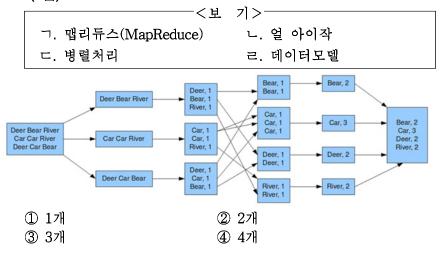
2과목	데 이 터 과 학 입 문	(36~60)	
출제위원 : 방송대 장영재			
출제범	위 : 교재 1~8장		

- 36. 데이터과학의 역사는 데이터 수집의 역사의 관점에서 살펴볼 수 있다. 다음 중 각 사례별 연결이 옳은 것은? (2점)
  - ① 로널드 피셔 실험계획법
  - ② 조지 갤럽 고객 거래 데이터
  - ③ 윌리엄 페어 공정과정 데이터
  - ④ 다구치 겐이찌 통제된 실험
- 37. 다음 보기 중 아래 그림과 관계가 깊은 것은 모두 몇 개인가?(3점)



38. 다음 <보기>의 내용과 가장 관계 깊은 것은? (2점)

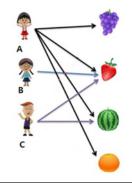
<보 기>----

넷플릭스 대상 영향, 데이터과학자, 데이터과학 경연대회 플랫폼

- ① 클라우드 컴퓨팅
- ② 협업필터링
- ③ 데이터웨어하우스
- ④ 캐글(Kaggle)
- 39. 다음 설명과 가장 관계 깊은 것은? (2점)

품질향상이라는 과제에서 시작된 장비 또는 기계에서 생산되는 데이터 수집으로 오늘날의 IoT(Internet of Things)와 유사하다고 볼 수 있다.

- ① 실험계획법
- ② 고객거래 데이터
- ③ 공정과정 데이터
- ④ 기계학습
- 40. 다음 중 아래 그림에 대해 옳은 설명으로만 짝지어진 것은? (4점)



- ㄱ. 협업필터링의 한 종류이다.
- 니. 기호정보를 바탕으로 많은 콘텐츠를 필터링하여 사용자 성향에 가장 가까운 콘텐츠를 추천한다.
- 다. 아이템기반 필터링에 따르면 B에게 포도, 수박, 오렌지를 추천하게 된다.
- ① ¬, ∟
- ② ∟, ⊏
- ③ ¬, ⊏
- ④ ¬, ∟, ⊏

41. 다음 중 빅 데이터의 특징을 기존의 데이터 및 데이터 분석과 비교하였을 때, 그 차이점으로 가장 옳지 **않은** 것은? (3점)

기존 데이터			빅데이터	
1	내부	vs	외부 혹은 소셜	
2	구조적	vs	비구조적	
3	실시간 생성데이터	vs	축적된 데이터	
4	중앙처리 서버	vs	분산처리	

- 42. 응용 프로그램에서 사용 가능하도록 운영체제나 프로그래밍 언어 가 제공하는 기능을 제어할 수 있도록 만들어 놓은 인터페이스를 의미하는 것은? (2점)
  - ① 파싱

② GUI

3 API

- ④ DBMS
- 43. 데이터베이스관리시스템에 관한 설명으로 옳지 않은 것은? (3점)
  - ① 자료와의 관계성을 정의하기 때문에 자료 통합을 억제할 수 있음
  - ② 데이터의 접근이 용이함
  - ③ 데이터 통제를 강화하여 오작동 방지나 보안 등의 면을 잘 관리
  - ④ 데이터의 논리적, 물리적 독립성을 확보할 수 있음
- 44. 온라인을 통해 다수의 참여에 의한 서비스, 아이디어를 얻는 과정으로, 캐글(Kaggle)에서 예측모형 혹은 데이터 분석결과를 얻기 위해 사용하는 방식과 관계가 깊은 것은? (2점)
  - ① 협업 필터링
  - ② 데이터 준설
  - ③ 예측 배송
  - ④ 크라우드소싱
- \*\* 다음 R 프로그램 등에서 사용되는 데이터 구조에 관한 질문에 답하시오.  $(45\sim47)$
- 45. 다음 중 데이터프레임에 대한 설명으로 옳은 것을 모두 고르면? (3점)
  - ㄱ. 형태(mode)가 일반화된 행과 열로 이루어져 있음
  - ㄴ. 하나의 객체에 여러 종류의 자료가 들어갈 수 있음
  - 다. 각 열은 각각 변수와 대응
  - ㄹ. 분석이나 모형 설정에 적합한 자료 객체
  - ① 7, ∟, ⊏
  - ② ㄴ, ㄷ, ㄹ
  - ③ 7, ⊏, ⊒
  - ④ 7, ∟, ⊏, ⊒
- 46. 아래와 같은 자료 구조에 관한 설명으로 옳은 것은 모두 몇 개인가? (3점)

- ㄱ. 동일한 형태의 원소로 이루어진 객체
- ㄴ. ", , 1"은 첫 번째 행렬의 첫 번째 열을 의미
- ㄷ. 서로 다른 길이의 성분이 존재할 수 있음
- ㄹ. 리스트를 확장시킨 객체를 의미
- ① 1개

② 2개

③ 3개

④ 4개

- 47. 다음 중 행렬에 대한 설명으로 옳지 <u>않은</u> 것은 모두 몇 개인가? (3점)
  - □. 동일한 형태(mode)로 구성된 2차원의 데이터 구조로 행의 차원과 열의 차원을 갖고 있음
  - ㄴ. 벡터의 속성을 포함하고 있음
  - 다. 문자형과 수치형 원소를 가질 수 있으나 논리형 원소는 가질 수 없음
  - ① 0개

② 1개

③ 2개

- ④ 3개
- 48. 다음 중 아래 보기의 설명과 가장 관계가 깊은 것은? (2점)

-<보 기><sup>.</sup>

KOSIS, ECOS, 통계 데이터베이스 형태로 제공, 공개 API로 제공

- ① 외부 데이터
- ② 내부 데이터
- ③ 검색 데이터
- ④ 웹크롤링
- 49. 다음 중 데이터베이스 관리 시스템의 관리 기능에 관한 설명으로 옳은 것을 모두 고르면? (3점)
  - ¬. 정의(Definition): 특정한 데이터를 검색하기 위한 질의, 데이터베이스의 갱신, 삽입, 삭제 등을 체계적으로 관리하기 위한인터베이스 제공
  - 도작(Manipulation): 모든 응용 프로그램이 요구하는 데이터 구조를 지원하기 위해 데이터베이스에 저장될 데이터에 대한 형식, 구조, 제약조건들을 명시하는 기능
  - 다. 제어(Control): 데이터베이스를 접근하는 갱신, 삽입, 삭제 등의 작업이 정확하게 수행되어 무결성을 유지하도록 하고 하드웨어나 소프트웨어의 오동작으로부터 시스템을 보호
  - 1 7

② ∟, ⊏

③ ⊏

- ④ ¬, ∟, ⊏
- 50. 다음 중 아래 그림과 관련한 설명 중 옳은 것을 모두 고르면? (4점)



- ㄱ. 데이터 분석 도입의 성공요소 중 하나이다.
- ㄴ. 분석대상에 대한 명확한 정의가 필요하다는 점을 고려한 것이다.
- 다. 분석 질문의 우선적 정의(question first) 방식이다.
- 리. 업무 간 장벽이 사라지고 상호 업무를 이해하고 원활하게 처리하기 위한 내재화 과정이다.
- ① 7. ㄴ. ㄸ
- ② ㄴ, ㄷ, ㄹ
- ② =, =, = ③ ¬, =, =
- ④ 기, ㄴ, ㄷ, ㄹ
- 51. 다음 중 분석 핵심 문제(기회)의 발굴 방법으로 옳은 것을 모두 고르면? (3점)
  - ㄱ. 분석업무 내재화법
- ㄴ. 상향식접근법
- ㄷ. 하향식접근법
- ㄹ. 벤치마킹 풀 작성법
- ① 7, ∟, ⊏
- ② ㄴ, ㄷ, ㄹ
- ③ 7, ⊏, ㄹ
- ④ 7, ∟, ⊏, 큳
- 52. 다음 중 한 수치의 변화를 통해 다른 수치의 변화를 예측할 수 있는 방법과 가장 거리가 먼 것은? (2점)
  - ① 트리맵
  - ② 관계 시각화
  - ③ 산점도
  - ④ 버블 차트

53. 다음 중 아래 통계적 문제해결 과정에 대한 설명으로 가장 옳지 않은 것은? (4점)



- ① (a)에 들어갈 과정은 통계적 해결책을 얻는 과정이다.
- ② (a)의 과정은 데이터를 바탕으로 모형을 구축하고 해결책을 찾는 것이다.
- ③ (a)의 과정에서는 전문적인 식견에서 면밀한 검토와 검증을 필요로 한다.
- ④ (a)의 과정은 분석목적의 일관성을 유지하며 의사결정 참여자들에게 이해하기 쉬운 언어로 전달하는 과정을 포함한다.
- 54. 다음 중 데이터 시각화의 방법이 가장 옳게 묶인 것은? (3점)
  - ① 시간시각화 지도매핑
  - ② 비교시각화 체르노프 페이스
  - ③ 공간시각화 파이차트
  - ④ 분포시각화 산점도
- 55. 다음 중 빅데이터 시각화 3단계 프로세스를 순서대로 나열한 것으로 가장 알맞은 것은? (3점)
  - ① 정보 탐색 → 정보시각표현 → 정보시각화
  - ② 정보 구조화 → 정보 시각화 → 정보 시각표현
  - ③ 정보 시각화  $\rightarrow$  정보 구조화  $\rightarrow$  정보 시각표현
  - 4 정보 시각화  $\rightarrow$  정보 시각표현  $\rightarrow$  정보 구조화
- 56. 다음 중 데이터마이닝 기법을 '교사' 역할에 해당하는 출력변수 존재 여부에 따라 분류한 것은? (2점)
  - ① 분류분석과 예측분석
  - ② 감독학습(관리학습)과 자율학습
  - ③ 정형분석과 비정형분석
  - ④ 모수적 방법과 비모수적 방법
- 57. 다음 데이터마이닝 기법을 모수적 모형과 알고리즘 접근방법으로 나눌 때 이에 관한 설명 중 가장 옳지 **않은** 것은? (3점)
  - ① 모수적 모형 접근 방법은 단순 선형 회귀분석 Y=a+bx와 같이 모수 a와 b를 과거 데이터로부터 적합하는 방법이다.
  - ② 모수적 모형 접근 방법은 결과에 대한 해석이 대체로 용이하며 결과가 복잡하지 않다.
  - ③ 로지스틱 모형은 알고리즘 접근 방법으로 분류한다.
  - ④ 알고리즘 접근 방법은 데이터 복잡성이 높아도 적용이 가능하다.
- 58. 다음 중 아마존의 예측배송과 구글의 스마트카와 가장 관계가 깊은 것은? (3점)
  - ① 데이터 기반의 맞춤형 개인화 서비스이다.
  - ② 협업 필터링 기반의 추천 서비스 방식이다.
  - ③ 크라우드소싱 기반의 웨어러블 디바이스이다.
  - ④ 위험조기포착을 통한 사회 안정을 위한 서비스이다.
- 59. 다음 중 엠마누엘 더만의 '모델링을 위한 선언'과 가장 관련이 깊은 것은? (3점)
  - ① 데이터분석의 경제적 가치 창출
  - ② 데이터과학자의 윤리적 책임
  - ③ 빅데이터 기반의 예측력 향상
  - ④ 데이터과학자의 하드스킬 함양
- 60. 다음 중 데이터과학자의 윤리 및 소양과 관련된 설명으로 가장 옳지 **않은** 것은? (3점)
  - ① 데이터과학자의 윤리적 활동을 위해 비영리단체의 활동이 확대되고 있다
  - ② 데이터과학자의 하드스킬 함양을 위하여서는 인문학적 소양을 갖추어야 한다.
  - ③ 인문학적 특성은 데이터로부터 통찰력을 얻고자 하는 데이터 과학의 목적에 부합한다.
  - ④ 데이터과학자는 데이터과학의 오용을 방지하는 감시자의 역할 도 수행할 수 있다.