

빅데이터의 이해, 시뮬레이션

2018학년도 2 학기

4 학년 2 교시

※ 정답 하나만을 골라 반드시 컴퓨터용 사인펜으로 OMR 답안지에 표기할 것.	학 과		감독관	㉠
	학 번	-	성 명	

1과목 빅 데이터 의 이 해 (1~25)

출제위원 : 방송대 이공희

출제범위 : 교재 1~10장 (R 프로그램 실습 제외)

1. 빅데이터의 출현배경에 해당하는 것을 모두 고르시오. (4점)

- ㉠ 스마트폰과 SNS의 보급 확대
㉡ 사물인터넷과 센서의 보급 확대
㉢ 데이터 처리를 위한 하드웨어의 가격상승
㉣ 새로운 데이터 유형을 다루는 신기술 등장

- ① 가, 나, 다
② 나, 다, 라
③ 가, 나, 라
④ 가, 다, 라

2. 다음 데이터 크기 단위 중 가장 크기가 큰 단위는? (2점)

- ① TB ② YB
③ ZB ④ PB

3. 다음 중 비정형데이터를 모두 고르시오. (3점)

- ㉠ 유튜브 동영상
㉡ 인구주택 총조사 결과
㉢ 블로그 댓글
㉣ 페이스북 사진

- ① 가, 나, 다
② 나, 다, 라
③ 가, 나, 라
④ 가, 다, 라

4. “데이터 과학은 (), ()과 관련 주제 전문지식이 종합된 분야이다.”에서 ()에 가장 적합한 단어들은? (4점)

- ① 수리능력, 경영기술
- ② 통계학, 해킹기술
- ③ 경영기술, 해킹기술
- ④ 통계학, 경영기술

5. 다음 중 데이터과학자가 가져야할 소프트스킬로 가장 적합한 것은?
(4점)

- ① R프로그램 작성 능력
- ② 데이터에 있는 가치를 발견할 수 있는 통찰력
- ③ 통계학
- ④ 텍스트마이닝 방법

6. 다음 중 빅데이터 경영이라고 할 수 있는 것은? (3점)

- ① 인터넷으로 은행 잔액조회 및 이체
- ② 인터넷으로 주민등록등본을 발급
- ③ 인터넷으로 도서를 단순 판매
- ④ 추천시스템을 통해 맞춤 상품을 추천

7. 필립 코틀러 등이 말하는 디지털 시대의 마케팅 트렌드가 아닌 것은? (3점)

- ① 고객의 취향과 요구를 감지, 대응하는데 주력하는 마케팅
- ② 고객의 마음을 얻는데 주력하는 마케팅
- ③ 새로운 고객을 확보하는데 주력하는 마케팅
- ④ 고객을 개별적으로 대하는 맞춤형 마케팅

8. 분산 시스템으로 높은 확장성을 갖는 컴퓨팅 자원을 인터넷으로 제공하는 서비스로 사용자가 컴퓨팅 자원의 일부를 임대하여 사용한 만큼 사용료를 지불하는 것을 무엇이라 부르는가? (4점)

- ① 클라우드 컴퓨팅
- ② 서버 컴퓨팅
- ③ 빅데이터 컴퓨팅
- ④ 아웃 컴퓨팅

9. 하둡 분산파일시스템에서 데이터 노드의 동작 상태를 실시간으로 관리하고 데이터 노드에 분산 저장되어 있는 메타데이터를 관리하는 일을 하는 것은? (4점)

- ① 슬레이브 노드
- ② 리듀스 노드
- ③ 리스크 노드
- ④ 마스터 노드

10. MIT Billion Price Project를 기반으로 작성된 온라인 물가지수에서 상품가격을 파악하는 방법으로 가장 바르게 기술된 것은? (2점)

- ① 웹 스크래핑(web scraping)
- ② 설문조사
- ③ 웹 조사
- ④ KOSIS 검색

11. 1800년~2000년 사이 과학자들의 활동시점과 영향력을 과거에 축적된 책에서의 출현빈도를 기반으로 파악하려고 할 때 이를 가능하게 하는 서비스는? (2점)

- ① 구글 트렌드
- ② Twitter API
- ③ 구글 Ngram
- ④ Facebook API

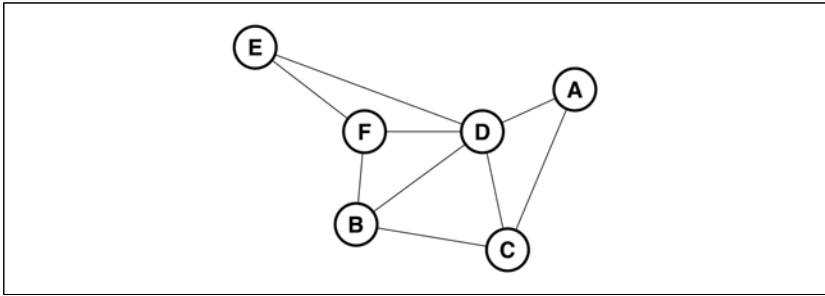
12. 텍스트 프로세싱에서 토큰화(tokenizing)는 무엇인가? (3점)

- ① 단어의 기본형을 추출하는 과정
- ② 텍스트 데이터를 작은 단위로 잘게 쪼개는 것
- ③ 하나의 어절에서 형태소를 분석하는 것
- ④ 텍스트 데이터를 시각화하는 것

13. 문서에 등장한 단어의 빈도수를 분석하여 어떤 단어가 많이 사용되었는지를 보여주는 시각화 방법은? (2점)

- ① 워드 트리
- ② 태그 링
- ③ 태그 클라우드
- ④ 워드 프로세싱

※ (14~15) 다음 소셜네트워크 그래프에 대해 답하시오.



14. 소셜네트워크의 엣지의 수는? (2점)
- ① 3

② 5

③ 7

④ 9
15. 다음 노드 중 디그리(degree)가 가장 큰 노드는? (2점)
- ① B

② C

③ D

④ F
16. 선형회귀 모형에서 설명변수의 수가 데이터의 수보다 큰 경우 이용되는 변수선택방법의 문제점에 대한 설명 중 바르게 기술된 것은? (3점)
- ① 추정결과가 불안정하고 회귀모형의 예측력이 저하된다.

② 추정결과는 불안정적이지만 회귀모형의 예측력이 개선된다.

③ 추정결과는 안정적이지만 회귀모형의 예측력이 저하된다.

④ 추정결과가 안정적이고 회귀모형의 예측력이 개선된다.
17. 벌점화 최소제곱추정량에 대한 설명 중 가장 바르게 기술된 것은? (2점)
- ① 편의가 없어지지만 분산이 커질 수 있다.

② 편의가 생기고 분산도 커질 수 있다.

③ 편의가 없이 분산을 크게 줄일 수 있다.

④ 편의가 생기지만 분산을 크게 줄일 수 있다.
18. 배깅방법에서 부트스트랩(bootstrap) 데이터에 대한 설명 중 가장 바른 것은? (3점)
- ① 주어진 데이터로부터 동일한 크기의 표본을 랜덤 복원 추출로 뽑은 데이터

② 주어진 데이터로부터 동일한 크기의 표본을 랜덤 비복원 추출로 뽑은 데이터

③ 주어진 데이터를 동일하게 나눈 데이터

④ 주어진 데이터를 크기 순서대로 정리한 데이터
19. 의사결정나무의 특징인 분산이 크다는 점을 고려한 양상을 기법으로 배깅보다 더 많은 무작위성을 주어 예측모형들을 생성한 후 이를 선형결합하여 최종 학습기를 만드는 방법은? (3점)
- ① 벌점화 방법

② 로지스틱 회귀분석

③ 랜덤 포레스트

④ 주성분 분석
20. k -평균 군집분석에 대한 설명 중 가장 바르게 기술된 것은? (2점)
- ① 계층적 군집분석에 비하여 계산량이 적은 지도학습법이다.

② 계층적 군집분석에 비하여 계산량이 많은 지도학습법이다.

③ 계층적 군집분석에 비하여 계산량이 적은 비지도학습법이다.

④ 계층적 군집분석에 비하여 계산량이 많은 비지도학습법이다.

※ (21~22) 편의점의 거래내역을 보고 물음에 답하시오.

고객번호	품목
1	빵, 라면, 사과
2	커피, 빵, 식기세척제, 사과
3	식기세척제, 세제
4	커피, 라면
5	빵, 세제, 커피

21. “빵을 구매하면 커피를 구매한다.”의 지지도는? (2점)
- ① $\frac{1}{5}$

② $\frac{2}{5}$

③ $\frac{3}{5}$

④ $\frac{4}{5}$
22. “빵을 구매하면 커피를 구매한다.”의 신뢰도는? (3점)
- ① $\frac{1}{2}$

② $\frac{2}{3}$

③ $\frac{3}{4}$

④ $\frac{2}{5}$
23. 협력적 정확방법에 대한 설명으로 가장 바른 것은? (3점)
- ① 로맨스영화를 좋아하는 고객 중 다수가 A사의 자동차를 소유하고 있다면, A사의 자동차를 소유한 고객에게 로맨스영화를 추천하는 것

② 로맨스영화를 좋아하는 고객 중 다수가 A사의 자동차를 소유하고 있다면, 로맨스 영화를 좋아하지만 자동차가 없는 고객에게 A사의 자동차를 추천하는 것

③ 로맨스영화를 좋아하는 고객 중 다수가 A사의 자동차를 소유하고 있다면, 로맨스 영화를 좋아하는 모든 고객에게 A사의 자동차를 추천하는 것

④ 로맨스영화를 좋아하는 고객 중 다수가 A사의 자동차를 소유하고 있다면, 자동차를 소유한 모든 고객에게 로맨스영화를 추천하는 것
24. 다음 중 프라이버시에 가장 속하지 않는 것은? (2점)
- ① 정치적 신념

② 질병정보

③ 범죄경력

④ 소속 직장 홈페이지 주소
25. 개인정보에 대한 이용 동의를 받는 방식 중 옵트아웃에 대한 설명 중 가장 바르게 기술된 것은? (3점)
- ① 거부의를 표시하지 않는 한 동의한 것으로 간주하는 것

② 사전에 동의를 받는 것

③ 거부의를 표시해도 동의한 것으로 간주하는 것

④ 문제가 발생될 경우만 사후에 동의를 받는 것