

## 1과목 빅데이터 분석 기획

1. 다음 중 획득된 다양한 정보를 구조화하여 유의미한 정보로 분류하고 일반화시킨 결과물을 무엇이라고 하는가?

- ① 데이터                      ② 정보
- ③ 지식                        ④ 지혜

[정답] ③

[참고] 수제비 1-2페이지

2. 다음 중 빅데이터의 특징으로 적합하지 않은 것은?

- ① 규모                        ② 다양성
- ③ 가치                        ④ 효율성

[정답] ④

[참고] 수제비 1-3페이지

3. 다음 중 분석조직 인력들을 현업 부서로 직접 배치해 분석 업무를 수행하는 조직 구조로 가장 적합한 것은 무엇인가?

- ① 분산 구조                  ② 기능 구조
- ③ 집중 구조                  ④ 계층 구조

[정답] ①

[참고] 수제비 1-10페이지

4. 다음 중 반정형 데이터 유형으로 가장 적합하지 않은 것은 무엇인가?

- ① 텍스트 문서              ② XML
- ③ JSON                      ④ 웹로그

[정답] ①

[참고] 수제비 1-72페이지

5. 다음 중 개인정보처리자가 정당한 처리 범위 내에서 정보주체의 동의 없이 가명정보를 처리할 수 있는 분야로 옳바르지 않은 것은?

- ① 특정 집단이나 대상 등에 관하여 작성한 수량적인 정보
- ② 기술의 개발과 실증, 기초연구, 응용 연구
- ③ 마케팅 등을 위해 특정 개인을 식별할 수 있는 통계
- ④ 공공의 이익을 위하여 지속적으로 열람할 가치가 있는 기록정보 보존

[정답] ③

[참고] Daily (1과목) 빅데이터 분석 기획 - 3

6. 개인정보의 일부를 삭제하거나 일부 또는 전부를 대체하는 등의 방법으로 추가정보 없이는 특정 개인을 알아볼 수 없도록 처리하는 것은 무엇인가?

- ① 익명처리                      ② 가명처리
- ③ 총계처리                      ④ 범주화

[정답] ②

[참고] 수제비 1-25페이지

7. 다음에서 설명하는 비식별 조치 방법은 무엇인가?

정약용, 35세 → 정씨, 30대

- ① 익명처리                      ② 총계처리
- ③ 데이터 범주화              ④ 데이터 마스킹

[정답] ③

[참고] 수제비 1-25페이지

8. 다음이 설명하는 것은 무엇인가?

자신에 관한 정보가 언제 누구에게 어느 범위까지 알려지고 또 이용되도록 할 것인지를 그 정보주체가 스스로 결정할 수 있는 권리

- ① 개인정보 이동권
- ② 개인정보 열람권
- ③ 개인정보 삭제권
- ④ 개인정보 자기 결정권

[정답] ④

[참고] 수제비 1-22페이지

9. 다음 분석 기획 유형 중 분석의 대상은 인지하고 있으나 방법을 모르는 경우에 적용하는 유형은 무엇인가?

- ① 최적화                      ② 솔루션
- ③ 통찰                        ④ 발견

[정답] ②

[참고] 수제비 1-41페이지

10. 다음 중 CRIPS-DM 분석 방법론의 절차는 무엇인가?

- ① 데이터 세트 선택→데이터 전처리→데이터 변환→데이터 마이닝→데이터 마이닝 결과 평가
- ② 업무 이해→데이터 이해→데이터 준비→모델링→평가→전개
- ③ 샘플링→탐색→수정→모델링→검증
- ④ 문제 인식→연구 조사→모델링→평가

[정답] ②

[참고] 수제비 1-47페이지

## 2과목 빅데이터 탐색

11. 아래와 같은 데이터가 관측이 되었을 때, 비조건부 평균 대치법으로 결측값을 대치할 때, 대치값으로 가장 알맞은 것은? (단, ‘?’는 결측값이다.)

10, ?, 15, 19, 12, 18, ?, ?, 16

- ① 14                          ② 15
- ③ 16                        ④ 17

[정답] ②

[해설]

- 수제비 빅데이터 분석기사 2-8

12. 통계기법을 이용한 데이터 이상값 검출 방법으로 가장 알맞지 않은 것은?

- ① ESD(Extreme Studentized Deviation)
- ② 디슨의 Q검정
- ③ 그룹스 T-검정
- ④ 확률밀도함수

[정답] ④

[해설]

- 수제비 빅데이터 분석기사 2-13

13. 다음 중 필터기법의 사례로 가장 알맞지 않은 것은?

- ① 정보소득(Information Gain)
- ② 피셔스코어(Fisher Score)
- ③ RFE(Recursive Feature Elimination)
- ④ 상관계수(Correlation Coefficient)

[정답] ③

[해설]

- 수제비 빅데이터 분석기사 2-23

14. 다변량의 신호를 통계적으로 독립적인 하부 성분으로 분리하여 차원을 축소하는 기법으로 가장 알맞은 것은?

- ① 독립성분분석              ② 요인분석
- ③ 다차원 척도법            ④ 주성분 분석

[정답] ①

[해설]

- 수제비 빅데이터 분석기사 2-26

15. 파생변수 생성방법의 예 중 변수결합에 해당하는 것은?

- ① 12시간을 24시간으로 변환
- ② 매출액과 방문 횟수로 평균 매출액 추출
- ③ 고객별 방문 횟수 집계
- ④ 남/여를 M/F 로 변환

[정답] ②  
[해설]  
- 수제비 빅데이터 분석기사 2-27

16. 언더 샘플링 기법 중 다수 클래스의 데이터를 토맥 링크 방법으로 제거한 후 Condensed Nearest Neighbor를 이용하여 밀집된 데이터를 제거하는 기법은?

- ① ENN                      ② SMOTE  
③ OSS                      ④ ADASYN

[정답] ③  
[해설]  
- 수제비 빅데이터 분석기사 2-30

17. EDA(Exploratory Data Analysis)의 4가지 주제로 가장 알맞지 않은 것은?

- ① Resistance  
② Residual  
③ Graphic Representation  
④ Accuracy

[정답] ④  
[해설]  
- 수제비 빅데이터 분석기사 2-43

18. 다음 중 변수의 속성에 따른 상관관계 분석 방법으로 가장 알맞은 것은?

- ① 수치적 데이터 - 카이제곱 검정  
② 순서적 데이터 - 스피어만 순위상관분석  
③ 명목적 데이터 - 스피어만 순위상관분석  
④ 명목적 데이터 - 피어슨 상관계수

[정답] ②  
[해설]  
- 수제비 빅데이터 분석기사 2-48

19. 다음 중 중위수로 가장 알맞은 것은?

3, 5, 7, 11

- ① 4                      ② 5  
③ 6                      ④ 7

[정답] ③  
[해설]  
- 수제비 빅데이터 분석기사 2-49

20. 박스 플롯의 구성요소 중 자료들의 하위 75%의 위치를 의미하는 것은?

- ① 제1사분위              ② 제3사분위  
③ 최솟값                  ④ 최댓값

[정답] ②  
[해설]  
- 수제비 빅데이터 분석기사 2-54

### 3과목 빅데이터 모델링

21. 다음 중 회귀 분석 시 회귀 모형의 가정사항으로 가장 옳지 않은 것은

- ① 선형성                      ② 상관성  
③ 등분산성                  ④ 독립성

[정답] ②  
[참고] 수제비 3-28페이지

22. 다음 중 회귀 모형의 검증 체크리스트에 대한 설명 중 가장 옳지 않은 것은?

- ① 추정된 회귀식은 유의수준 5% 하에서 F-통계량의 p-값이 0.05보다 클 경우 통계적으로 유의하다고 볼 수 있다.  
② 회귀계수의 유의성은 해당 계수의 T-통계량을 이용한다.  
③ 모형의 설명력은 결정계수를 확인한다.  
④ 데이터는 선형성, 독립성, 등분산성, 비상관성, 정상성 가정을 만족시켜야 한다.

[정답] ①  
[참고] 수제비 3-28페이지

23. 게임에서 이길 확률이 3/10, 질 확률이 7/10일 경우 게임에서 이길 승산(Odds)은 얼마인가?

- ① 3/7                      ② 3/10  
③ 7/10                    ④ 7/3

[정답] ①

[참고] 수제비 3-35페이지

24. 의사결정나무에서 과대적합을 방지하기 위하여 더 이상 가지가 분리가 일어나지 않고 현재의 마디가 끝마디가 되도록 하는 규칙은 무엇인가?

- ① 가지치기                ② 정지규칙  
③ 분류규칙                ④ 분리기준

[정답] ②

[참고] 수제비 3-40페이지

25. 다음 중 의사결정나무에서 분류규칙을 선택하기 위하여 사용하는 척도가 아닌 것은 무엇인가?

- ① 카이제곱 통계량  
② 지니 지수  
③ 엔트로피 지수  
④ 분산

[정답] ④

[참고] 수제비 3-40페이지

26. 다음 그림의 지니 지수는 얼마인가?



- ① 2/5                      ② 3/5  
③ 13/25                ④ 12/25

[정답] ④

[참고] 수제비 3-41페이지

27. 다음 중 의사결정나무의 알고리즘인 CART에 대한 설명으로 가장 옳지 않은 것은?

- ① 각 독립변수를 이분화하는 과정을 반복하여 이진트리 형태를 형성함으로써 분류를 수행하는 방법이다.

② 변수가 범주형일 경우 불순도의 척도로 엔트로피 지수를 이용한다.

③ 가장 널리 이용되는 의사결정나무 알고리즘이다.

④ 개별 입력변수 뿐만 아니라 입력 변수들의 선형 결합 중에서 최적의 분리를 구할 수 있다.

[정답] ②

[참고] 수제비 3-43페이지

28. 다음 중 의사결정나무에 대한 설명으로 가장 옳지 않은 것은?

① 나무 구조에 의해서 모형이 표현되기 때문에 모형을 사용자가 쉽게 이해 가능하다.

② 두 개 이상의 변수가 결합하여 목표변수에 어떻게 영향을 주는지 쉽게 파악이 가능하다.

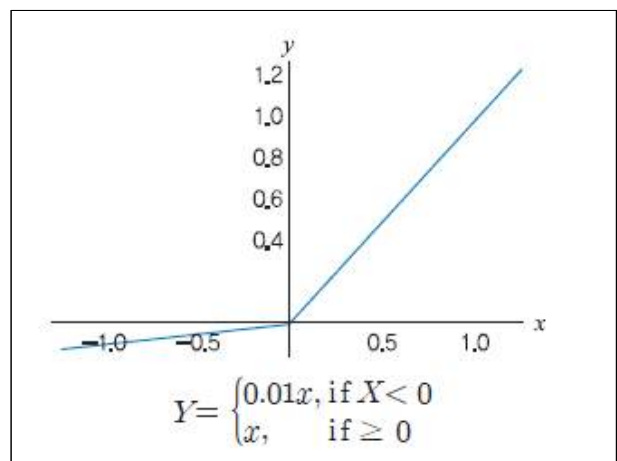
③ 선형성이나 정규성 또는 등분산성 등의 가정이 필요한 모수적 모형이다.

④ 연속형 변수를 비연속적인 값으로 취급하기 때문에 분리의 경계점 근방에서 예측 오류가 클 가능성이 있다.

[정답] ③

[참고] 수제비 3-46페이지

29. 다음의 그림이 나타내는 뉴런의 활성화 함수는 무엇인가?



① 계단 함수

② 시그모이드 함수

③ tanh 함수

④ Leaky ReLU

[정답] ④  
[참고] 수제비 3-51페이지

30. 서포트 벡터 머신의 구성 요소 중에서 완벽한 분리가 불가능할 때 선형적으로 분류를 위해 허용된 오차를 위한 변수는 무엇인가?

- ① 결정 경계
- ② 슬랙 변수
- ③ 마진
- ④ 커널 트릭

[정답] ②  
[참고] 수제비 3-53페이지

#### 4과목 빅데이터 결과 해석

31. 다음 중 알고리즘의 정확도를 판단하기 위한 AUC(Area Under ROC)의 수치 중 가장 좋은 모형은 무엇인가?

- ① 0.9                      ② 0.7
- ③ 0.1                      ④ 0.5

[정답] ①  
[해설]  
- 수제비 빅데이터 분석기사 4-9

32. 다음 중 부트스트랩(Bootstrap)에 대한 설명으로 옳지 않은 것은?

- ① 무작위 복원추출 기법으로, 전체 데이터에서 중복을 허용하여 데이터 크기만큼 샘플을 추출하고 이를 학습 데이터로 한다.
- ② 주어진 자료에서 단순 랜덤 복원추출 방법으로 동일한 크기의 표본을 여러 개 생성하는 샘플링 방법이다.
- ③ 부트스트랩을 통해 100개의 샘플을 추출하더라도 샘플에 한 번도 선택되지 않는 원 데이터가 발생할 수 있는데 전체 샘플의 약 63.2%가 이에 해당한다.

④ 한 번도 포함되지 않는 샘플들은 평가(Test)에 사용한다.

[정답] ③  
[해설]  
- 수제비 빅데이터 분석기사 4-16

33. 다음 중 취합(Aggregation) 방법론으로 가장 거리가 먼 것은?

- ① 다수결(Majority Voting)
- ② 배깅(Bagging)
- ③ 부스팅(Boosting)
- ④ 페이스팅(Pasting)

[정답] ③  
[해설]  
- 수제비 빅데이터 분석기사 4-35

34. 교차검증 방법 중 모집단으로부터 조사의 대상이 되는 표본을 무작위로 추출 하는 기법은 무엇인가?

- ① Holdout Cross Validation
- ② LOOCV
- ③ k-Fold Cross Validation
- ④ Random Sampling

[정답] ④  
[해설]  
- 수제비 빅데이터 분석기사 4-z13

35. 다음 혼동 행렬(Confusion Matrix)에서 F1 스코어는 얼마인가?

		예측값		합계
		True	False	
실제값	True	30	70	100
	False	20	80	100
합계		50	150	200

- ①  $\frac{2}{5}$                       ②  $\frac{3}{5}$
- ③  $\frac{1}{5}$                       ④  $\frac{4}{5}$

[정답] ①

[해설]

- 수제비 빅데이터 분석기사 4-8

36. 다음 중 경사 하강법의 단점을 개선해 주는 기법으로 옳지 않은 것은?

- ① Momentum                      ② AdaBoost  
③ AdaGrad                        ④ Adam

[정답] ②

[해설]

- 수제비 빅데이터 분석기사 4-31

37. 다음 중 데이터 시각화의 절차로 올바르게 연결된 것은?

- ① 구조화 단계→시각화 단계→시각표현 단계  
② 시각화 단계→구조화 단계→시각표현 단계  
③ 시각표현 단계→시각화 단계→구조화 단계  
④ 시각표현 단계→구조화 단계→시각화 단계

[정답] ①

[해설]

- 수제비 빅데이터 분석기사 4-49

38. 다음이 설명하는 데이터 시각화 기능은 무엇인가?

- 데이터에 숨겨져 있는 관계와 패턴을 찾기 위한 시각적 분석 기능  
- 데이터의 의미미하고 흥미로운 요소를 이용자가 직접 탐색함

- ① 설명 기능  
② 탐색 기능  
③ 보안 기능  
④ 표현 기능

[정답] ②

[해설]

- 수제비 빅데이터 분석기사 4-47

39. 다음 중 빅데이터 시각화 도구로 가장 옳지 않은 것은 무엇인가?

- ① 태블로(Tableau)  
② 인포그램(Infogram)  
③ 데이터래퍼(Data Wrapper)  
④ 하이브(Hive)

[정답] ④

[해설]

- 수제비 빅데이터 분석기사 4-48

40. 다음 중 드롭아웃의 유형으로 옳지 않은 것은 무엇인가?

- ① 시간적 드롭아웃  
② 간헐적 드롭아웃  
③ 초기 드롭아웃  
④ 공간적 드롭아웃

[정답] ②

[해설]

- 수제비 빅데이터 분석기사 4-29