인공지능

1. 다음 설명에 해당하는 기계 학습의 유형은?

> 어떤 환경 안에서 정의된 에이전트가 현재의 상태를 인식하여, 선택 가능한 행동들 중 보상을 최대화하는 행동 혹은 행동 순서를 선택하는 방법

① 강화 학습(reinforcement learning)
② 지도 학습(supervised learning)
③ 비지도 학습(unsupervised learning)
④ 앙상블 학습(ensemble learning)

2. 이진 분류기의 성능평가에 대한 설명으로 옳은 것은?
① 분류기가 양성으로 판정한 것들 중에서 실제 양성인 것의 비율을 재현율(recall)이라고 한다.
② 실제 음성인 것들 중에서 분류기가 음성으로 판정한 것의 비율을 특이도(specificity)라고 한다.
③ 실제 양성인 것들 중에서 분류기가 양성으로 판정한 것의 비율을 위양성율(false positive rate)이라고 한다.
④ 분류기는 민감도(sensitivity) 값이 작을수록 좋고 위양성율 값이 클수록 좋다.

3. 운전면허 시험 성적에서 합격 6명, 불합격 4명으로 구성된 데이터 세트의 지니 계수(Gini index) 값은?
① 0.25
② 0.48
③ 0.52
④ 1

4. 선형 회귀에서 평균제곱오차(MSE)를 최소화하는 계수를 찾는 방법은?
① 주성분 분석법
② 선형 판별 분석법
③ 최소우도법
④ 최소제곱법

5. 시간성을 갖는 데이터에 유용하고 문맥 의존성을 효율적으로 처리할 수 있는 신경망은?
① CNN
② KNN
③ MLP
④ RNN

6. 탐색 방법에 대한 설명으로 옳지 않은 것은?
① 맹목적 탐색은 문제의 상태 공간 정보를 이용한다.
② 깊이 우선 탐색은 너비 우선 탐색에 비해 메모리에 대한 부담이 적지만 최단 경로를 찾는다는 보장이 없다.
③ 너비 우선 탐색은 목표 상태를 찾을 때까지 생성된 모든 노드를 메모리에서 관리하기 때문에 메모리 비용이 크다.
④ 반복적 깊이 심화 탐색은 탐색 깊이 한계를 0부터 점차 증가시키면서 깊이 우선으로 탐색하는 방법이고 최단 경로 찾는 것을 보장한다.

7. 추천 시스템(recommender system)의 유형 중 협업 기반 추천과 내용 기반 추천에 대한 설명으로 옳지 않은 것은?
① 내용 기반 추천은 메모리 기반 방법과 모델 기반 방법이 있고 콜드 스타트(cold start) 문제가 발생한다.
② 아이템 기반 추천은 이전에 구매한 아이템을 기반으로 그 상품과 유사한 다른 상품을 추천하는 방법이다.
③ 사용자 기반 추천은 나와 비슷한 성향을 지닌 사용자의 데이터를 기반으로 상품을 추천하는 방법이다.
④ 추천 시스템은 콘텐츠의 내용에 기반하거나 사람들의 성향 정보를 취득한 후 개인화된 맞춤 항목을 추천한다.

8. 규칙 기반 시스템에 대한 설명으로 옳지 않은 것은?
① 경합 해소(conflict resolution)는 여러 규칙을 동시에 실행할 경우 모순이 발생할 수 있기 때문에 실행 가능한 규칙들 중에서 한 개를 선택하는 것을 의미한다.
② 전향 추론(forward chaining)은 규칙의 조건부를 만족시키는 사실이 있을 때 규칙의 결론부를 실행하거나 처리한다.
③ 후향 추론(backward chaining)은 특정 사실을 확인하기 위해 해당 사실을 결론부에 포함하는 규칙의 조건부가 만족하는지 확인해 가는 방식으로 추론한다.
④ 추론 엔진(inference engine)은 경합 해소 → 패턴 매칭(pattern matching) → 규칙 실행(rule execution)의 과정을 반복하여 추론한다.

9. A^* 알고리즘에 대한 설명으로 옳지 않은 것은?
① A^* 알고리즘의 평가함수는 $f(n) = g(n) + h(n)$으로 표현한다.
② 휴리스틱 함수 $\hat{h}(n)$이 실제 남은 비용을 과대평가하지 않으면 허용성(admissibility)을 만족한다.
③ 시작 노드에서 노드 n까지의 실제 비용 $g(n)$은 탐색이 진행됨에 따라 감소할 수 있다.
④ 휴리스틱 함수 $\hat{h}(n)$이 허용성을 만족하면 A^* 알고리즘은 항상 최적해를 찾는다.

10. 이진 분류 문제에 대한 혼동행렬(confusion matrix)에서 각각 계산한 민감도, 정확도(accuracy), 위양성율 값을 모두 곱한 값은?

		예측값	
		양성	음성
실젯값	양성	40	10
	음성	15	35

① $\frac{3}{80}$ ② $\frac{32}{275}$
③ $\frac{3}{55}$ ④ $\frac{9}{50}$

11. 기계 학습에 대한 설명으로 옳지 않은 것은?
① 군집화와 밀도추정 문제는 일반적으로 비지도 학습 방식을 사용한다.
② 지도 학습의 학습 데이터는 입력과 기대 출력 정보의 쌍으로 구성된다.
③ 데이터 증강은 입력 데이터를 저차원 데이터로 압축하는 인코더 형태의 모델을 통해 학습한다.
④ 프로그램을 명시적으로 작성하지 않고 컴퓨터가 학습할 수 있는 능력을 갖추게 하는 기술이다.

12. 생성적 적대 신경망(generative adversarial networks)에서 생성자(generator)와 판별자(discriminator)에 대한 설명으로 옳은 것만을 모두 고르면?

ㄱ. 생성자와 판별자는 적대적인 관계에서 경쟁하면서 훈련한다.
ㄴ. 생성자는 진짜 데이터를 만들면서 판별자가 가짜라고 속을 때까지 학습한다.
ㄷ. 판별자는 훈련 데이터를 진짜 데이터로 판별하고 생성자가 만든 데이터는 가짜 데이터로 판별하도록 학습한다.

① ㄱ, ㄴ ② ㄱ, ㄷ
③ ㄴ, ㄷ ④ ㄱ, ㄴ, ㄷ

13. 역학조사에서 흡연 여부와 암 유병 여부를 조사하였더니 다음 표와 같이 계산되었다. 흡연하는 사람일 사건을 A로 나타내고 암환자일 사건을 B로 나타내었을 때, 결합 확률 P(A∩B)의 값은?

	비흡연	흡연	합계(명)
정상인	35	15	50
암환자	7	3	10
합계(명)	42	18	60

① 0.05
② 0.15
③ 0.25
④ 0.33

14. 빅데이터 분석을 위한 데이터 마이닝에 대한 설명으로 옳지 않은 것은?
① 정제되지 않은 대용량의 데이터로부터 의미 있는 상관관계, 패턴, 추세 등을 발견한다.
② 트랜잭션에서 항목 간의 불순도를 표현하기 위해 최소지지도와 최소신뢰도를 사용한다.
③ 저장된 데이터를 정제(cleaning), 통합(integration)하여 잡음과 불일치를 제거하고 데이터 웨어하우스에 저장한다.
④ 대규모 데이터에서 암묵적인, 이전에 알려지지 않은, 잠재적으로 유용할 것 같은 정보나 지식을 추출하는 체계적인 과정이다.

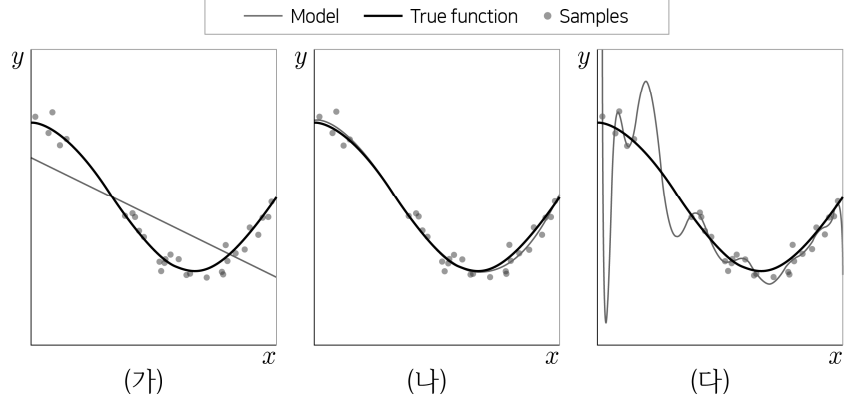
15. 다음 설명의 (가)에 들어갈 학습 방법은?

(가)은 학습하는 방법을 학습하는 것을 의미한다. 특정 작업을 인간과 같이 빠르게 학습하는 능력을 만들기 위해 동일한 분포를 이루는 여러 작업에 대한 학습 방법을 학습한 뒤, 유사한 작업을 적은 데이터로 빠르고 효과적으로 학습할 수 있게 한다.

① 연합 학습 ② 메타 학습
③ 강화 학습 ④ 표현 학습

16. 결정 트리 학습 알고리즘에 대한 설명으로 옳지 않은 것은?
① C4.5, C5.0, CHAID, CART 등의 알고리즘이 있다.
② 엔트로피는 어떤 집단에 각 부류의 데이터들이 균등한 비율로 섞여 있을수록 작아지는 특성이 있다.
③ ID3 알고리즘은 범주형 속성값을 갖는 학습 데이터로부터 엔트로피 개념을 사용하여 결정 트리를 만든다.
④ 정보 이득(information gain)은 특정한 속성이 원하는 분류 방식에 부합하게 데이터를 나누는지를 측정할 수 있는 척도이다.

17. 다음 그래프에 대한 설명으로 옳지 않은 것은? (단, x는 특징, y는 레이블이다)



① (가)는 모델의 표현력이 부족해 훈련 오차와 검증 오차가 모두 높게 유지되는 상황을 나타낸다.
② (나)는 훈련 오차와 검증 오차가 모두 낮고 서로 차이가 작아, 모델이 데이터에 일반화되고 있음을 보여 준다.
③ (다)는 훈련 오차가 낮고 검증 오차는 높게 유지되므로, 성능을 향상하기 위해서 모델 복잡도를 더 높여야 한다.
④ 기계 학습에서 과적합은 학습 데이터에서 성능이 뛰어나지만 일반화에 대한 성능을 보장할 수 없다.

18. 딥러닝 모델의 출력 노드가 3개(A, B, C)로 구성되어 있고, 각 출력 노드의 값이 A = 3.0, B = 2.0, C = 1.0이라고 하면 활성화 함수로 소프트맥스(softmax) 함수를 사용하였을 때 출력 노드 B를 변환한 근삿값은? (단, 자연 상수 e = 3.0으로 가정한다)

① 0.23
② 0.33
③ 0.67
④ 0.83

19. 워드 임베딩(word embedding)에 대한 설명으로 옳지 않은 것은?

① N-gram은 문자열에서 N개의 연속된 요소이다.
② 원-핫 인코딩(one-hot encoding)은 단어 개수가 많아질수록 벡터를 저장하기 위한 공간이 계속해서 늘어난다는 단점이 있다.
③ TF-IDF는 단어마다 가중치를 부여하여 단어를 벡터로 변환하는 방법으로 TF는 특정 단어가 등장한 문서의 수를 의미한다.
④ 워드투벡터(Word2Vec)는 비슷한 컨텍스트에 등장하는 단어들이 유사한 의미를 지닌다는 이론에 기반을 두고 단어를 벡터로 표현해 주는 기법이다.

20. 컨볼루션 신경망(CNN)에 대한 설명으로 옳지 않은 것은?

① 학습 알고리즘은 기본적으로 경사 하강법을 사용하는 역전파 알고리즘이다.
② 컨볼루션 신경망을 분류 문제에 사용하는 경우, 마지막 층은 소프트맥스층으로 구성된다.
③ 컨볼루션 신경망에서 활성화 함수로 시그모이드를 사용하면 ReLU 함수에 비해 계산시간이 줄어든다.
④ 풀링층(pooling layer)은 합성곱층의 출력 데이터를 입력으로 받아서 특징맵의 크기를 줄이거나 특징을 추출하는 용도로 사용된다.

21. 로지스틱 회귀에 대한 설명으로 옳은 것만을 모두 고르면?

ㄱ. 이진 분류에 사용하지 않는 알고리즘이다.
ㄴ. 예측값을 변환할 때 시그모이드 함수를 사용한다.
ㄷ. 종속변수에 로그를 적용하는 로짓(logit) 변환을 사용한다.
ㄹ. 최적해를 구하기 위해서 최소제곱법을 사용한다.

① ㄱ, ㄴ　② ㄴ, ㄷ
③ ㄴ, ㄹ　④ ㄷ, ㄹ

22. 퍼지 집합론에서 영역 X의 특정 원소 x가 있을 때, 퍼지 소속 함수(membership function) $\mu_A(x)$는 x가 집합 A에 속하는 정도이다. 퍼지 집합 연산 중 합집합으로 옳은 것은?

① $\mu_A(x) + \mu_B(x)$
② $\mu_A(x) \times \mu_B(x)$
③ $\min(\mu_A(x), \mu_B(x))$
④ $\max(\mu_A(x), \mu_B(x))$

23. 이진 분류 모델의 실젯값과 예측값이 다음과 같이 계산되었을 때, F1 점수(score)의 근삿값은? (단, 1을 positive로 가정한다)

실젯값	1	0	1	0	1	0	1	0
예측값	1	0	0	0	1	1	0	0

① 0.33
② 0.5
③ 0.57
④ 0.75

24. 품사 태깅(POS tagging)에 대한 설명으로 옳지 않은 것은?

① 구문에 따라 주어진 문장에서 단어들의 역할을 파악하여 계층적인 트리 구조로 변환한다.
② 문장의 각 단어에 품사의 범주 및 문법적 기능에 대한 태그를 붙인다.
③ 컴퓨터가 읽을 수 있는 형태로 텍스트를 모아놓은 말뭉치(corpus)를 사용한다.
④ 규칙 기반 품사 태깅은 언어 정보를 이용하여 수작업으로 만든 규칙들을 사용한다.

25. 연관 규칙 마이닝(association rule mining)에 대한 설명으로 옳은 것만을 모두 고르면?

ㄱ. 향상도가 1.0보다 크면, 항목집합 사이의 연관성 정도와 유용성이 높다는 것을 의미한다.
ㄴ. 신뢰도는 전체 거래 건수 대비 항목집합 X와 Y를 모두 포함하는 거래 건수의 비율이다.
ㄷ. 지지도는 항목집합 X를 포함하는 거래 중에서 Y도 포함하는 거래 건수의 비율이다.
ㄹ. 트랜잭션에서 어프라이어리(Apriori) 알고리즘으로 빈발 항목집합들을 찾은 다음, 이들로부터 연관 규칙을 만든다.

① ㄱ
② ㄱ, ㄹ
③ ㄴ, ㄷ
④ ㄴ, ㄷ, ㄹ