이는 모형의 과적합 문제를 해결하기 위한 단계로 잘못된 가설을 가정하게 되는 2종 오류의 발생을 방지하는 데 있다

5. 데이터 분할에 관한 설명 중 적절하지 않은 것은? * 1점

- ☐ 훈련 데이터에 대한 학습만을 바탕으로 모델의 설정(Hyperparameter)를 튜닝하게 되면 과대적합(overfitting)이 일어날 가능성이 매우 크다.
- ☐ 모델이 너무 간단하여 정확도가 낮은 모델을 과소적합(Underfitting) 되었다고 말한다.
- ☐ 과대적합이나 과소적합의 문제를 최소화하고 모델의 정확도를 높이는 가장 좋은 방법은 더 많고 다양한 데이터를 확보하고, 확보한 데이터로부터 더 다양한 특징(feature)들을 찾아서 학습에 사용하는 것이다.
- ☐ test set 결과가 일반적으로 training set 결과보다 좋다.



6. 분류모형에서 일부 범주형의 관측치가 현저히 부족하여 모형이 학습하기 힘든 문제를 무엇이라 하는가? * 1점

- ☐ 클래스 불균형
- ☐ 집중 샘플링
- ☐ ROSE 샘플링
- ☐ 차원 축소

1. 아래 confusion matrix의 오분류율(error rate)을 구하면? *

1점

Confusion matrix		예측값	
		TRUE	FALSE
실제값	TRUE	40	60
	FALSE	60	40

- ☐ 0.6
- ☐ 0.4
- ☐ 0.5
- ☐ 0.3



2. 보기의 표를 보고 정분류율(accuracy)를 구하시오

1점

Confusion matrix		Predicted class	
		1	0
Actual class	1	a	b
	0	c	d

내 답변

3. 오분류표 중 정밀도와 재현률의 조화평균을 나타내며 정밀도와 재현율에 같은 가중치를 부여하여 평균한 지표를 무엇이라 하는가? * 1점

- ☐ F1
- ☐ Precision
- ☐ Recall
- ☐ Specificity



4. [오분류표]를 이용하여 F1값을 구하여라 (소수점 아래 셋째 자리에서 반올림 * 1점 하여 둘째 자리까지 기입하세요)

n=165	Predicted:		
	NO	YES	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

내 답변

5. 다음 오분류표를 이용하여 F1의 값을 구하면? *

1점

Confusion matrix		예측값	
		TRUE	FALSE
실제값	TRUE	40	60
	FALSE	60	40

- ☐ 0.3
☐ 0.4
☐ 0.15
☐ 0.5



6. 다음 오분류표를 이용하여 재현율을 정의하면? *

1점

confusion matrix		실제값	
		Y	N
예측값	Y	True Positive	False Positive
	N	False Negative	True Negative

- ☐ $TP/(TP+FN)$
- ☐ $TP/(TP+FP)$
- ☐ TP/FP
- ☐ TN/FN

7. 오분류표를 활용한 평가지표 F1지표 중 민감도(sensitivity)와 같은 지표는? * 1점

- ☐ recall
- ☐ specificity
- ☐ precision
- ☐ kappa



8. 다음 중 오분류표의 평가지표 중 True로 예측한 관측치 중 실제 True인 지표를 무엇이라 하는가? * 1점

- ☐ Precision
- ☐ Accuracy
- ☐ F1
- ☐ Kappa

9. 두 평가자의 평가가 얼마나 일치하는지 평가하는 값으로 0~1 사이의 값을 가진다. P(e)는 두 평가자의 평가가 우연히 일치할 확률을 뜻하는 모델 평가 메트릭을 무엇이라 하는가? * 1점

내 답변

10. 오분류표의 평가지표 중 F2의 의미는? *

1점

- ☐ 재현율(Recall)에 정밀도(Precision)에 1/4배의 가중치를 부여
- ☐ 재현율(Recall)에 정밀도(Precision)에 4배의 가중치를 부여
- ☐ 재현율(Recall)에 정밀도(Precision)에 1/2만큼 가중치를 부여
- ☐ 재현율(Recall)에 정밀도(Precision)에 2배만큼의 가중치를 부여



11. 아래 오분류표를 이용하여 구한 특이도 값은 얼마인가?

1점

Confusion matrix		예측값	
		TRUE	FALSE
실제값	TRUE	40	60
	FALSE	60	40

- ☐ 0.4
- ☐ 0.3
- ☐ 0.15
- ☐ 0.5

12. ROC 그래프의 X축에는 FP Ratio, Y축은 민감도를 나타낸다. 가장 이상적인 * 1점
X, Y 축의 값은?

- ☐ 0, 1
- ☐ 1, 0
- ☐ 1, 1
- ☐ 0, 0

13. FP-Ratio(1-특이도), 민감도를 나타내어 이 두 평면 값의 관계로 하는 모형 평가를 무엇이라 하는가? 1점

내 답변



1. K-평균 군집화와 달리 군집 수 k를 설정할 필요가 없는 탐색적 모형을 무엇이 * 1점
라 하는가?

- ☐ 계층적 군집분석
- ☐ 비계층적 군집분석
- ☐ 판별분석
- ☐ 밀도기반 군집분석

2. 군집간의 거리에 따라 데이터를 연결하기 보다는 군집내 편차들의 제곱합에 * 1점
근거를 두고 군집들을 병합시키는 방법을 무엇이라 하는가?

내 답변

3. 계층적 군집에서 군집내의 오차제곱합에 기초하여 군집을 수행하는 군집 방 * 1점
법은 무엇인가?

- ☐ 단일연결법
- ☐ 완전연결법
- ☐ 평균연결법
- ☐ 와드연결법



4. 비계층적 군집분석인 K-means 군집분석의 수행순서는? *

1점

가) 초기 군집의 중심으로 k 개의 객체를 임의로 선택한다
 나) 각 자료를 가장 가까운 군집 중심에 할당한다
 다) 각 군집 내의 자료들의 평균을 계산하여 군집의 중심을 갱신한다
 라) 군집 중심의 변화가 거의 없을 때까지 나와 다를 반복한다

- ☐ 가-나-다-라
- ☐ 나-가-다-라
- ☐ 다-나-가-라
- ☐ 라-가-나-다

5. 군집내 거리와 군집 간의 거리를 기준으로 군집 분할 성과를 측정하는 방식으 * 1점
 로 클러스터 안의 데이터들이 다른 클러스터와 비교해서 얼마나 비슷한가를 나
 타내는 군집평가를 무엇이라 하는가?

- ☐ Silhouette
- ☐ Scree Plot
- ☐ F1 Score
- ☐ ROC Curve

6. 군집분석에서 중요한 지표로서, 거리가 가까울수록 높고 멀수록 낮은 지표이 * 1점
 자 완벽히 분리된 경우 1이 되는 지표는?

내 답변



7. 어느 점을 기준으로 반경 x 내에 점이 n 개 이상 있으면 하나의 군집으로 인식 하는 방식을 의미하며, 임의적 모양의 군집분석을 무엇이라 하는가? * 1점

- ☐ kNN
- ☐ PAM
- ☐ K-means
- ☐ DBSCAN

☐ 옵션 1

☐ 옵션 1

8. 아래 데이터 세트(dataset)에서 a, b 간의 맨해튼 거리는 얼마인가? * 1점

구분	a	b
Score	90	80
Time	60	75

- ☐ 25
- ☐ 20
- ☐ 15
- ☐ 10



9. 거리를 활용한 측도에 대한 설명으로 틀린 것은? *

1점

- ☐ 유클리드는 두 점 사이의 거리로, 가장 직관적이고 일반적인 거리의 개념이다
- ☐ 맨해튼 거리는 두 점의 좌표 간의 절대값 차이를 구하는 것이다
- ☐ 마할라노비스는 변수의 표준화를 고려하고, 변수 간의 상관성을 고려하지 않는다
- ☐ 표준화, 마할라노비스 거리는 통계적 거리의 개념이다

10. 거리에 대한 개념으로 두 벡터 사이의 사잇각을 계산해서 유사한 정도를 구하는 것을 무엇이라 하는가? *

1점

- ☐ 코사인 유사도
- ☐ binary
- ☐ 민코프스키
- ☐ 표준화 거리

11. 계층적 군집은 두 개체 간의 거리에 기반하므로 거리측정에 대한 정의가 필요하다. 다음 중 `dist()` 함수에서 지원하지 않은 거리는?

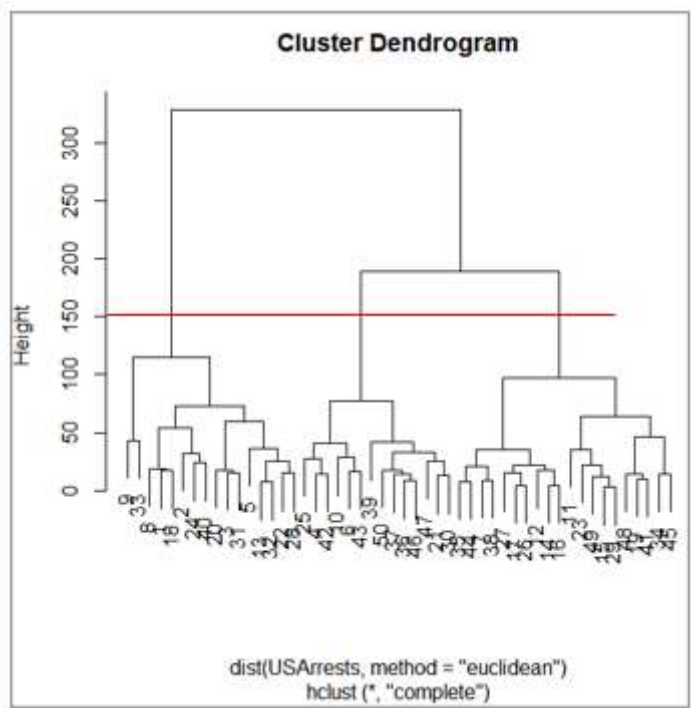
1점

- ☐ 유클리드
- ☐ 맨하튼
- ☐ 민코프스키
- ☐ cosine



12. 덴드로그램에서 Height 150에서 나눌 경우 군집의 수는? *

1점



내 답변

13. k 평균 군집에 대한 설명 중 적절하지 않은 것은? *

1점

- ☐ 초기값 선택이 최종 군집 선택에 영향을 미친다.
- ☐ 초기 군집수를 결정하기 어렵다.
- ☐ 한 개체가 속해 있던 군집에서 다른 군집으로 이동해 재배치가 가능하지 않다.
- ☐ 각 군집내의 자료들의 평균을 계산하여 군집의 중심을 갱신한다.



14. 다음 군집분석(Cluster analysis) 관한 설명 중 옳바르지 않은 것은?

1점

- ☐ 비계층적 군집분석 기법의 경우 사용자가 사전 지식 없이 그룹의 수를 정해주는 일이 많기 때문에 결과가 잘 나오지 않을 수 있다.
- ☐ 군집분석은 신뢰성과 타당성을 점검하기 어렵다
- ☐ 군집 결과에 대한 안정성을 검토하는 방법으로 지도학습과 동일한 교차타당성을 이용한다.
- ☐ 계층적 군집분석은 이상치에 민감하다.

1. SOM에 대한 설명으로 가장 적절한 것은? *

1점

- ☐ 지도학습이다
- ☐ 인공신경망과 같은 역전파 알고리즘을 이용한다.
- ☐ 다수의 입력층과 다수의 출력층으로 구성이 되어 있다.
- ☐ 출력 뉴런들은 승자 뉴런이 되기 위해 경쟁하고 오직 승자만이 학습한다.

2. SOM에 대한 설명이 잘못된 것은? *

1점

- ☐ SOM은 역전파 알고리즘을 사용한다.
- ☐ 차원축소와 군집화를 동시에 수행하는 기법이다
- ☐ 주요 기능 중에 데이터의 특징을 파악하여 유사 데이터를 클러스터링한다.
- ☐ 대표적인 비지도학습이다.



3. 기법 활용 분야가 다른 것은? *

1점

- ☐ SOM
- ☐ 로지스틱 회귀분석
- ☐ 신경망
- ☐ 의사결정 나무

1. 연관분석의 특징으로 옳지 않은 것은? *

1점

- ☐ 조건반응(if then)으로 표현되는 연관분석의 결과를 이해하기 쉽다.
- ☐ 비목적성 분석 기법이다.
- ☐ 세분화 분석 품목 없이 연관 규칙을 찾을 수 있다.
- ☐ 분석 계산이 간편하다.

2. 조건-결과(if-then) 유형의 패턴을 발견하는데 사용하는 데이터마이닝 기법은? * 1점

- ☐ SOM
- ☐ 연관규칙
- ☐ 다차원척도
- ☐ 의사결정나무



3. 연관분석의 대표적 알고리즘 apriori의 단점을 보완하기 위해 트리와 노드 링 * 1점
크라는 특별한 자료 구조를 사용하는 알고리즘은?

- ☐ FP-Growth
- ☐ arules
- ☐ kohonen
- ☐ spade

4. 교차판매/물건배치 등에 이용되는 분석 기법은? *

1점

- ☐ 연관분석
- ☐ 주성분분석
- ☐ 회귀분석
- ☐ SOM



5. 빵 --> 우유의 신뢰도는? *

1점

장바구니	item
1	빵, 맥주, 우유
2	빵, 우유, 계란
3	맥주, 우유
4	빵, 맥주, 계란
5	빵, 맥주, 우유, 계란

- ☐ 0.75
- ☐ 0.65
- ☐ 0.6
- ☐ 0.7

6. A 제품구입 -> B 제품 구입의 연관규칙 측정지표 중 지지도(support)란? *

1점

- ☐ A와 B가 동시에 포함된 거래수/전체 거래수
- ☐ A와 B가 동시에 포함된 거래수/A가 포함된 거래수
- ☐ A와 B가 동시에 포함된 거래수/B가 포함된 거래수
- ☐ A가 포함된 거래수/전체 거래수



7. 아래의 거래데이터에서 추출된 연관규칙 중 하나인 콜라 --> 맥주의 지지도 * 1점은?
는?

거래번호	판매상품
1	소주, 콜라, 맥주
2	소주, 콜라, 와인
3	소주, 주스
4	콜라, 맥주
5	주스

내 답변

8. 연관규칙의 측정지표 중 전체 거래 중에서 품목 A, B가 동시에 포함되는 거래 * 1점의 비율을 무엇이라 하는가?

- ☐ 지지도
- ☐ 향상도
- ☐ 신뢰도
- ☐ ROC



9. 어떤 슈퍼마켓 고객 6명의 장바구니별 구입품목이 다음과 같다고 하자, 연관 * 1점 규칙(콜라-->맥주)의 지지도는?

거래번호	판매상품
1	소주, 콜라, 맥주
2	소주, 콜라, 와인
3	소주, 주스
4	콜라, 맥주
5	소주, 콜라, 맥주, 와인
6	주스

- ☐ 0.6
- ☐ 0.4
- ☐ 0.5
- ☐ 0.3



10. 연관규칙 "딸기 -> 사과 " 의 향상도는? 다음은 쇼핑물의 거래 내역이다. * 1점

항목	거래수
딸기	100
사과	100
배	50
[딸기, 사과]	500
[딸기, 배]	300
[사과, 배]	200
[딸기, 사과, 배]	100
전체 거래건수	1450

- ☐ 0.96
- ☐ 0.75
- ☐ 0.90
- ☐ 0.70



11. 다음은 데이터베이스에 있는 속성(Attribute)들 간의 연관성 결과이다. 이에 * 1점
대한 설명 중 가장 적절한 것은?

item	구매항목
1	햄버거, 피자, 라면
2	라면, 치킨, 김밥, 피자
3	치킨, 햄버거, 피자
4	라면, 치킨, 김밥

- ☐ 햄버거 → 피자의 향상도가 1이 넘으면 햄버거와 피자 사이에 연관성이 높다고 할 수 있다
- ☐ 향상도 1이면 햄버거와 피자는 서로 독립이 아닌 관계이다
- ☐ 향상도가 1보다 작으면 햄버거와 피자는 서로 양의 상관관계를 의미한다
- ☐ 햄버거 → 피자의 지지도는 햄버거와 피자가 동시에 포함된 거래수를 분자로 햄버거를 포함하는 거래수를 분모로 나타낼 수 있다

12. 연관 규칙 지표에 대한 설명 중 옳은 것은? *

1점

- ☐ 향상도가 1보다 크면 이 규칙은 결과를 예측하는 데 있어서 우수하다는 것을 의미
- ☐ 향상도가 1이면 두 품목 간에 상호 연관성이 높다는 것으로 해석한다
- ☐ 향상도가 1보다 작으면 품목 간에 독립적인 관계로 본다
- ☐ 지지도는 조건부확률로 품목 a를 구매한 사람이 품목 b도 구매한다고 해석한다 (a → b)



13. 연관규칙의 향상도 설명이 옳은 것은? *

1점

- ☐ 향상도가 1보다 크면 이 규칙은 결과를 예측하는 데 있어서 우수하다는 것을 의미
- ☐ 향상도가 1이면 두 품목은 연관성이 높다는 의미
- ☐ 향상도가 1보다 작으면 두 품목은 서로 양의 상관관계를 의미
- ☐ 품목 A가 포함된 거래 중에서 품목 A, B를 동시에 포함하는 거래일 확률

14. 품목B를 구매한 고객 대비 품목 A를 구매한 후 품목 B를 구매하는 고객에 대 * 1점
한 확률을 의미하는 연관규칙의 측정지표를 무엇이라 하는가?

내 답변

제출

양식 지우기

Google Forms를 통해 비밀번호를 제출하지 마세요.

이 콘텐츠는 Google이 만들거나 승인하지 않았습니다. [악용사례 신고](#) - [서비스 약관](#) - [개인정보처리방침](#)

Google 설문지

