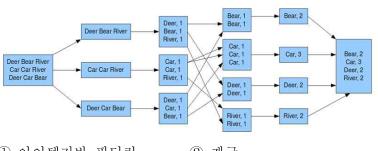
| 2과목            | 데 이 터 과 학 입 문 | $(36 \sim 60)$ |  |
|----------------|---------------|----------------|--|
| 출제위원 : 방송대 장영재 |               |                |  |
| 출제범            | 위 : 교재 1~8장   |                |  |

- 36. 여러 사람들의 콘텐츠 평가 데이터, 즉 기호 정보(taste information) 를 이용하여 많은 콘텐츠를 필터링하여 사용자 성향에 가장 가까운 콘텐츠를 추천하는 시스템을 의미하는 것은? (2점)
  - ① 데이터 클리닝
  - ② 협업 필터링
  - ③ 데이터마이닝
  - ④ 맵리듀스
- 37. 다음 그림은 빅데이터 시대의 주요 기술 중 하나로서 엄청난 규 모의 데이터를 처리, 분석할 수 있는 알고리즘을 나타낸다. 이와 가장 관계가 깊은 것은? (3점)



- ① 아이템기반 필터링
- ② 캐글
- ③ 맵리듀스
- ④ R
- 38. 다음 분석방법들을 활용하여 탐지하고 예측할 수 있는 분야와 가장 관계가 <u>먼</u> 것은? (2점)

예측 모델링에 의한 위험도 점수, 관계망 분석, 데이터 시각화

- ① 신용카드 사기
- ② 품질보증(warranty) 관련 사기
- ③ 자금세탁
- ④ 독감 트렌드
- 39. 분석이나 의사결정 지원을 위한 본격적인 데이터 축적, 관리 및 활용의 기반을 다지기 시작한 것으로, 비즈니스 목적에 부합 하는 데이터 분석을 수행하는 상향식(bottom-up)뿐 아니라 수 집된 다양한 데이터의 활용방안을 고민하는 하향식(top-down) 접근이 요구되기도 하는 것은? (2점)
  - ① 데이터 클리닝
  - ② 데이터웨어하우스
  - ③ 데이터마이닝
  - ④ 기계학습
- 40. 데이터마이닝과 관련하여 알고리즘 접근 방식에 대한 설명으로 가장 옳은 것은**? (3**점)
  - ① 특별히 정해진 방식을 따르지 않고 계산된 결과에 의해서만 분석되는 방식이다.
  - ② 알고리즘에 의해 데이터로 병합(merging)한다는 표현을 사 용하다
  - ③ 결과의 해석이 용이하지만 데이터 복잡성이 높으면 적용이 불가능하다.
  - ④ 알고리즘의 속성이나 방법에 대한 이론적 근거를 모르고 남 용하는 경우, 과도적합(over-fitting) 등의 우를 범할 수 있다.

41. 다음 중 빅데이터의 특징을 기존의 데이터 및 데이터 분석과 비교하였을 때, 그 차이점으로 가장 옳은 것은? (3점)

|   | 기존 데이터     |     | 빅데이터     |
|---|------------|-----|----------|
| 1 | 외부         | vs. | 내부 혹은 소셜 |
| 2 | 비구조적       | VS. | 구조적      |
| 3 | 실시간 생성 데이터 | vs. | 축적된 데이터  |

42. 한 개 이상의 자료가 논리적으로 연결되어 축적된 것을 의미 하는 것으로서 이 축적 과정에서 구조화 방법을 이용함으로써 자료 검색과 갱신의 효율성을 최대한 확보할 수 있도록 구축 하는 것은? (2점)

VS.

① 파싱

- ② 웹크롤링
- ③ 데이터베이스

④ 중앙처리 서버

④ 데이터병합화

분산처리

- 43. 데이터베이스관리시스템에 관한 설명으로 가장 옳은 것은? (3점)
  - ① 자료와의 관계성을 정의하기 때문에 자료 통합을 억제할 수 있음
  - ② 데이터의 접근이 매우 어려움
  - ③ 데이터 통제를 강화하여 오작동 방지나 보안 등의 면을 잘 관리
  - ④ 데이터의 논리적, 물리적 독립성을 억제
- 44. 데이터베이스 설계과정에서 데이터의 구조를 표현하기 위해 사용되는 도구로서 현실세계의 데이터를 컴퓨터에서의 데이터 로 표현하는 개념적 도구로 이해할 수 있는 것은? (2점)
  - ① 데이터 개념화
  - ② 데이터 준설
  - ③ 데이터 현실화
  - ④ 데이터베이스 모델(데이터 모델)
- ※ 다음 R 프로그램 등에서 사용되는 데이터 구조에 관한 질문에 답하시오. (45~47)
- 45. 다음 설명과 가장 관계가 깊은 데이터구조는? (3점)
  - 가. 형태(mode)가 일반화된 행과 열로 이루어져 있음
  - 나. 하나의 객체에 여러 속성(mode)의 자료가 들어갈 수 있음
  - 다. 각 열은 각각 변수와 대응
  - 라. 분석이나 모형 설정에 적합한 자료 객체
  - ① 벡터
- ② 행렬
- ③ 배열

- ④ 데이터프레임
- 46. 아래와 같은 배열이 존재할 때, 이와 관련하여 ",, 2 "가 의 미하는 것은? (3점)

| 행렬 1        |
|-------------|
| % 될 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 |

- ① 배열의 두 번째 행
- ② 배열의 두 번째 열
- ③ 배열의 두 번째 행렬
- ④ 배열의 두 번째 행렬의 두 번째 원소

- **47.** 다음 중 리스트에 대한 설명으로 옳지 **않은** 것은 모두 몇 개인가**? (3**점)
  - 가. 서로 다른 형태(mode)의 데이터로 구성된 객체
  - 나. 리스트를 구성하는 성분은 길이가 다를 수 있음
  - 다. 문자형과 수치형 자료만 한 리스트 내에 존재할 수 있음
  - ① 0개

② 1개

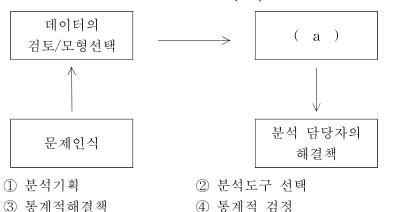
③ 2개

- ④ 3개
- 48. 다음 중 데이터를 사용하기에 적절한 환경을 뜻하는 데이터 품질과 가장 관련이 <u>없는</u> 것은? (3점)
  - ① 접근 가능성
- ② 적시성
- ③ 적합성
- ④ 환가성
- **49.** 다음 중 데이터 분석 도입의 세 가지 성공요소에 속하지 <u>않는</u> 것은**? (3**점)
  - ① 분석 질문의 우선적 정의
  - ② 분석업무의 집중화
  - ③ 분석업무의 내재화
  - ④ 선택과 집중
- 50. 다음 데이터분석과 관련한 설명 중 옳은 것은 모두 몇 개인가? (4점)
  - 가. 분석기회를 정의하는 단계는 주요한 의사결정의 대상이 되는 분석의 핵심문제를 발견하고 성공적인 성과물을 만 들 수 있도록 청사진을 제시하는 단계를 의미한다.
  - 나. 유저스토리란 업무목적을 효과적으로 달성할 수 있도록 데이터의 품질을 지속적으로 관리하는 프로세스이다.
  - 다. 목표가치의 구체화란 목표를 측정 불가능한 형태로 개념 화, 추상화하는 과정을 의미하는 것이다.
  - ① 0개

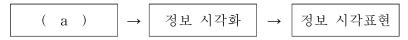
② 1개

③ 2개

- ④ 3개
- 51. 다음 중 의사결정 지원 시스템의 구성 요소로 보기 <u>힘든</u> 것은? (3점)
  - ① 데이터베이스 시스템
  - ② 모델베이스 시스템
  - ③ 데이터 품질관리 시스템
  - ④ 사용자
- 52. 다음 중 도식적인 형태 안에 추상적 혹은 정량적으로 표현된 속성(attributes)이나 변수(variables)를 포함한 정보와 가장 관련이 깊은 것은? (3점)
  - ① 데이터 시각화
- ② 데이터마이닝
- ③ 데이터 계량화
- ④ 데이터 모델링
- 53. 다음은 통계적 문제해결 과정을 도식화한 것이다. (a)에 들어갈 내용으로 가장 알맞은 것은? (3점)



- 54. 다음 중 데이터 시각화의 목적으로 옳게 묶인 것은? (3점)
  - ① 데이터 수집과 분석
  - ② 데이터베이스 구축과 관리
  - ③ 데이터 분석과 의사소통
  - ④ 데이터 수집과 의사소통
- 55. 다음은 빅데이터 시각화 3단계 프로세스를 도식화한 것이다. (a)에 들어갈 내용으로 가장 알맞은 것은? (3점)



- ① 정보 탐색
- ② 정보 분산화
- ③ 정보 구조화
- ④ 정보 선택
- 56. 다음 중 성공적인 시각화의 조건으로 적당하지 않은 것은? (2점)
  - ① 정보전달
  - ② 아름다움
  - ③ 설득력
  - ④ 보안성
- 57. 다음 중 데이터 내 변수들의 비교를 통해 데이터를 전체적으로 조망할 수 있는 비교시각화 기법과 가장 거리가 **먼** 것은? (3점)
  - ① 히트맵
  - ② 체르노프 페이스
  - ③ 평행좌표계
  - ④ 트리맵
- 58. 다음 데이터마이닝 기법 중 지도학습이나 자율학습에 속하지 않는 비정형 기법만으로 묶인 것은? (3점)
  - ① 사회연결망분석과 회귀분석
  - ② 연관성 분석과 텍스트마이닝
  - ③ 사회연결망분석과 텍스트마이닝
  - ④ 연관성 분석과 사회연결망분석
- 59. 다음 중 아래의 설명과 가장 관련 깊은 데이터과학에서의 위기 요인은? (3점)

어떤 사람이 특정한 사회적·경제적 특성을 지닌 집단에 속한다는 이유로 자신의 신용과 상관없이 부당하게 대출이 거절될 수 있음

- ① 사생활침해
- ② 정보오용과 분석결과 맹신
- ③ 정보유출
- ④ 연구부정행위
- 60. 다음 중 'DataKind'나 'RunMyCode' 등과 가장 관계가 깊은 설명은? (3점)
  - ① 데이터과학 및 데이터분석의 윤리적 활용 및 투명성을 위한 비영리단체의 활동이 확대되고 있다
  - ② 데이터과학 및 데이터분석의 효과적인 시스템 구축을 위한 자료 제공 기업의 활동이 확대되고 있다
  - ③ 데이터과학 및 데이터분석을 돕는 관리 시스템의 활용이 급증하고 있다.
  - ④ 데이터분석 결과의 검증을 통해 데이터과학을 발전시키기 위한 소프트웨어 제작 활동이 확대되고 있다