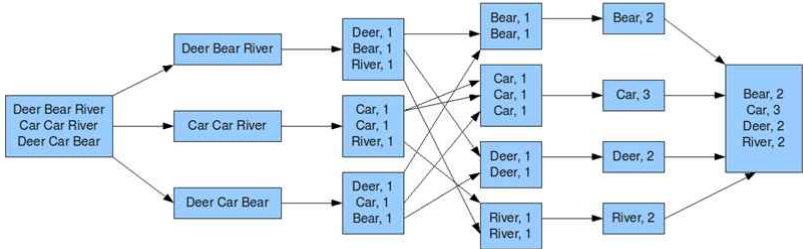


2과목	데이터과학입문	(36~60)
출제위원 : 방송대 장영재		
출제범위 : 교재1~8장		

36. 다음 중 현대적 의미의 데이터 분석의 기초를 세웠으며 현대의 실험계획법(design of experiment)을 창시한 사람은? (2점)

- ① 조지 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 먼 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 않은 것은? (3점)

- ① 알고리즘에 의해 정해진 방식에 따라 계산된 결과로 분석되는 방식이다.
- ② 의사결정나무, 신경망, 배깅, 덤퍼닝 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 않은 것은? (3점)

- 데이터 vs. 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 하이퍼링크를 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. (45~47)

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 존재다. 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

- ①  $\frac{2}{5}$
- ②  $\frac{1}{8}$
- ③  $\frac{7}{2}$
- ④ 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

