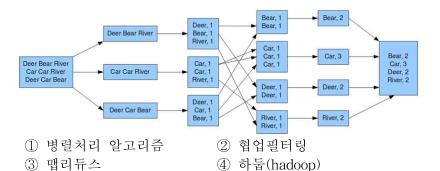
출제범위:교재 1~8장

 2과목
 데이터과학입문
 (36~60)

 출제위원: 방송대 장영재

- 36. 다음 중 서비스를 이용하는 고객으로부터 제공되는 데이터를 축적, 분석하여 의사결정을 시작한 사례와 가장 거리가 <u>먼</u> 것은? (3점)
 - ① 과거 결과에 대한 데이터를 활용하여 미래의 행동을 예측
 - ② 사람들의 생각을 질문이나 문항을 통해 데이터로 수집
 - ③ 고객의 행동 및 기본 정보를 취득하여 의사결정에 활용
 - ④ 데이터 분석 결과로 인간의 편향적인 의사결정을 대체
- 37. 다음 그림은 빅데이터 시대의 주요 기술 중 하나로서 엄청난 규모의 데이터를 처리, 분석할 수 있는 알고리즘을 나타낸다. 이와 가장 거리가 먼 것은? (2점)



38. 다음과 같은 범죄 활동을 탐지하고 예측할 수 있는 분석방법들과 가장 거리가 **먼** 것은**? (2**점)

신용카드 사기, 품질보증(warranty) 관련 사기, 자금세탁

- ① 예측 모델링에 의한 위험도 점수
- ② 관계망 분석
- ③ 데이터 시각화
- ④ 실험계획법
- **39**. 다음 보기 중 데이터웨어하우스에 관한 설명으로 옳은 것은 모두 몇 개인가**? (4**점)
 - 가. 분석이나 의사결정 지원을 위한 본격적인 데이터 축적, 관리 및 활용의 기반이 됨
 - 나. 비즈니스 목적에 부합하는 데이터 분석을 수행하는 상향식 (bottom-up)접근이 요구됨
 - 다. 수집된 다양한 데이터의 활용방안을 고민하는 하향식 (top-down)접근이 요구됨
 - ① 0개

② 1개

③ 2개

- ④ 3개
- 40. 다음 데이터마이닝과 관련한 설명 중 가장 옳지 않은 것은? (3점)
 - ① 모수적 모형 접근방법과 알고리즘 접근방법으로 나눌 수 있다.
 - ② 대표적인 알고리즘 접근방법으로 의사결정나무, 신경망, 로지스틱 회귀모형을 꼽을 수 있다.
 - ③ 모수적 방법은 결과의 해석이 용이하며 결과가 대체로 복잡하지 않다.
 - ④ 알고리즘 접근방법의 경우 알고리즘의 속성이나 방법에 대한 이론적 근거를 모르고 남용하는 경우, 과도적합(over-fitting) 등의 우를 범할 수 있다.
- 41. 다음 중 아래 보기와 가장 관계가 깊은 데이터과학 응용 사례는? (2점)
 - 가. 사용자의 행동 분석 및 예측모형
 - 나. 구매하려는 제품 종류 및 가격을 고려한 구매패턴 분석
 - 다. 최적 제품가격 제시 모형
 - ① 구글의 스마트카
- ② 프레드폴
- ③ 넷플릭스 추천시스템
- ④ 아마존의 예측배송

- 42. 다음 중 구글의 독감트렌트, 바이오인포매틱스(bioinformatics), 착용장치(wearable device) 등과 가장 관계가 깊은 것은? (2점)
 - ① 범죄예방 시스템
 - ② 맞춤형 개인화 서비스
 - ③ 건강한 삶
 - ④ 연결된 사회
- **43.** 다음 중 데이터베이스 모델에 관한 설명으로 가장 거리가 <u>먼</u> 것은**?** (**3**점)
 - ① 데이터베이스 설계과정에서 데이터 구조를 표현
 - ② 현실세계의 데이터를 컴퓨터에서의 데이터로 표현
 - ③ 데이터 통제를 강화하여 오작동 방지나 보안 등의 면을 잘 관리
 - ④ 계층형, 네트워크형, 관계형 등이 존재
- 44. 다음 보기의 내용과 가장 관계가 깊은 것은? (2점)
 - 가. 질적인, 혹은 양적인 변수들의 가치 집합
 - 나. 정보의 조합
 - 다. 어원에 따르면 주어진 것들(things given)을 의미
 - ① 기계학습
- ② 통계
- ③ 컴퓨터
- ④ 데이터
- ※ 다음 R 프로그램 등에서 사용되는 데이터 구조에 관한 질문에 답하시오. (45~48)
- 45. 다음 중 리스트에 관한 설명으로 옳은 것은 모두 몇 개인가? (3점)
 - 가. 리스트를 구성하는 성분은 길이가 다를 수 없다.
 - 나. 서로 다른 형태(mode)의 데이터로 구성된 객체이다.
 - 다. 일반화된 행렬로서 모든 자료 간 사칙 연산이 가능하다.
 - 라. 문자형, 수치형, 논리형 자료가 혼합되어 한 리스트 내에 존 재할 수 있다.
 - ① O개

② 1개

③ 2개

- ④ 3개
- 46. 아래와 같은 배열이 존재할 때, [1,2,2]의 원소와 [2,1,1]의 원소의 합을 구하면? (3점)

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

- ① 16
- ③ 19

- 4 21
- 47. 다음 중 벡터에 대한 설명으로 옳지 <u>않은</u> 것은 모두 몇 개인가? (3점)
 - 가. 한 개 이상의 원소로 구성된 자료구조이다.
 - 나. 하나의 벡터의 원소는 한 가지 형태(mode)만 가능하다.
 - 다. (TRUE, FALSE)과 같이 참(TRUE), 거짓(FALSE) 등의 값을 원소로 갖는 벡터를 진위형 벡터라고 한다.
 - ① 0개

② 1개

③ 2개

④ 3개

48. 다음 중 아래와 같이 길이가 같은 두 벡터 a, b가 열방향으로 병합되었을 때 생성되는 자료구조에 대한 설명으로 가장 옳지 않은 것은? (4점)

	1		8	
	0		9	
ามา	2	ull		
1 124	3	땐	10	
벡 터	4	터	11	
a	5	벡 터 b	12	
	6		13	
	7		14	

- ① 벡터 a, b가 병합되어 생성되는 자료구조는 행렬이다.
- ② 생성되는 자료구조의 [1, 2]원소와 [4, 2]원소의 곱은 24이다.
- ③ 생성되는 자료구조는 일반적으로 수치형, 문자형, 논리형 중한 가지 형태의 원소만 갖는다.
- ④ 생성되는 자료구조는 2차원의 데이터 구조이다.
- 49. 다음 중 수집하는 기업이나 기관의 목적에 따라 데이터베이스에 축적되어 이용되며 외부에 공개되지 않고 분석업무 및 의사결정 등에 사용되는 것은? (2점)
 - ① DBMS
- ② API
- ③ KOSIS
- ④ 내부데이터
- 50. 다음 중 웹문서 데이터 수집 방법에 관한 설명으로 가장 적절한 것은? (3점)
 - ① 크롤링 최적화란 검색로봇이 해당 웹문서의 내용을 가급적 간결하게 축소할 수 있도록 하는 것을 의미한다.
 - ② 인덱싱 최적화는 검색로봇에 의해 수집된 문서가 원하는 검색어에 대해 최대한 상위에 노출되도록 하는 것을 의미한다.
 - ③ 인덱싱 최적화는 사용자가 지정하거나 필요한 정보만을 추 출하여 가공하고 저장하는 것을 의미한다.
 - ④ 크롤링 최적화란 운영체제(Operating System; OS)나 프로그 래밍 언어가 제공하는 기능을 제어할 수 있도록 만들어 놓는 과정이다.
- 51. 다음 데이터베이스와 관련된 설명 중 옳은 것은 모두 몇 개인가? (4점)
 - 가. 데이터의 내용이 중복되지 않은 가운데 다양한 접근 방식 및 검색과 갱신의 효율성을 확보할 수 있다.
 - 나. 여러 사람이 데이터에 동시 접근하여 자료를 공유한다는 장점이 존재하지만 이로 인해 최신 데이터로의 갱신에 제 약이 따른다.
 - 다. 접근을 효율적으로 통제, 관리하여 보안을 유지할 수 있다. 라. 보조기억장치 활용이 최대한 배제된다.
 - ① 0개

② 1개

③ 2개

- ④ 3개
- 52. 다음 중 좋은 품질의 데이터에 관한 설명으로 가장 옳지 <u>않은</u> 것은? (3점)
 - ① 업계 표준을 준수
 - ② 완전하고 일관성 있어야 함
 - ③ 적시성이 담보되어야 함
 - ④ 다양한 형태로 입수되어야 함
- 53. 다음 중 의사결정 지원 시스템과 가장 거리가 먼 것은? (3점)
 - ① 데이터 기반의 의사결정을 데이터베이스와의 연관성을 고려하여 지원한다.
 - ② 분석모형과 데이터를 제공함으로써 의사결정과정이 더욱 효과적으로 이루어질 수 있도록 돕는다.
 - ③ 킨(Keen)과 모턴(Morton)이 제안한 바 있다.
 - ④ 단순하게 데이터를 수집, 저장, 분배하기 위한 시스템이다.

- 54. 다음 중 의사결정 지원 시스템의 구성요소로 옳게 묶인 것은? (3점)
 - ① 데이터베이스 시스템, 프레젠테이션, API, 데이터 시각화
 - ② 데이터웨어하우스, 모델베이스 시스템, 사용자 인터페이스, API
 - ③ 데이터베이스 시스템, 모델베이스 시스템, 사용자 인터페이스, 사용자
 - ④ 데이터웨어하우스, 모델베이스 시스템, 사용자, 데이터 시각화
- 55. 다음 중 소프트웨어 시스템, 라이브러리, 서지 데이터베이스에서 코드의 라인들과 같은 비수치 정보, 인터넷의 네트워크 관계 등 큰 범위의 집합에 대한 시각적 표현방법의 연구영역을 의미하는 것은? (2점)
 - ① 정보 탐색
 - ② 정보 추출
 - ③ 정보 시각화
 - ④ 인포그래픽
- 56. 다음 중 보기의 방법들과 가장 관계 깊은 시각화 영역은? (3점)

파이차트, 도우넛차트, 트리맵, 누적연속그래프, 히스토그램

- ① 시간시각화
- ② 분포시각화
- ③ 관계시각화
- ④ 비교시각화
- 57. 다음 중 데이터마이닝에서 실제 출력변수가 존재하여 입출력 변수 간의 관계를 근사시키는 모형 또는 규칙을 학습하는 기법 을 의미하는 것은? (2점)
 - ① 출력학습
 - ② 자율학습
 - ③ 근사학습
 - ④ 감독학습
- 58. 다음 중 데이터마이닝의 특징에 해당하는 것으로 묶인 것은? (3점)

가. 업무 활용성

나. 경험적 방법

다. 추론(inference)의 목적

라. 일반화(generalization)

- ① 가, 나
- ② 가, 나, 라
- ③ 가, 다, 라
- ④ 가, 나, 다, 라
- 59. 다음 중 아래의 설명과 가장 관련 깊은 데이터과학자의 기술적 필요역량을 일컫는 말은? (3점)

빅데이터에 대한 이론적 지식, 분석기술에 대한 숙련, IT 전 문성

- ① 통찰력(Insight)
- 2 Hard Skill
- ③ 스토리텔링 및 커뮤니케이션
- ④ 컨설팅 능력
- 60. 다음 중 데이터과학자가 필요로 하는 인문학적 소양과 가장 거리가 <u>면</u> 것은? (3점)
 - ① 데이터마이닝을 통한 표준화된 분석
 - ② 도메인지식(domain knowledge)을 바탕으로 한 새로운 관점
 - ③ 직관과 전략, 경영 프레임워크, 경영, 사회적 가치의 이해
 - ④ 가설 수립 및 검증에 필수적인 비판적 시각