

2020년 - C형 인쇄용(50문제)

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



* 필수항목

1. 빅데이터가 만들어 내는 본질적인 변화에 대한 설명 중 적절하지 않은 것은? * 1점

- ☐ 사전처리에서 사후 처리시대로의 변화
- ☐ 표본조사에서 전수조사로의 변화
- ☐ 질보다 양을 강조하는 변화
- ☐ 상관관계에서 인과관계로의 변화

2.기업의 성과분석 현황에 대한 설명 중 적절하지 못한 것은? * 1점

- ☐ 성과가 높은 기업과 성과가 낮은 기업이 큰 차이를 보이는 부분은 분석에 대한 태도와 분석의 응용 부분이다.
- ☐ 성과가 높은 기업들이 가치 분석적 통찰력을 갖췄다고 대답한 비율이 낮다.
- ☐ 성과가 높은 기업도 분석 역량을 활용하지 못하고 있다.
- ☐ 성과가 낮은 기업들은 일상 업무에 데이터 분석을 활용하지 못하고 있다.



3. 다음 중 데이터 사이언티스트의 필요 역량으로 가장 부적절한 것은 무엇인가? * 1점

- ☐ 통찰력 있는 분석
- ☐ 설득력 있는 전달
- ☐ 다 분야간 협력
- ☐ 네트워크 최적화

4. 다음 데이터의 특징 중 가장 옳바르지 않은 것은? * 1점

- ☐ 암묵지는 시행착오와 오랜 경험을 통해 개인에게 습득된 무형의 지식이다.
- ☐ 데이터는 존재론적 특징과 함께 당위적 특징의 성격을 가지고 있다.
- ☐ 데이터는 개별 데이터 자체로는 의미가 중요하지 않은 객관적인 사실을 의미한다.
- ☐ 데이터의 유형은 암묵지와 형식지로 구분한다.

5. 데이터에 관한 구조된 데이터로 다른 데이터를 설명해 주는 데이터를 무엇이 라 하는가? * 1점

- ☐ 메타데이터
- ☐ 데이터 사전
- ☐ 데이터웨어하우스
- ☐ 데이터베이스



6. 다음은 빅데이터 활용 기술에 관한 설명이다. 적절하지 않은 것은? *

1점

- ☐ 택배 차량을 어떻게 배치하는 것이 비용에 효율적인가?: 유형분석
- ☐ 응급실에서 의사를 어떻게 배치하는 것이 가장 효율적인가?: 유전알고리즘
- ☐ 맥주 구매자가 기저귀를 더 많이 구매하는가?: 연관분석
- ☐ 사용자의 만족도가 충성도에 어떤 영향을 미치는가?: 회귀분석

7. 다음 중 딥러닝(Deep Learning)과 관련된 분석기법은? *

1점

- ☐ ANN
- ☐ 로지스틱 회귀분석
- ☐ 연관분석
- ☐ 주성분분석

8. 아래는 용어와 의미를 서로 연결한 것이다. 다음 중 용어-의미가 잘못 연결된 것을 모두 나열한 것은? *

1점

- OLTP - 다차원의 데이터를 대화식으로 분석하기 위한 소프트웨어
- BI(business Intelligence) - 경영 의사결정을 위한 통계적이고 수학적인 분석에 초점을 둔 기법
- BA(Business Analytics) - 데이터 기반 의사결정을 지원하기 위한 리포트 중심의 도구
- Data Mining - 대용량 데이터로부터 의미 있는 관계, 규칙, 패턴을 찾는 과정

- ☐ OLTP
- ☐ OLTP, BI
- ☐ OLTP, BI, BA
- ☐ OLTP, BI, BA, Data Mining



9. 빅데이터가 만들어 내는 본질적인 변화에 대한 설명이다. 1, 2에 적절한 단어 * 1점은 무엇인가? 1. XXXX, 2. XXXX 로 답해 주세요.

(1) 은 어떤 현상에 대하여 현상을 발생시킨 원인과 그 결과 사이의 관계를 말하고,
(2) 는 어떤 두 현상이 관계가 있음을 말하지만 어느 쪽이 원인인지 알 수 없다.

내 답변

10. 빅데이터 비즈니스 측면에서 “공동 활용의 목적으로 구축된 유무형의 구조 * 1점 물”을 의미하는 빅데이터 기능을 무엇이라 하는가?

내 답변

1. 메타 데이터와 데이터 사전의 관리 원칙을 수립하고, 빅데이터의 경우 데이터 * 1점 생명주기 관리방안 수립에 해당되는 데이터 거버넌스 체계를 무엇이라 하는가?

- ☐ 데이터 표준화
- ☐ 데이터 관리체계
- ☐ 데이터 저장소 관리
- ☐ 표준화 활동



2. 마스터플랜 수립 시 적용 범위 및 방식의 고려요소가 아닌 것은? *

1점

- ☐ 업무 내재화 적용 수준
- ☐ 분석 데이터 적용 수준
- ☐ 투자 비용 수준
- ☐ 기술 적용 수준

3. CRISP-DM 분석절차에서 “위대한 실패”가 발생하는 구간은? *

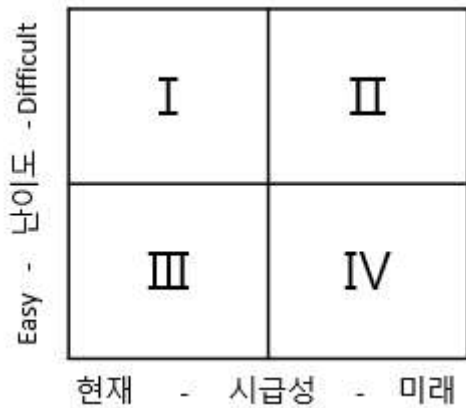
1점

- ☐ Evaluation – Business Understanding
- ☐ Modeling – Data Preparation
- ☐ Data Preparation – Data Understanding
- ☐ Evaluation - Deployment



4. 다음 중 포트폴리오 사분면 분석에 대한 설명 중 적절하지 않은 것은? *

1점



- ☐ 사분면 분석에서 가장 우선적인 분석 과제 적용이 필요한 영역은 I 사분면이다.
- ☐ 분석 과제를 바로 적용하기 어려워 우선순위가 낮은 영역은 II 사분면이다.
- ☐ 시급성에 기준을 둔다면, III → IV → II 사분면 순이다.
- ☐ 난이도에 기준을 둔다면, III → I → II 사분면 순이다.

5. 다음 중 데이터 거버넌스에 대한 설명 중 적절하지 않은 것은? *

1점

- ☐ 데이터 거버넌스의 구성요소는 원칙, 조직, 프로세스 이다.
- ☐ 기업은 데이터 거버넌스 체계를 구축함으로써 데이터의 가용성, 유용성, 통합성, 보안성, 안정성 등을 확보할 수 있다.
- ☐ 데이터 거버넌스란 모든 데이터에 대하여 정책 및 지침, 표준화, 운영조직 및 책임 등의 표준화된 관리체계를 수립하고 운영하는 프레임워크 및 저장소 구축을 의미한다.
- ☐ 데이터 거버넌스는 독자적으로 수행해야만 하고 전사 차원의 IT 거버넌스나 EA의 구성요소로 구축되는 경우는 없다.



6. CRISP - DM 분석방법론의 데이터 준비 단계의 Task가 아닌 것은? *

1점

- ☐ 데이터 정제
- ☐ 데이터 통합
- ☐ 데이터 탐색
- ☐ 분석용 데이터셋 선택

7. 분석 과제 발굴의 접근방식에 대한 설명 중 적절하지 않은 것은? *

1점

- ☐ 문제가 무엇인지(what)를 알고 답을 구하는 방식을 상향식 접근방식이라 한다.
- ☐ 디자인 씽킹 프로세스는 상향식 접근방식의 발산과 하향식 접근방식의 수렴단계를 반복적으로 수행하게 된다.
- ☐ 분석과제발굴의 상향식과 하향식 접근방법은 실제 분석과정에서 혼용되어 활용되는 경우가 많다
- ☐ 데이터를 활용하여 생각하지 못했던 인사이트를 도출하고 시행착오를 통해서 개선해 가는 상향식 접근방식의 유용성이 점차 증가하고 있는 추세이다.

8. 분석 기획 고려 사항 중 장애 요소에 대한 부적절한 설명은? *

1점

- ☐ 데이터 유형에 따라서 적용 가능한 솔루션 및 분석 방법이 다르기 때문에 유형에 대한 분석이 선행적으로 이루어져야 한다.
- ☐ 유사 분석 시나리오 및 솔루션이 있다면 이를 최대한 활용하는 것이 중요하다.
- ☐ 장애요소들에 대한 사전 계획 수립이 필요하다.
- ☐ 이해하기 쉬운 모델보다는 복잡하고 정교한 모형이 더 효과적이다.



9. 빈 칸에 알맞은 용어는? *

1점

식별된 비즈니스 문제를 데이터의 문제로 변환하여 정의하는 단계이다. 앞서 수행한 문제탐색의 단계가 무엇을 어떤 목적으로 수행해야 하는지에 대한 관점이었다면 () 단계에서는 이를 달성하기 위해 필요한 데이터 및 기법을 정의하기 위한 데이터 분석의 문제로의 변환을 수행하게 된다.

내 답변

10. 아래 내용은 데이터 분석 단계의 어떤 Task에 대한 설명인가? *

1점

분석용 데이터를 이용한 가설 설정을 통하여 통계모델을 만들거나 기계학습을 이용한 데이터의 분류, 예측, 군집 등의 기능을 수행하는 과정을 의미한다.

내 답변

1. 아래의 불순도 측정 결과를 사용해서 구한 지니 지수는 얼마인가? *

1점



- ☐ 0.5
- ☐ 0.32
- ☐ 0.48
- ☐ 0.38



2. 어떤 슈퍼마켓에서 고객 5명의 장바구니 구입품목이 다음과 같다고 한다. 연 * 1점
관규칙 딸기→사과에 대한 지지도는?

구입품목	거래건수
딸기, 배	100
사과	200
딸기, 사과	150
수박, 배, 사과	250
메론, 딸기, 사과	300

- ☐ 45%
- ☐ 65%
- ☐ 15%
- ☐ 25%

3. 다음 확률분포에서 확률변수 x 의 기댓값은? *

1점

x	1	2	3
$f(x)$	$1/6$	$1/2$	$1/3$

- ☐ $13/6$
- ☐ $4/6$
- ☐ 2
- ☐ 1



4. 다음 중 아래 오분류표를 이용하여 F1값 구하면? *

1점

오분류표		예측치	
		TRUE	FALSE
실제값	TRUE	40	60
	FALSE	60	40

- ☐ 0.4
- ☐ 0.6
- ☐ 0.8
- ☐ 1.0

5. Hitters dataset의 일부이다. 다음 설명 중 적절하지 않은 것은? *

1점

> summary(Hitters)

```

      AtBat      Hits      HmRun    NewLeague      Salary
Min.   : 16.0   Min.   :  1   Min.   : 0.00   A:176   Min.   :  67.5
1st Qu.:255.2   1st Qu.: 64   1st Qu.: 4.00   N:146   1st Qu.: 190.0
Median :379.5   Median : 96   Median : 8.00           Median : 425.0
Mean   :380.9   Mean   :101   Mean   :10.77           Mean   : 535.9
3rd Qu.:512.0   3rd Qu.:137   3rd Qu.:16.00           3rd Qu.: 750.0
Max.   :687.0   Max.   :238   Max.   :40.00           Max.   :2460.0
NA's   :59

```

- ☐ Salary 변수 분포는 왼쪽꼬리가 긴 분포를 가진다.
- ☐ NewLeague 변수는 범주형 자료이다.
- ☐ Hits 변수에는 결측값이 없음을 알 수 있다.
- ☐ HmRun 변수의 최대값은 40이다.

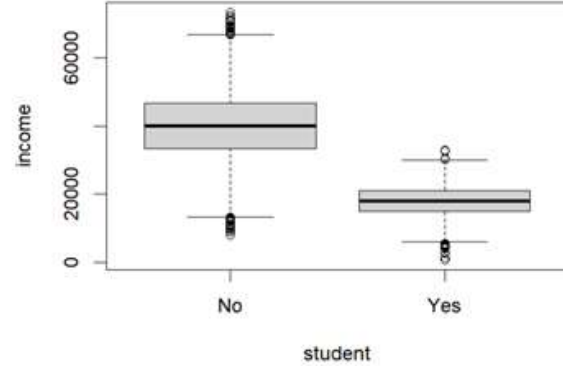
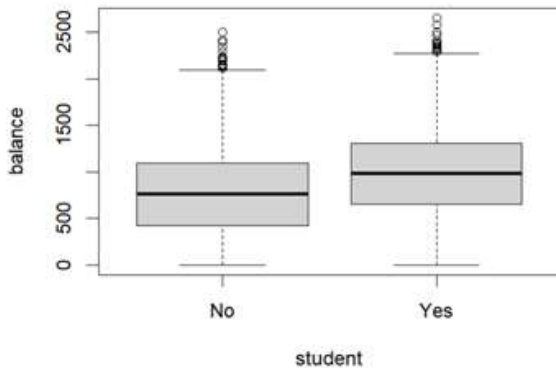


6. Default dataset에 대한 설명 중 적절하지 않은 것은? *

1점

```
> summary(Default)
```

default	student	balance	income
No :9667	No :7056	Min. : 0.0	Min. : 772
Yes: 333	Yes:2944	1st Qu.: 481.7	1st Qu.:21340
		Median : 823.6	Median :34553
		Mean : 835.4	Mean :33517
		3rd Qu.:1166.3	3rd Qu.:43808
		Max. :2654.3	Max. :73554



- ☐ 학생인 경우가 학생이 아닌 경우보다 balance가 낮은 경향을 보인다.
- ☐ 학생인 경우 income이 학생이 아닌 경우 보다 편차가 작다
- ☐ default와 student 변수는 범주형 데이터이다
- ☐ balance, income에 이상치가 있음을 알 수 있다

7. 군집 간의 거리에 기반하는 다른 연결법과는 달리 군집 내의 오차 제곱합 (error sum square)에 기초하여 군집을 수행하는 계층적 군집분석의 거리측정을 무엇이라 하는가?

* 1점

- ☐ 중심연결법
- ☐ 평균연결법
- ☐ 와드연결법
- ☐ 최단연결법



8. 이산형 확률변수 x 의 기댓값은? *

1점

- ☐ $E(x) = \sum x f(x)$
- ☐ $E(x) = [x f(x)]$
- ☐ $E(x) = E[(x - \mu)^2]$
- ☐ $E(x) = x^3 - x^2$

9. 다음 2개의 좌표에 대한 맨해튼 거리를 구하면? *

1점

$a(100, 5), b(50, 7)$

- ☐ 52
- ☐ 43
- ☐ 138
- ☐ 95

10. 다음 중 순위만 제공할 뿐 양적인 비교는 불가능한 척도는 무엇인가? *

1점

- ☐ 명목척도
- ☐ 순서척도
- ☐ 구간척도
- ☐ 비율척도



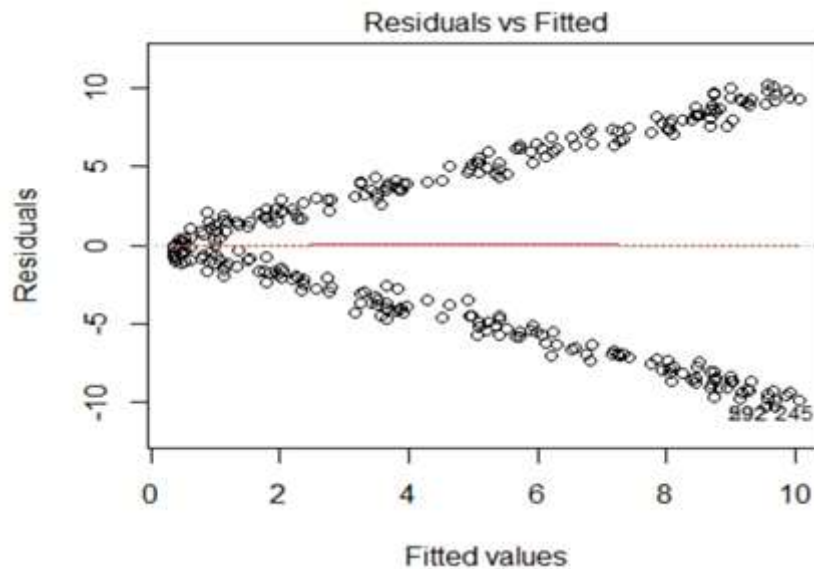
11. 다음 두 개의 확률변수 X, Y의 공분산에 대한 설명 중 옳지 않은 것은? *

1점

- ☐ 공분산이 양수이면 X가 증가할 때 Y도 증가한다.
- ☐ 공분산이 음수이면 X가 증가할 때 Y는 감소한다.
- ☐ 공분산이 0이면 두 변수간에는 아무런 선형관계가 없으며 두 변수는 서로 독립적인 관계이다.
- ☐ $-1 \leq \text{Cov}(X, Y) \leq 1$, 공분산의 크기는 -1~1 사이의 범위를 갖는다.

12. 다음 잔차도는 회귀분석의 어떤 가정을 위배하고 있는가? *

1점



- ☐ 선형성
- ☐ 비상관성
- ☐ 등분산성
- ☐ 독립성



13. 다음 중 중심극한정리에 대한 설명 중 적절하지 않은 것은? *

1점

- ☐ 모집단의 분포가 정규분포에 가까워져야 표본평균의 분포가 정규분포로 근사하게 된다
- ☐ 중심극한정리가 성립하기 위해서는 표본크기가 최소 30 이상이어야 한다.
- ☐ 모집단의 분포와 상관없이 표본의 크기가 커짐에 따라 정규분포에 가까워지게 된다.
- ☐ 표본크기가 N 인 확률표본의 표본평균은 N 이 충분히 크면 근사적으로 정규분포를 따른다

14. 다음 중 이상값 검색을 활용한 응용시스템으로 가장 적절한 것은? *

1점

- ☐ 장바구니 분석시스템
- ☐ 교차 판매 시스템
- ☐ 부정사용방지 시스템
- ☐ 추천 시스템

15. 앙상블 모형에 대한 설명 중 적절하지 않은 것은? *

1점

- ☐ 부스팅(Boosting)은 여러 모델이 순차적 학습을 하며, 붓스트랩 표본을 구성하는 재표본 과정에서 각 자료에 동일한 확률을 부여한다.
- ☐ 랜덤 포레스트는 배깅(Bagging)에 랜덤 과정을 추가한 방법으로 노드 내 데이터를 자식 노드로 나누는 기준을 정할 때 설명변수의 일부분만을 고려함으로 성능을 높이는 방법을 사용한다.
- ☐ 배깅(Bagging)은 원 데이터 집합으로 부터 중복을 허용하는 크기가 같은 표본을 여러 번 단순임의 복원 추출하여 각 표본에 대한 분류기를 생성 후 그 결과를 앙상블 하는 방법이다.
- ☐ 배깅은 반복추출 방법을 사용하기 때문에 같은 데이터가 한 표본에 여러 번 추출될 수도 있고, 어떤 데이터는 추출되지 않을 수도 있다.



16. 다음 중 최근 시계열 데이터에 많은 가중치를 부여하는 분석 방법은?

1점

- ☐ 자기회귀
- ☐ 이동평균
- ☐ 지수평활법
- ☐ 자기회귀이동평균

17. 당뇨 환자 25명의 약물 섭취 전, 후 당 수치 변화 평균을 조사하여 치료 효과 등을 분석하는 방법은? * 1점

- ☐ 독립표본 t 검정
- ☐ 분산분석
- ☐ ANOVA
- ☐ 대응표본 t 검정

18. 다음 의사결정나무의 분리기준에 대한 설명 중 적절하지 않은 것은? *

1점

- ☐ 카이제곱 통계량 p값이 가장 작은 예측변수 선택
- ☐ 엔트로피 지수가 가장 작은 예측변수 선택
- ☐ 지니 지수를 크게 하는 예측변수 선택
- ☐ 분산의 감소량을 최대화하는 기준 선택



19. 다음 중 주성분분석(PCA)에 결과분석에 설명 중 올바르지 않은 것은? *

1점

```
> temp <- prcomp(iris[, -5], scale=TRUE)
> summary(temp)
```

Importance of components:

	PC1	PC2	PC3	PC4
Standard deviation	1.7084	0.9560	0.38309	0.14393
Proportion of Variance	0.7296	0.2285	0.03669	0.00518
Cumulative Proportion	0.7296	0.9581	0.99482	1.00000

- ☐ 두 번째 주성분 변수는 전체 데이터의 22.85%를 설명할 수 있다
- ☐ 전체데이터의 100%를 설명하기 위해 4개의 주성분이 필요하다
- ☐ 2개의 주성분을 이용해서 전체 데이터의 99% 설명이 가능하다.
- ☐ 1개의 주성분만 사용한다면 잃게 되는 정보량이 27.04%이다

20. 신경망에서 결괏값(출력)을 내보낼 때 사용하는 함수로, 가중치 값을 학습할 때 에러가 적게 나도록 돕는 기능을 하는 것은 무엇인가? *

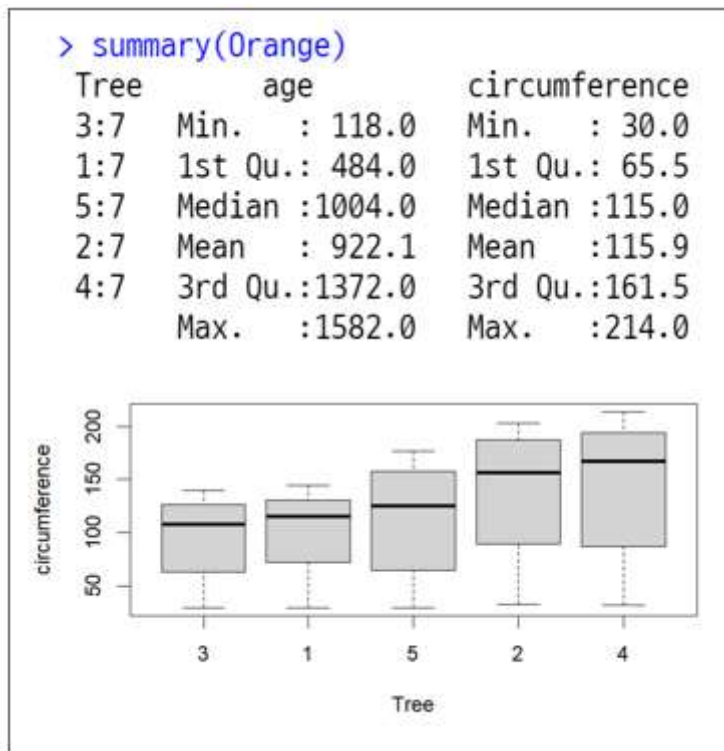
1점

- ☐ 활성화 함수
- ☐ 로짓함수
- ☐ 비용함수
- ☐ 임계함수



21. 오렌지 나무의 dataset에 대한 설명 중 적절하지 않은 것은? *

1점



- ☐ 결측치가 존재하지 않는다
- ☐ age의 최소값은 118.0이다
- ☐ Tree는 연속형 변수이다.
- ☐ 3번 Tree의 중앙값이 가장 낮다

22. 다음 중 회귀분석에 대한 설명으로 부적절한 것은? *

1점

- ☐ 표본회귀선의 유의성 검정은 회귀선의 기울기의 계수가 $\beta \neq 0$ 은 귀무가설, $\beta = 0$ 은 대립가설로 설정한다.
- ☐ 일반선형회귀는 종속변수가 연속형 변수일 때 가능하다
- ☐ 회귀분석의 모형 검정은 F-Test, T-Test이다
- ☐ 로지스틱 회귀분석의 모형 탐색 방법은 최대우도법이다.



23. 아래의 confusion matrix에서 오분류율(Error Rate)을 구하면? *

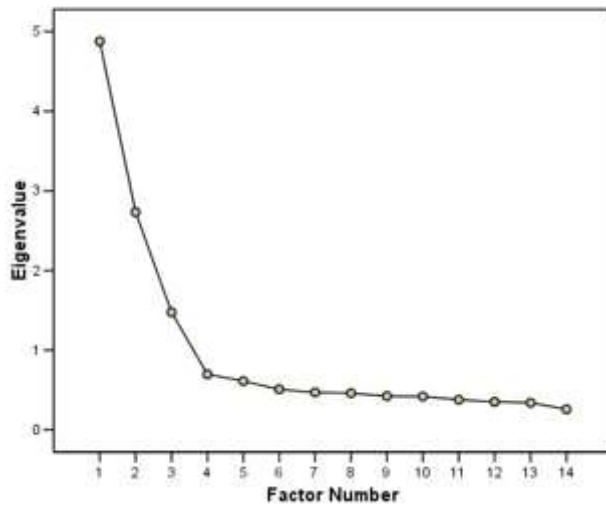
1점

		예측치	
		TRUE	FALSE
실제값	TRUE	40	10
	FALSE	20	30

- ☐ 70%
- ☐ 30%
- ☐ 20%
- ☐ 25%



24. 다음 아래 그림은 주성분 분석의 Scree plot이다. 이 그림을 통해 가정할 수 있는 주성분 변수의 개수는? * 1점



- ☐ 1
- ☐ 4
- ☐ 8
- ☐ 15



25. 아래 표는 주성분 분석의 결과이다. 제 1주성분(PC1) 함수식을 작성하시오. 1점

```
> fit <-prcomp(USArrests, scale=TRUE)
> summary(fit)
Importance of components:

                PC1      PC2      PC3      PC4
Standard deviation  1.5749  0.9949  0.59713  0.41645
Proportion of Variance 0.6201 0.2474 0.08914 0.04336
Cumulative Proportion 0.6201 0.8675 0.95664 1.00000
> round(fit$rotation, 3)
      PC1      PC2      PC3      PC4
Murder  -0.536  0.418 -0.341  0.649
Assault  -0.583  0.188 -0.268 -0.743
UrbanPop -0.278 -0.873 -0.378  0.134
Rape     -0.543 -0.167  0.818  0.089
```

내 답변

26. 인공신경망의 활성화 함수에서 목표치가 다범주인 경우 각 범주에 속할 사 * 1점
후 확률을 제공하는 함수는 무엇인가?

내 답변

27. 회귀모형의 계수를 추정하는 방법으로 잔차제곱합(SSR)을 최소화 하는 계 * 1점
수를 찾는 방법을 무엇이라고 하는가?

내 답변



28. 원 자료로부터 붓스트랩 샘플을 추출하고, 각 붓스트랩 샘플에 대해 트리를 * 1점
형성해 나가는 과정은 배깅과 유사하나, 노드 내 데이터를 자식 노드로 나누는
기준을 정할 때 모든 예측변수에서 최적의 분할을 선택하는 대신, 설명변수의
일부분만을 고려함으로 성능을 높이는 방법을 무엇이라 하는가?

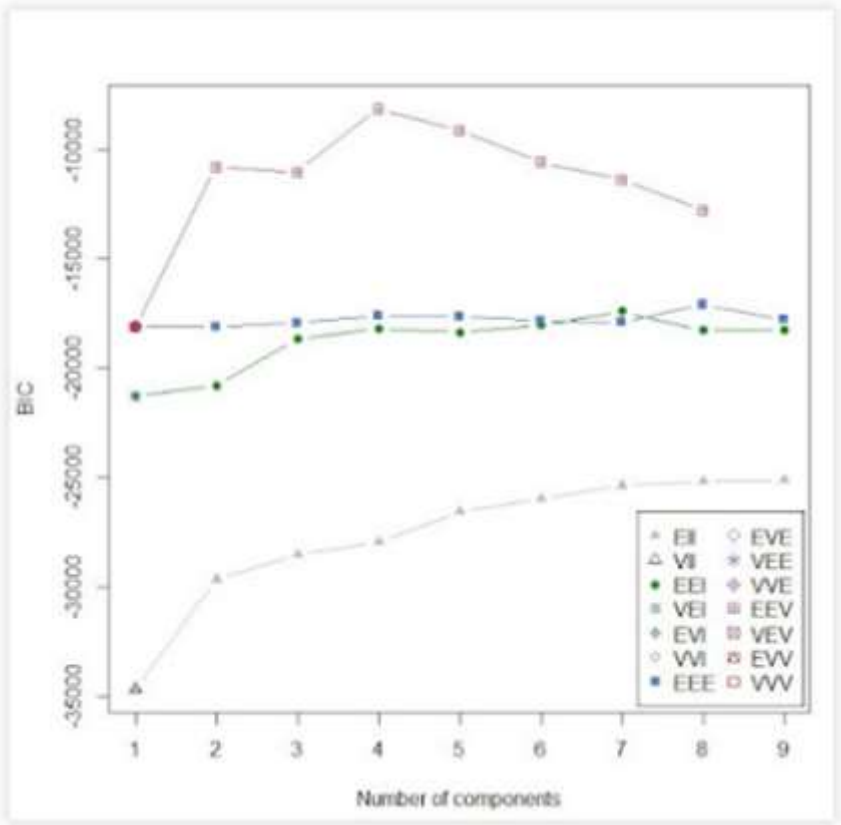
내 답변

29. 로지스틱 회귀분석에서 어떤 일이 일어날 확률을 일어나지 않을 확률로 나 * 1점
누어 log를 취하여 값의 범위를 전체 실수 범위로 확장하는 변환 방법을 무엇이
라고 하는가?

내 답변



30. 다음은 모형 기반 군집분석의 결과이다. 아래 그림을 통한 최적의 군집수는? * 1점



내 답변

제출

양식 지우기

Google Forms를 통해 비밀번호를 제출하지 마세요.

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

Google 설문지

