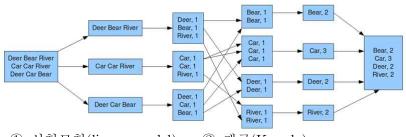
2과목
 데이터과학입문
 (36~60)

 출제위원: 방송대 장영재

 출제범위: 교재1~8장

- 36. 다음 중 현대적 의미의 데이터 분석의 기초를 세웠으며 현대의 실험계획법(design of experiment)을 창시한 사람은? (2점)
 - ① 죠지 갤럽
- ② 로널드 피셔
- ③ 다구찌
- ④ 윌리엄 페어
- 37. 다음 그림은 빅데이터 시대의 주요 기술 중 하나로서 엄청난 규모의 데이터를 처리, 분석할 수 있는 맵리듀스 알고리즘이다. 이와 가장 관계가 깊은 것은? (3점)



- ① 선형모형(linear model)
- ② 캐글(Kaggle)
- ③ 클러스터링
- ④ 하둡(hadoop)
- 38. 아래 보기는 안전한 사회 구현을 위해 데이터분석을 활용하는 사례이다. 이러한 분야와 가장 관계가 <u>먼</u> 것은? (2점)

____<보 기>___

위험 조기포착, 범죄 패턴 분석 및 예측, 위험감지 분석

- ① 히트맵(Heatmap)
- ② 프레드폴(PredPol)
- ③ 스마트 빅보드
- ④ 링크드인
- 39. 다음 중 맞춤형 개인화 서비스와 가장 관계가 깊은 것은? (2점)
 - ① 예측배송 시스템
- ② 구글 독감 트렌드
- ③ 페이스북
- ④ 프레드폴
- 40. 데이터마이닝과 관련하여 알고리즘 접근 방식에 대한 설명으로 가장 옳지 <u>않은</u> 것은? (3점)
 - ① 알고리즘에 의해 정해진 방식에 따라 계산된 결과로 분석되는 방식이다.
 - ② 의사결정나무, 신경망, 배깅, 딥러닝 등의 방법이 이에 속한다.
 - ③ 알고리즘에 의해 데이터로 분해(decomposition)한다는 표현을 사용한다.
 - ④ 알고리즘의 속성이나 방법에 대한 이론적 근거를 모르고 남용하는 경우, 과도적합(over-fitting) 등의 우를 범할 수 있다.
- 41. 다음 중 빅데이터의 특징을 기존의 데이터 및 데이터 분석과 비교하였을 때, 그 차이점으로 가장 옳지 <u>않은</u> 것은? (3점)

데이터

빅데이터

- ① 내부
- vs. 외부 혹은 소셜
- ② 구조적
- vs. 비구조적
- ③ 축적된 데이터 vs. 실시간 생성 데이터
- ④ 분산처리 서버 vs. 클라우드 환경
- 42. URL의 패턴을 분석하거나 HTML 문법을 알고 있을 경우 HTML 코드에서 원하는 데이터를 수집할 수 있다. 이 때 사용되는 방법과 가장 관계가 깊은 것은? (2점)
 - ① 파싱

- ② 웹크롤링
- ③ 데이터베이스
- ④ 데이터병합화

- 43. 다음 중 웹스크래핑(web scraping)에 관한 설명으로 가장 옳은 것은? (3점)
 - ① 웹문서를 가공하여 정보를 추출하는 과정이다.
 - ② 웹스크래핑을 하기 위해서는 웹크롤링 과정을 거쳐야 한다.
 - ③ 기초가 되는 URL seed들을 저장한 뒤 웹페이지의 하이퍼링 크를 인식하여 URL을 갱신한다.
 - ④ 개인이 검색을 위해 입력한 자료를 축적한 데이터를 얻는 과정이다.
- 44. 다음 중 다수의 사용자가 데이터베이스 내의 데이터에 접근할 수 있도록 도와주는 소프트웨어의 집합을 의미하는 것은? (2점)
 - ① 데이터 마이닝
- ② 데이터베이스 모델
- ③ 릴레이션
- ④ 데이터베이스 관리 시스템
- % 다음 R 프로그램 등에서 사용되는 데이터 구조에 관한 질문에 답하시오. $(45{\sim}47)$
- 45. 다음 중 데이터프레임에 관한 설명으로 옳은 것은 모두 몇 개인가? (3점)
 - 가. 형태(mode)가 일반화된 행과 열로 이루어져 있음
 - 나. 행렬과 달리 형태가 일반화 되어 길이가 다른 성분이 존재
 - 다. 데이터프레임의 각 열은 각각의 변수와 대응
 - 라. 분석이나 모형 설정에 적합한 자료 객체
 - ① 1개

② 2개

③ 3개

- ④ 4개
- 46. 아래와 같은 배열 A에 대하여 'A[2,1,2]/A[1,2,1]'과 같은 연산을 실행할 때, 출력되는 값은? (4점)

배열 A

행렬 1 1 4 7 10 2 5 8 11 3 6 9 12 행렬 2 13 16 19 22 14 17 20 23 15 18 21 24

 $\bigcirc \frac{2}{5}$

 $2\frac{1}{8}$

 $3\frac{7}{2}$

- 4
- 47. 다음 중 리스트에 대한 설명으로 옳은 것은 모두 몇 개인가? (3점)
 - 가. 서로 다른 형태(mode)의 데이터로 구성된 객체
 - 나. 리스트를 구성하는 성분은 길이가 다를 수 있음
 - 다. 문자형, 수치형, 논리형 자료가 혼합되어 한 리스트 내에 존재할 수 있음
 - ① 0개

② 1개

③ 2개

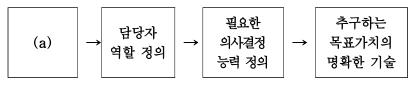
④ 3개

- 48. 다음 중 낮은 품질의 데이터의 예와 그의 해결방안이 적절하게 연결되지 **않은** 것은? (3점)
 - ① 결측치나 속성이 알려지지 않은 데이터 결측치의 보완 문제 등에 주의
 - ② 오염된 데이터 데이터클리닝 과정을 고려
 - ③ 잘못된 해석을 초래하는 데이터 데이터 설명을 충분히 숙지
 - ④ 왜곡되고 잘못된 인코딩에 의한 데이터 중요성을 지니는 부 분을 선별
- 49. 다음 중 데이터 분석기획과 관련하여 분석의 핵심문제를 찾는 일련의 과정에 속하지 **않는** 것은? (3점)
 - ① 하향식 접근방식
 - ② 선택과 집중
 - ③ 상향식 경로접근 방식
 - ④ 업무별 분석 대상 풀로부터 벤치마킹
- **50.** 다음 데이터 품질관리와 관련한 설명 중 옳은 것은 모두 몇 개인가? (4점)
 - 가. 좋은 품질의 데이터란 데이터의 적시성이 담보되는 데이 터를 의미한다.
 - 나. 좋은 품질의 데이터란 업계 표준을 준수하는 완전하고 일 관성 있는 정확한 데이터를 의미한다.
 - 다. 좋은 데이터 품질은 상품이나 용역, 서비스의 공급측면에 서 비용 증가의 단점은 있으나 오류나 오차를 줄이는 궁 정적 측면이 있다.
 - 라. 데이터베이스관리시스템이란 데이터의 품질관리를 효과적 으로 돕는 관리 시스템이다.
 - ① 1개

② 2개

③ 3개

- ④ 4개
- 51. 다음 유저스토리의 작성 과정에서 (a)에 들어갈 내용으로 가장 알맞은 것은? (3점)



- ① 측정가능한 형태로 목표 정의
- ② 핵심분석 역량 강화
- ③ 분석기회의 식별
- ④ 사용자 편람 제공
- 52. 다음 중 의사결정자가 의사결정을 효과적으로 할 수 있도록 데이터에 기반한 정보를 제공하는 시스템의 구성요소와 가장 거리가 <u>먼</u> 것은? (3점)
 - ① 데이터베이스 시스템
- ② 모델베이스 시스템
- ③ 유저스토리
- ④ 사용자
- 53. 다음 프레젠테이션에 관한 설명 중 옳은 것은 모두 몇 개인가?(4점)
 - 가. 프레젠테이션의 3요소인 3P는 장소(Place), 청중(People), 제안(Proposal) 등이다.
 - 나. 데이터기반의 올바른 프레젠테이션을 위해서는 데이터 자체의 이해보다는 데이터 환경변화를 잘 이해하고 분석할 능력이 요구된다.
 - 다. 프레젠테이션을 단지 시각화에만 국한시킬 경우 분석목표 에 대한 이해부족이나 잘못된 의사결정에 이를 수 있다.
 - ① 0개
- ② 1개

③ 2개

④ 3개

- 54. 다음 중 데이터 시각화의 효과와 가장 거리가 먼 것은? (3점)
 - ① 정보습득 시간의 절감
 - ② 보는 이들의 흥미 유발
 - ③ 데이터 분석의 구체화
 - ④ 정보의 빠른 확산
- 55. 다음 빅데이터 시각화 3단계 프로세스에 관한 설명 중 바르게 짝지어지지 **않은** 것은? (3점)
 - ① 1단계 정보 구조화 : 데이터를 수집, 정제하는 과정
 - ② 2단계 정보 시각화 : 분석도구에서 만든 결과물에 별도의 그래픽 요소를 추가
 - ③ 3단계 정보 시각표현 : 시각화의 의도를 강화해서 전달
 - ④ 3단계 정보 시각표현 : 전문 그래픽 도구를 활용
- 56. 다음 중 방대한 데이터를 정제하여 통계 및 수학적 기술, 패턴 인식 기술 등을 사용함으로써 의미 있는 연관성과 패턴 및 추세를 발견하는 과정을 의미하는 것은? (2점)
 - ① 분석기획
 - ② 데이터시각화
 - ③ 데이터베이스 모델링
 - ④ 데이터마이닝
- 57. 다음 중 데이터의 특성에 맞게 전체의 관점에서 부분이 차지하는 비율 및 관계를 보여주는 시각화 기법과 가장 거리가 먼 것은? (2점)
 - ① 파이차트
 - ② 체르노프 페이스
 - ③ 히스토그램
 - ④ 트리맵
- 58. 다음 중 입출력 간의 관계를 결정하는 시스템에 대한 유용한 근사 시스템을 구하는 것을 목표로 하는 데이터마이닝 기법과 가장 거리가 <u>면</u> 것은? (2점)
 - ① 회귀분석
 - ② 의사결정나무
 - ③ 장바구니 분석
 - ④ 신경망
- 59. 다음 중 데이터과학에서의 위기요인으로서 아래의 설명과 가장 관련 깊은 것은? (3점)

인과성이 과학적으로 완전히 밝혀지지 않았는데도 유전자 분석을 통해 질병과의 연관성이 높다는 이유로 의사가 환자의 수술을 거부할 수 있다.

- ① 정보유출
- ② 연구부정행위
- ③ 사생활침해
- ④ 정보오용과 분석결과 맹신
- 60. 다음 중 'DataKind'나 'RunMyCode' 등과 가장 거리가 <u>먼</u> 설명은? (3점)
 - ① DataKind는 데이터 분석의 윤리적 가치 함양에 필요한 지원을 하고 있다.
 - ② RunMyCode는 데이터 및 분석알고리즘의 투명성을 위한 활동이다.
 - ③ 데이터과학 및 데이터 분석과 관련한 비영리단체의 활동이다.
 - ④ 데이터분석 결과를 바탕으로 한 경연, 검증 등을 통해 소프트 웨어 품질을 제고하고 있다.