

데이터과학입문, 인간과과학

2019학년도 2 학기

1 학년 2 교시

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

1과목	데 이 터 과 학 입 문 (1~25)
출제위원 : 방송대 장영재	
출제범위 : 교재 1~8장(강의 포함)	

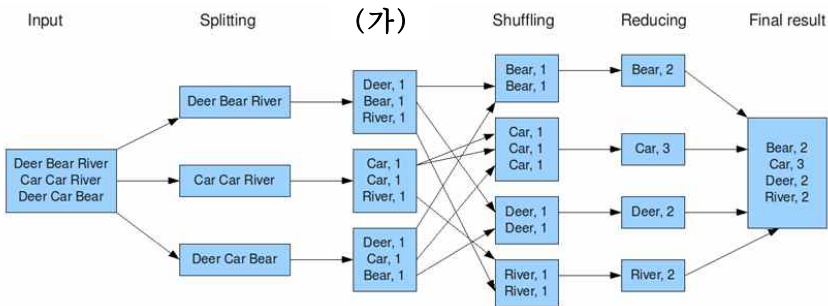
1. 데이터과학의 역사를 데이터 수집의 역사의 관점에서 살펴볼 때, 다음 중 아래 <보기>와 가장 관계 깊은 인물은? (2점)

<보 기>

- ㄱ. 과거 결과에 대한 데이터를 활용하여 미래의 행동을 예측
 ㄴ. 고객의 행동 및 기본 정보를 취득하여 의사결정에 활용
 ㄷ. 데이터 분석 결과로 인간의 편향적인 의사결정을 대체

- ① 로널드 피셔 ② 조지 갤럽
 ③ 윌리엄 페어 ④ 다구치 겐이찌

2. 아래는 하둡의 맵리듀스에 관한 그림이다. 다음 중 단계 (가)를 지칭하는 것으로 가장 옳은 것은? (3점)



- ① Randomization ② Mapping
 ③ Ordering ④ Serving

3. 다음 <보기>의 내용과 가장 관계 깊은 것은? (2점)

<보 기>

기호정보에 바탕을 둔, 사용자 성향, 추천 시스템

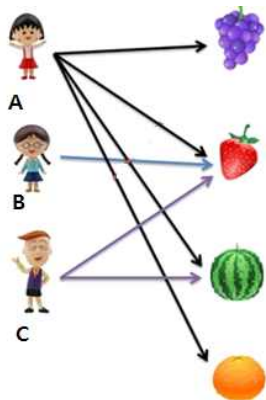
- ① 클라우드 컴퓨팅 ② 협업필터링
 ③ 데이터웨어하우스 ④ 캐글(Kaggle)

4. 다음 설명과 가장 관계 깊은 것은? (2점)

로담스테드 연구소, 곡물관련 실험 결과, 통제된 실험

- ① 실험계획법 ② 고객거래 데이터
 ③ 공정과정 데이터 ④ 기계학습

5. 다음 중 아래 그림에 대해 옳은 설명으로만 짝지어진 것은? (4점)



- ㄱ. 협업필터링의 한 종류이다.
 ㄴ. 사용자들의 선호도만 고려한 모형기반 협업 필터링이다.
 ㄷ. 위 사례의 경우 C에게 포도와 오렌지를 추천할 수 있다.

- ① ㄱ, ㄴ ② ㄴ, ㄷ
 ③ ㄱ, ㄷ ④ ㄱ, ㄴ, ㄷ

6. 다음 중 모수적 모형 접근법과 알고리즘 접근방법을 비교하였을 때, 적절하게 짝지어지지 않은 것은? (3점)

	모수적 모형 접근방법	알고리즘 접근방법
① 특징	모수를 과거 데이터로부터 적합	알고리즘에 의해 정해진 방식에 의해 계산된 결과로 분석
② 장점	결과에 대한 해석이 대체로 용이	데이터 복잡성이 높아도 적용 가능
③ 단점	기계만 인지할 수 있는 결과로 해석이 어려움	가정이나 설정한 식에 부합하지 않는 데이터의 경우, 성능이 낮을 수 있음
④ 모형	선형회귀, 로지스틱회귀	의사결정나무, 신경망, 배깅

7. 기초가 되는 URL seed들을 저장한 뒤 웹페이지의 하이퍼링크를 인식하여 URL을 갱신하며 반복적으로 웹링크를 찾는 과정을 의미하는 것으로 가장 적절한 것은? (2점)

- ① 파싱 ② API
 ③ 웹크롤링 ④ DBMS

8. 다음 중 데이터 모델에 관한 설명으로 옳은 설명을 모두 고르면? (3점)

- ㄱ. 데이터베이스 설계과정에서 데이터의 구조를 표현하기 위해 사용되는 도구
 ㄴ. 현실세계의 데이터를 컴퓨터에서의 데이터로 표현하는 개념적 도구
 ㄷ. 데이터를 구조화하는 방법론
 ㄹ. 구성요소로는 목적(Purpose), 청중(People), 장소(Place) 등 3P를 꼽을 수 있음

- ① ㄱ, ㄴ, ㄷ ② ㄴ, ㄷ, ㄹ
 ③ ㄱ, ㄷ, ㄹ ④ ㄱ, ㄴ, ㄷ, ㄹ

9. 다음 중 데이터를 사용하기에 적절한 환경, 즉 적합성, 적시성, 정확성, 완전성, 적절성, 접근 가능성 등을 이르는 말은? (2점)

- ① 빅데이터 ② 데이터 준설
 ③ 데이터 품질 ④ 클라우드소싱

※ 다음 R 프로그램 등에서 사용되는 데이터 구조에 관한 질문에 답하시오. (10~11)

10. 다음 중 행렬에 대한 설명으로 옳은 것을 모두 고르면? (3점)

- ㄱ. 동일한 형태(mode)로 구성된 2차원의 데이터 구조로 행의 차원과 열의 차원을 갖고 있음
 ㄴ. 벡터의 속성을 포함하고 있음
 ㄷ. 논리형과 수치형 원소를 가질 수 있으나 문자형 원소는 가질 수 없음

- ① ㄱ, ㄴ ② ㄴ, ㄷ
 ③ ㄱ, ㄷ ④ ㄱ, ㄴ, ㄷ

11. 다음 중 리스트에 관한 설명으로 옳은 것은 모두 몇 개인가? (3점)

- ㄱ. 서로 다른 형태(mode)의 데이터로 구성된 객체
 ㄴ. 문자형, 수치형, 논리형 자료가 혼합되어 한 리스트 내에 존재할 수 있음
 ㄷ. 일반화된 행렬이라 볼 수 있음
 ㄹ. 벡터를 확장시킨 객체로서 모든 원소가 같은 길이의 행과 같은 길이의 열의 구조를 가짐

- ① 1개 ② 2개
 ③ 3개 ④ 4개

12. 다음 중 신용카드 사기, 보험사기, 인터넷 뱅킹 사기, 품질보증 관련 사기, 자금세탁, 정보유출이나 사이버해킹 같은 유형의 사기나 부정행위에 대응하여 데이터 과학자가 사용할 수 있는 분석방법을 모두 고르면? (3점)

- ㄱ. 위험도 예측 모델링
- ㄴ. 관계망분석
- ㄷ. 고도의 데이터 시각화

- ① \neg , \perp
② \perp , \top
③ \neg , \top
④ \neg , \perp , \top

13. 다음 중 아래의 설명과 가장 관계가 깊은 것은? (2점)

저장된 데이터, 접근 및 공유 가능한 공유데이터, 업무운영을 위한 운영데이터

- ① 데이터베이스 ② 내부 데이터
③ 검색 데이터 ④ 웹크롤링

14. 다음 중 데이터베이스 관리 시스템의 관리 기능과 그에 관한 설명으로 옳게 짝지어진 것은? (3점)

ㄱ. 정의	a. 특정한 데이터를 검색하기 위한 질의, 데이터베이스의 갱신, 삽입, 삭제 등을 체계적으로 관리하기 위한 인터페이스 제공
ㄴ. 조작	b. 모든 응용 프로그램이 요구하는 데이터 구조를 지원하기 위해 데이터베이스에 저장될 데이터에 대한 형식, 구조, 제약조건들을 명시하는 기능
ㄷ. 제어	c. 데이터베이스를 접근하는 갱신, 삽입, 삭제 등의 작업이 정확하게 수행되어 무결성을 유지하도록 하고 하드웨어나 소프트웨어의 오동작으로부터 시스템을 보호

- $$\begin{array}{ll} \textcircled{1} \neg - c & \textcircled{2} \neg - a \\ \textcircled{3} \sqsubset - b & \textcircled{4} \sqsubset - a \end{array}$$

15. 다음 중 데이터 분석 도입의 성공요소 중 질문의 우선적 정의 방식(question first)을 옳은 순서대로 나열한 것은? (4점)

(A)	(B)	(C)	(D)
필요한 데이터 및 정보의 정의	질 문	필요한 분석의 정의	경영정책과 연계성이 있는 데이터의 입수

- ① C-A-B-D ② A-C-B-D
③ B-C-A-D ④ D-C-A-B

16. 다음 중 아래 <보기>의 세 가지 방법들과 가장 관계가 깊은 것은? (3점)

-<보 기>

상향식접근법, 하향식접근법, 벤치마킹 풀 작성법

- ① 데이터 품질관리
- ② 분석 핵심 문제(기회)의 발굴
- ③ 기업가치의 평가
- ④ 질문의 우선적 정의 방식

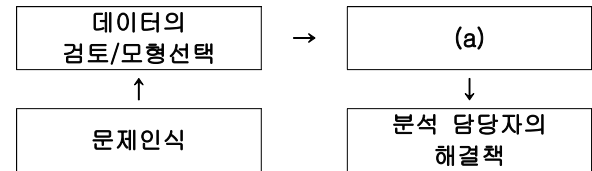
17. 다음 중 분석질문을 구체화하는 방법으로서 아래 <보기>의 항목들과 가장 관계가 깊은 것은? (3점)

- <보 기>

최초 분석기회에 관한 질문, 필요한 사항과 연계된 질문,
최종 분석 목적을 달성하는 데 필요한 사항을 찾기 위한 질문

- ① 분석기획
- ② 연속 질문 방식
- ③ 프리젠테이션
- ④ 업무요소별 질문 방식

18. 아래 통계적 문제해결 과정의 (a) 단계에 대한 설명으로 옳은 것을 모두 고르면? (4점)



- ㉠. (a)의 과정은 데이터를 바탕으로 모델을 구축하고 해결책을 찾는 것이다.
- ㉡. (a)의 과정에서는 전문적인 식견에서 면밀한 검토와 검증을 필요로 한다.
- ㉢. (a)의 과정은 분석목적의 일관성을 유지하며 의사결정 참여자들에게 이해하기 쉬운 언어로 전달하는 과정을 포함한다.

- ① \neg , \perp
② \perp , \sqsubset
③ \neg , \sqsubset
④ \neg , \perp , \sqsubset

19. 다음 중 데이터 시각화의 단계 중 데이터를 수집하고 정제하는 과정으로서 데이터 세트를 만들기 위한 분석도구가 필요한 단계로 가장 적절한 것은? (3점)

- ① 정보구조화 ② 정보시각화
③ 상호작용 ④ 정보시각표현

20. 다음 중 비교 시각화에 해당하는 것을 모두 고르면? (3점)

㉠. 히트맵 ㉡. 체르노프페이스 ㉢. 스타차트 ㉣. 히스토그램

- ① \neg, \perp, \exists

③ \neg, \perp, \top

② \perp, \top, \exists

④ $\neg, \perp, \top, \exists$

21. 다음 중 데이터마이닝 기법의 특징에 관한 설명으로 가장 적절하지 않은 것은? (2점)

- ① 경험적 방법
- ② 대용량의 관측 가능한 자료
- ③ 수리적 특성이 반드시 규명됨
- ④ 컴퓨터 집약적 기법

22. 다음 데이터마이닝 기법을 감독학습과 자율학습으로 나눌 때 이에 관한 설명 중 가장 옳지 않은 것은? (3점)

- ① 감독학습의 목표는 입출력 간의 관계를 결정하는 시스템에 대한 유용한 근사 시스템을 구하는 것으로 정의할 수 있다.
- ② 감독학습에서는 학습에 사용되는 실제 출력변수가 존재한다.
- ③ 감독학습의 사례로서 장바구니 분석을 꼽을 수 있으며 입출력 간의 관계를 결정하는 시스템을 찾기보다는 데이터에 존재하는 여러 가지 형태의 특징을 추출하는 데 목적을 둔다.
- ④ 자율학습은 생명정보학의 마이크로어레이 데이터 분석에서와 같이 데이터 중에서 유사한 특성을 가지는 그룹의 존재를 찾는 문제와 관련이 있다.

23. 다음 중 데이터 기반의 맞춤형 개인화 서비스와 관련이 깊은 것으로만 묶인 것은? (3점)

ㄱ. 스마트 빅 보드 ㄴ. 구글의 스마트카
ㄷ. 아마존의 예측배송 ㄹ. 프레드폴(Predpol)

- ① \neg, \perp

③ \neg, \perp, \exists

② \perp, \exists

④ \neg, \exists, \exists

24. 다음 중 ‘소프트웨어공학자보다 통계학에 대해 더 잘 알면서, 통계학자보다 소프트웨어공학을 잘 아는 사람’이란 의미와 가장 관계가 깊은 것은? (2점)

- ① 애널리틱스 ② 하드스킬
③ 소프트스킬 ④ 데이터과학자

25. 다음 중 데이터 분석의 윤리적 측면, 투명성과 건전성 등과 관련이 깊은 항목을 모두 고르면? (3점)

ㄱ. DataKind 프로젝트	ㄴ. RunMyCode
ㄷ. MapReduce	ㄹ. 더만의 모델링을 위한 선언

- | | |
|-------------------------|-------------------------------|
| ① \neg, \perp | ② \perp, \top |
| ③ \neg, \perp, \vdash | ④ $\neg, \perp, \top, \vdash$ |