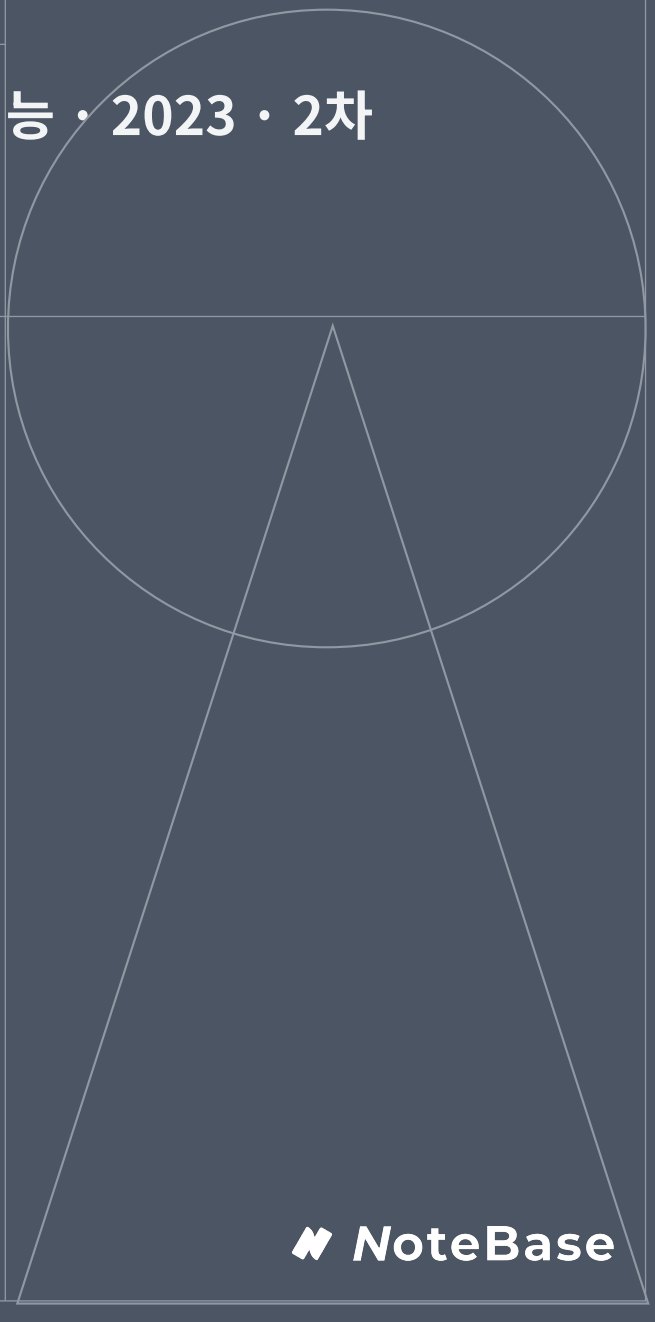
기출OX 해설집

공무원/국가직/7급 · 인공지능 · 2023 · 2차



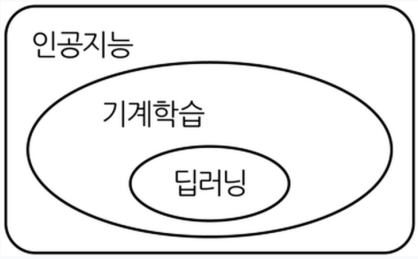
NoteBase

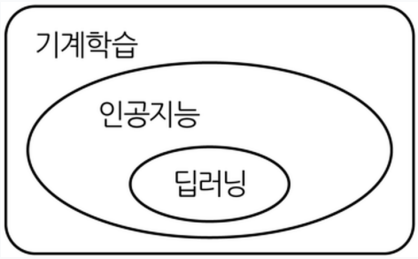
기출OX 해설집

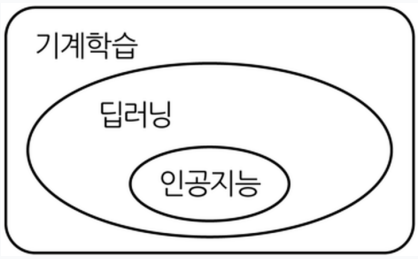
공무원/국가직/7급 · 인공지능 · 2023 · 2차

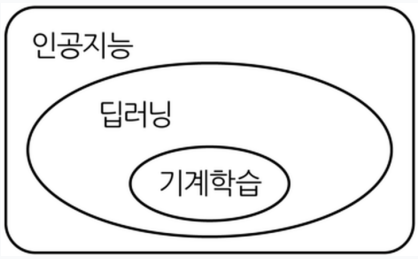
공무원/국가직/7급 · 인공지능 · 2023 · 2차

문 1. 기술 간 포함관계로 옳은 것은?

① 

② 

③ 

④ 

정답 1번

출제 포인트

이 문제는 인공지능, 기계학습, 딥러닝의 포함관계를 도식으로 구별할 수 있는지를 묻는다.

NoteBase

선지별 해설

1번: O. 인공지능이 가장 큰 범주이고 그 안에 기계학습이 있으며, 딥러닝은 기계학습의 하위 분야이므로 도식의 포함관계가 정확하다.
2번: X. "기계학습이 가장 바깥 집합"이라는 배치가 틀리다. 올바르게는 인공지능이 가장 바깥 집합이고 그 안에 기계학습, 그 안에 딥러닝이 들어간다.
3번: X. "딥러닝이 기계학습을 포함하고 인공지능이 가장 안쪽 집합"이라는 배치가 틀리다. 올바르게는 인공지능이 가장 넓고, 기계학습이 그 부분집합이며, 딥러닝은 기계학습의 부분집합이다.
4번: X. "딥러닝 안에 기계학습이 들어간다"는 포함방향이 뒤집혀 있다. 올바르게는 기계학습이 딥러닝을 포함하며, 둘 다 인공지능에 포함된다.

오답 포인트

기계학습과 딥러닝의 상하관계를 뒤집거나, 인공지능과 기계학습을 같은 범주로 보는 경우에 오답이 나온다. 이 문항은 집합의 크기 순서를 묻는 문제이므로 이름이 익숙한 기술일수록 더 큰 범주라는 식으로 판단하면 틀린다.

다시 나오면 이렇게 판단

먼저 가장 넓은 개념이 무엇인지부터 본다. 인공지능은 지능적 작업을 수행하게 하는 전체 분야이고, 기계학습은 그중 데이터로 학습하는 방법이며, 딥러닝은 기계학습 안의 신경망 기반 방법이므로 인공지능 ⊃ 기계학습 ⊃ 딥러닝의 순서로 판단한다.

암기 포인트

포함관계는 인공지능 ⊃ 기계학습 ⊃ 딥러닝이다.

NoteBase

문 2. 자연어처리 도구로 옳지 않은 것은?

① BERT

② GPT

③ Transformer

④ OpenCV

정답 4번

출제 포인트

이 문제는 자연어처리에서 쓰이는 대표 모델·구조와 컴퓨터비전 라이브러리를 구분하는 문제이다.

핵심 개념

BERT는 문맥을 양방향으로 학습하는 자연어처리 사전학습 언어모델이다. 키워드는 양방향 인코더이다.
GPT는 다음 토큰 생성 중심의 자연어처리 생성형 사전학습 언어모델이다. 키워드는 생성형 디코더이다.
Transformer는 어텐션 메커니즘을 기반으로 한 시퀀스 처리 신경망 구조로, 현대 자연어처리 모델의 기반이다. 키워드는 어텐션 기반 구조이다.
OpenCV는 이미지와 영상 처리용 컴퓨터비전 라이브러리이다. 키워드는 비전 처리이다.

선지별 해설

① O: BERT는 대표적인 자연어처리 언어모델이므로 자연어처리 도구로 볼 수 있다.
② O: GPT는 텍스트 생성과 질의응답 등에 쓰이는 대표적 자연어처리 모델이므로 자연어처리 도구에 해당한다.
③ O: Transformer는 자연어처리 모델 다수의 기반이 되는 핵심 구조이므로 시험 문맥에서는 자연어처리 도구 범주에 포함된다.
④ X: 틀린 부분은 OpenCV를 자연어처리 도구로 본 점이다. OpenCV는 주로 이미지, 영상, 객체 검출 등 컴퓨터비전 작업에 쓰인다. 올바르게는 OpenCV는 컴퓨터비전 라이브러리이다.

오답 포인트

Transformer가 모델 이름인지 구조 이름인지 헷갈리면 ③을 배제하기 쉽다. 이 문항은 엄밀한 제품 분류보다 인공지능 분야별 대표 도구·모델을 구분하는 방식이므로, 비전 전용에 가까운 OpenCV를 골라야 한다.

다시 나오면 이렇게 판단

선지를 보고 먼저 텍스트 처리 계열인지, 영상 처리 계열인지 분류하면 된다. 언어모델·어텐션 기반 구조는 자연어처리 쪽이고, 이미지·영상 처리 라이브러리는 컴퓨터비전 쪽이다.

NoteBase

암기 포인트

BERT·GPT·Transformer는 NLP, OpenCV는 CV이다.

NoteBase

문 3. 약한 인공지능(weak AI) 및 강한 인공지능(strong AI)에 대한 설명으로 옳지 않은 것은?

① 약한 인공지능은 특정 문제를 해결할 수 있는 수준의 인공지능을 말한다.

② 강한 인공지능은 인간의 모든 지능적 요소를 가진 인공지능을 말한다.

③ 알파고는 강한 인공지능에 속한다.

④ 강한 인공지능은 앨런 튜링이 제시한 튜링 테스트를 통과할 수 있다.

정답 3번

출제 포인트

이 문제는 약한 인공지능과 강한 인공지능의 정의를 구별하고 실제 시스템이 어느 범주에 속하는지를 판단하는 문제이다.

핵심 개념

약한 인공지능은 특정 과업에 특화되어 그 문제를 효율적으로 해결하도록 설계된 인공지능이다. 강한 인공지능은 인간처럼 범용적으로 이해, 추론, 학습할 수 있는 일반 지능을 가진 인공지능을 뜻한다. 알파고는 바둑이라는 한정된 과업에 최적화된 시스템이므로 약한 인공지능에 속한다.

선지별 해설

1번 O: 약한 인공지능은 번역, 추천, 게임, 음성 인식처럼 특정 문제를 해결하는 데 특화된 인공지능을 말한다. 범용 지능이 아니라 과업 특화형이라는 점에서 옳다.
2번 O: 강한 인공지능은 인간의 여러 지적 능력을 폭넓게 수행할 수 있는 일반 인공지능을 가리킨다. 인간의 모든 지능적 요소를 가진 인공지능이라는 설명과 맞는다.
3번 X: 틀린 부분은 "알파고는 강한 인공지능에 속한다"는 표현이다. 실제로는 알파고는 바둑 대국이라는 특정 영역에 특화된 시스템이므로 약한 인공지능에 속한다.
4번 O: 튜링 테스트는 기계가 인간처럼 지능적 대화를 수행하는지를 평가하는 기준이다. 강한 인공지능은 인간 수준의 일반 지능을 전제하므로 이러한 테스트를 통과할 수 있는 인공지능으로 설명된다.

오답 포인트

실제 성능이 매우 뛰어난 시스템을 보면 곧바로 강한 인공지능으로 연결하기 쉽다. 그러나 출제자는 성능의 높고 낮음이 아니라 특정 과업에만 특화되었는지, 인간처럼 범용적으로 사고할 수 있는지가 구분 기준임을 묻고 있다.

다시 나오면 이렇게 판단

먼저 해당 인공지능이 한 가지 과업에 특화된 시스템인지, 여러 지적 과업을 인간처럼 일반적으로 수행하는 지능인지 구분한다. 특정 영역 최적화이면 약한 인공지능, 범용적 인간 수준 지능이면 강한 인공지능으로 판단한다.

NoteBase

암기 포인트

약한 인공지능은 특정 과업, 강한 인공지능은 범용 지능, 알파고는 약한 인공지능이다.

NoteBase

문 4. 규칙 기반 전문가 시스템을 구성하는 요소가 아닌 것은?

① 규칙 집합으로 표현된 지식베이스

② 사용자와의 상호 작용을 지원하는 사용자 인터페이스

③ 규칙과 사실을 연결하는 추론 엔진

④ 규칙 생성을 위한 자연어 처리 라이브러리

정답 4번

출제 포인트

이 문제는 규칙 기반 전문가 시스템의 전형적 구성 요소와 구현에 활용될 수 있는 주변 기술을 구분하는 문제이다.

핵심 개념

규칙 기반 전문가 시스템은 규칙 형태의 지식을 저장한 지식베이스와, 그 규칙을 사실에 적용하는 추론 엔진, 그리고 사용자와 상호작용하는 인터페이스를 중심으로 구성된다.
자연어 처리 라이브러리는 입력 처리나 규칙 생성 보조에 활용될 수는 있지만, 고전적 규칙 기반 전문가 시스템의 필수 구성 요소로 보지는 않는다.

선지별 해설

① O
규칙 집합으로 표현된 지식베이스는 전문가 시스템의 핵심 구성 요소이다. 전문가의 지식을 IF-THEN 형태의 규칙으로 저장해 두고 추론에 사용한다.

② O
사용자 인터페이스는 사용자가 사실을 입력하고 결과를 확인하도록 돕는 구성 요소이다. 전문가 시스템은 사용자와의 질의응답이 필요하므로 인터페이스가 포함된다.

③ O
추론 엔진은 규칙과 사실을 연결해 결론을 도출하는 장치이다. 전방 추론이나 후방 추론을 수행하는 부분이므로 핵심 구성 요소에 해당한다.

④ X
틀린 부분은 "규칙 생성을 위한 자연어 처리 라이브러리"이다. 자연어 처리 기술은 규칙 작성이나 입력 해석을 보조할 수는 있지만, 규칙 기반 전문가 시스템의 전형적 필수 구성 요소는 아니다. 올바르게는 지식베이스, 추론 엔진, 사용자 인터페이스, 설명 기능 등이 핵심 구성 요소이다.

NoteBase

오답 포인트

최근 인공지능 기술과 고전적 전문가 시스템 구조를 섞어서 보면 ④가 그럴듯해 보인다. 그러나 이 문항은 시스템을 운영하는 기본 아키텍처를 묻는 것이므로, 구현 편의 기술인지 핵심 구성 요소인지를 구분해야 한다.

다시 나오면 이렇게 판단

전문가 시스템 구성 요소 문제는 지식 저장, 추론 수행, 사용자 상호작용처럼 시스템이 직접 동작하는 축을 먼저 본다. 자연어 처리, 데이터 수집 도구, 외부 라이브러리처럼 구축이나 보조에 쓰이는 기술은 기본 구성 요소와 구분한다.

암기 포인트

전문가 시스템은 지식베이스-추론 엔진-사용자 인터페이스가 기본 축이고, 자연어 처리 라이브러리는 보조 기술이다.

NoteBase

문 5. 트리 탐색의 결과로 전역 최적값이 보장되는 기법은?

① 너비 우선 탐색(breadth-first search)

② 허용성(admissibility)을 갖지 않는 A* 알고리즘

③ 언덕 오르기 방법(hill climbing method)

④ 탐색 깊이 제한이 있는 깊이 우선 탐색(depth-first search)

정답 1번

출제 포인트

이 문제는 대표적 탐색 기법들 중 전역 최적해를 보장하는 조건이 무엇인지 구분하는 문제이다.

핵심 개념

너비 우선 탐색은 깊이가 가장 낮은 노드부터 순서대로 확장하는 탐색이며, 각 단계 비용이 동일한 트리 탐색에서는 가장 먼저 찾은 해가 최단 경로가 되어 최적해를 보장한다.
A* 알고리즘은 휴리스틱이 허용적일 때 최적해를 보장하며, 허용성이 없으면 이 보장이 깨진다.
언덕 오르기 방법은 현재 상태보다 좋아 보이는 방향으로만 이동하는 국소 탐색이므로 지역 최적해에 멈출 수 있다.
깊이 제한이 있는 깊이 우선 탐색은 탐색 순서와 제한 깊이에 크게 좌우되므로 최적해 보장이 없다.

선지별 해설

① O: 너비 우선 탐색은 같은 깊이의 노드를 모두 살핀 뒤 다음 깊이로 내려간다. 따라서 단계 비용이 동일한 전형적 트리 탐색에서는 가장 얕은 해를 먼저 찾아 전역 최적해를 보장한다.
② X: 틀린 부분은 "허용성(admissibility)을 갖지 않는"이다. 허용성이 없는 휴리스틱은 실제 비용을 과대평가할 수 있어 A*의 최적성 보장이 사라진다. 올바르게는 "허용성을 갖는 A* 알고리즘은 최적해를 보장한다."이다.
③ X: 언덕 오르기 방법은 현재보다 좋아지는 방향만 선택하므로 지역 최적점이나 평탄면에서 멈출 수 있다. 틀린 부분은 전역 최적을 기대하는 해석이다. 올바르게는 "언덕 오르기 방법은 전역 최적해를 보장하지 않고 지역 최적해에 머물 수 있다."이다.
④ X: 깊이 제한이 있는 깊이 우선 탐색은 제한 깊이 밖의 더 좋은 해를 보지 못할 수 있고, 같은 깊이 안에서도 탐색 순서에 따라 먼저 만난 해를 택하게 된다. 틀린 부분은 최적성 보장을 기대하는 점이다. 올바르게는 "깊이 제한이 있는 깊이 우선 탐색은 전역 최적해를 보장하지 않는다."이다.

NoteBase

오답 포인트

"A*는 항상 최적"이라고 외워 두고 ②를 고르기 쉽지만, 그 보장은 허용적 휴리스틱이라는 조건이 붙는다. 또한 ③처럼 값이 계속 좋아지는 방향으로 가는 탐색은 효율적으로 보이지만 지역 최적에 갇히기 쉬워 전역 최적 보장과는 다르다.

다시 나오면 이렇게 판단

전역 최적 보장 여부를 묻는 문제에서는 먼저 완전탐색 계열인지, 휴리스틱의 최적성 조건이 붙는지, 지역 탐색인지부터 구분한다. 지역 탐색이나 깊이 제한 탐색은 보통 최적성 보장이 없고, A*는 허용성 조건을 확인해야 하며, 너비 우선 탐색은 동일 비용 조건에서 최적해를 보장한다.

암기 포인트

BFS는 같은 비용이면 최단 경로를 보장하고, A*는 허용적일 때만 최적이며, 언덕 오르기와 깊이 제한 DFS는 최적 보장이 없다.

NoteBase

문 6. 생성 모델에 대한 설명으로 옳지 않은 것은?

① 생성 모델을 통해 특정 연령, 성별, 안경 유형과 같은 원하는 속성으로 유명인의 얼굴을 바꿀 수 있다.

② 자기 부호화기(auto encoder)로 영상을 생성할 경우, 인코더와 디코더 설계 시 합성곱(convolution)층을 사용할 수 있다.

③ 생성적 적대 신경망(또는 적대적 생성 신경망, GAN, Generative Adversarial Networks)이 성공적으로 학습을 완료하면 분별망은 생성망이 만든 영상을 식별할 수 있다.

④ 변이형 자기 부호화기(variational auto encoder)는 학습 데이터의 분포를 따르는 새로운 영상을 생성할 수 있다.

정답 3번

출제 포인트

이 문제는 생성 모델에서 자기부호화기, GAN, VAE의 역할과 학습 결과를 정확히 구분할 수 있는지를 묻는 문제이다.

핵심 개념

자기부호화기는 입력을 잠재표현으로 압축한 뒤 다시 복원하는 구조이며, 영상 처리에서는 합성곱 기반 인코더와 디코더를 자주 사용한다.
GAN은 생성망과 분별망이 경쟁적으로 학습하는 구조이며, 학습이 잘 되면 생성망이 만든 데이터가 실제 데이터와 비슷해진다.
변이형 자기부호화기는 잠재공간을 확률분포로 학습하고 그 분포에서 샘플링해 새로운 데이터를 생성한다.

선지별 해설

1번: O. 생성 모델은 잠재공간 조작이나 조건 입력을 통해 나이, 성별, 안경과 같은 속성을 바꾸는 얼굴 편집에 활용될 수 있다. 유명인의 얼굴도 데이터와 조건이 주어지면 원하는 속성 방향으로 변형할 수 있다.
2번: O. 영상용 자기부호화기는 공간적 구조를 보존해야 하므로 인코더와 디코더에 합성곱층을 사용하는 합성곱 자기부호화기 형태를 사용할 수 있다.
3번: X. 틀린 부분은 "분별망은 생성망이 만든 영상을 식별할 수 있다"이다. GAN이 성공적으로 학습되면 분별망은 생성 영상과 실제 영상을 뚜렷하게 구분하지 못한다. 올바르게는 생성망이 실제와 유사한 영상을 만들어 분별망이 쉽게 식별하지 못하는 상태에 가까워진다.
4번: O. 변이형 자기부호화기는 잠재변수를 확률분포로 모델링하므로 학습 데이터 분포를 따르는 새로운 영상을 샘플링해 생성할 수 있다.

NoteBase

오답 포인트

3번은 분별망의 역할과 학습 완료 후 상태를 혼동하면 틀리기 쉽다. 분별망은 학습 과정에서는 가짜를 구별하려고 하지만, 성공적인 GAN의 균형점에서는 생성 결과가 좋아져 분별망이 쉽게 구별하지 못한다는 점이 핵심이다.

다시 나오면 이렇게 판단

생성 모델 문제는 먼저 모델의 목적을 대응시키면 풀린다. 자기부호화기는 복원 구조인지, GAN은 생성망과 분별망의 경쟁 학습인지, VAE는 확률적 잠재공간에서 샘플링해 새 데이터를 만드는지 순서대로 확인하면 된다.

암기 포인트

자기부호화기는 복원, GAN은 경쟁 학습, VAE는 분포 샘플링 생성이다.

NoteBase

문 7. 분류(class)가 2개인 혼동행렬(confusion matrix)에서 사용하는 성능 측정 기준에 대한 설명으로 옳은 것은?

		실젯값	
		긍정	부정
예측값	긍정	TP	FP
	부정	FN	TN

① 특이도(specificity) = $\frac{TN}{FN+TN}$

② 민감도(sensitivity) = $\frac{TP}{TP+FP}$

③ 정밀도(precision) = $\frac{TN}{TN+FP}$

④ 재현율(recall) = $\frac{TP}{TP+FN}$

정답 4번

출제 포인트

이 문제는 이진 분류 혼동행렬에서 특이도, 민감도, 정밀도, 재현율의 정의를 정확히 구분하는지를 묻는 문제이다.

핵심 개념

특이도는 실제 부정 중 부정으로 맞힌 비율로 $TN/(TN+FP)$이다.
민감도와 재현율은 같은 개념으로 실제 긍정 중 긍정으로 맞힌 비율이며 $TP/(TP+FN)$이다.
정밀도는 긍정으로 예측한 것 중 실제 긍정의 비율로 $TP/(TP+FP)$이다.
혼동 포인트는 분모가 무엇이냐는 점이다. 실제값 기준이면 민감도·특이도이고, 예측값 기준이면 정밀도이다.

NoteBase

선지별 해설

① X
특이도는 실제 부정 집합을 기준으로 계산하므로 분모가 $TN + FP$이어야 한다.
틀린 표현은 $FN + TN$이다.
올바르게는 특이도 = $TN/(TN + FP)$이다.

② X
민감도는 실제 긍정 중에서 얼마나 찾아냈는지를 보는 지표이므로 분모가 $TP + FN$이어야 한다.
틀린 표현은 $TP + FP$이다.
올바르게는 민감도 = $TP/(TP + FN)$이다.
해당 식 $TP/(TP + FP)$는 정밀도이다.

③ X
정밀도는 긍정으로 예측한 것의 정확성을 보므로 $TP/(TP + FP)$로 계산한다.
틀린 표현은 분자 TN과 분모 $TN + FP$의 조합이다.
올바르게는 정밀도 = $TP/(TP + FP)$이다.
제시된 식 $TN/(TN + FP)$는 특이도이다.

④ O
재현율은 실제 긍정 중에서 모델이 긍정으로 맞힌 비율이므로 $TP/(TP + FN)$이다.
재현율은 민감도와 같은 식을 사용한다.

오답 포인트

②와 ④는 민감도와 재현율이 같은 식이라는 점을 모르면 서로 다른 개념처럼 보이기 쉽다. ①과 ③도 TN이 들어간다고 모두 같은 지표가 아니며, 실제값 기준인지 예측값 기준인지 구분하지 않으면 분모를 뒤바꾸게 된다.

다시 나오면 이렇게 판단

먼저 분모가 실제값 기준인지 예측값 기준인지 본다. 실제 긍정 기준이면 민감도·재현율 $TP/(TP + FN)$, 실제 부정 기준이면 특이도 $TN/(TN + FP)$, 예측 긍정 기준이면 정밀도 $TP/(TP + FP)$로 판단한다.

암기 포인트

민감도=재현율은 실제 긍정을 얼마나 놓치지 않았는가, 정밀도는 예측한 긍정이 얼마나 맞았는가, 특이도는 실제 부정을 얼마나 잘 걸렀는가이다.

NoteBase

문 8. 날씨 상태의 변화를 나타내는 동질 마르코프 체인(homogeneous Markov chain) 전이 확률 행렬(transition probability matrix)이 다음과 같을 때, 오늘 날씨가 맑을 경우 내일 날씨가 흐리고 모레 날씨도 흐릴 확률은?

오늘 날씨 \ 내일 날씨	맑음	흐림
맑음	0.8	0.2
흐림	0.3	0.7

① 0.04

② 0.14

③ 0.49

④ 0.70

정답 2번

출제 포인트

이 문제는 동질 마르코프 체인에서 두 날 연속 사건의 결합확률을 전이확률의 곱으로 구하는 문제이다.

조건 정리

주어진 값: 오늘이 맑음일 때 내일 흐림일 확률은 0.2이고, 오늘이 흐림일 때 내일 흐림일 확률은 0.7이다.
구해야 할 값: 오늘이 맑음일 때 내일이 흐리고 모레도 흐릴 확률이다.
주의할 조건: 모레의 확률은 내일 상태가 이미 흐림으로 바뀐 뒤의 전이확률 0.7을 써야 한다.

사용할 공식

마르코프 성질에 따라 두 단계 결합확률은 단계별 전이확률의 곱으로 계산된다.
$P(X_1 = C, X_2 = C \mid X_0 = S) = P(X_1 = C \mid X_0 = S) \times P(X_2 = C \mid X_1 = C)$
여기서 S는 맑음, C는 흐림이다.
이 공식을 쓰는 이유는 내일 상태가 정해지면 모레 확률은 오늘이 아니라 내일 상태에만 의존하기 때문이다.

계산 과정

1. 오늘이 맑음이므로 첫 단계에서 맑음 → 흐림 확률 0.2를 잡는다.
2. 내일이 흐림이 된 뒤 모레도 흐릴 확률은 흐림 → 흐림 확률 0.7이다.
3. 따라서 전체 확률은 $0.2 \times 0.7 = 0.14$이다.

NoteBase

선지별 해설

① X: 0.04는 0.2 × 0.2처럼 두 단계 모두 맑음 → 흐림 확률 0.2를 반복 적용한 값이다. 틀린 표현은 모레 확률에도 0.2를 쓴 점이다. 올바르게는 내일이 흐림이 된 뒤에는 흐림 → 흐림 확률 0.7을 써서 0.14가 된다.
② O: 오늘이 맑음일 때 내일 흐림일 확률 0.2와, 내일이 흐림일 때 모레도 흐림일 확률 0.7을 곱하면 0.14가 된다.
③ X: 0.49는 0.7 × 0.7로 흐림 상태에서 시작한 것처럼 계산한 값이다. 틀린 표현은 처음 상태가 맑음인데도 흐림 → 흐림 확률 0.7을 첫 단계에 쓴 점이다. 올바르게는 첫 단계는 0.2, 둘째 단계는 0.7이어서 0.14가 된다.
④ X: 0.70은 흐림 → 흐림 한 단계 확률만 본 값이다. 틀린 표현은 내일 흐림이 될 첫 단계 0.2를 누락한 점이다. 올바르게는 두 단계를 모두 반영해 0.14가 된다.

오답 포인트

출제자는 모레 확률을 계산할 때도 처음 행의 0.2를 반복해서 쓰거나, 반대로 흐림 → 흐림 확률 0.7만 보고 첫 단계를 빼먹는 실수를 노린다. 마르코프 체인에서는 각 단계에서 현재 상태의 행을 다시 선택해야 한다.

다시 나오면 이렇게 판단

전이확률 문제는 시작 상태를 먼저 정하고, 한 단계 이동할 때마다 현재 상태가 어느 행으로 바뀌는지 추적한 뒤 필요한 전이확률을 순서대로 곱해 판단한다.

암기 포인트

두 단계 마르코프 확률은 첫 전이확률 × 다음 상태에서의 전이확률이다.

NoteBase

문 9. 선형 회귀에서 손실함수(loss function)가 다음과 같을 때, 경사하강법에 대한 내용으로 옳지 않은 것은? (단, 학습률은 0.05이고 W는 가중치(weight), b는 편향(bias)이다)

$$Loss(W, b) = \frac{1}{n}\sum_{i=1}^{n}((Wx_i + b) - y_i)^2$$

① $\frac{\partial Loss(W,b)}{\partial W} = \frac{2}{n}\sum_{i=1}^{n}((Wx_i + b) - y_i)$

② $W_{new} = W - 0.05 \times \frac{\partial Loss(W,b)}{\partial W}$

③ $b_{new} = b - 0.05 \times \frac{\partial Loss(W,b)}{\partial b}$

④ $b_{new} = b - 0.1 \times \frac{1}{n}\sum_{i=1}^{n}((Wx_i + b) - y_i)$

정답 1번

출제 포인트

평균제곱오차 손실함수를 W와 b로 편미분한 뒤 학습률 0.05를 적용한 경사하강법 업데이트식을 판단하는 문제이다.

조건 정리

주어진 식은 다음과 같다.
$Loss(W, b) = \frac{1}{n}\sum_{i=1}^{n}((Wx_i + b) - y_i)^2$
구해야 할 내용은 각 선지의 편미분식과 업데이트식의 타당성이다.
주의할 조건은 W로 미분할 때 연쇄법칙에 따라 x_i가 함께 곱해지고, b로 미분할 때는 내부식의 미분값이 1이 된다는 점이다.

사용할 공식

편미분과 경사하강법의 기본식은 다음과 같다.
$\frac{\partial}{\partial W}(Wx_i + b - y_i)^2 = 2(Wx_i + b - y_i)x_i$
$\frac{\partial}{\partial b}(Wx_i + b - y_i)^2 = 2(Wx_i + b - y_i)$
$W_{new} = W - \eta\frac{\partial Loss}{\partial W}, \quad b_{new} = b - \eta\frac{\partial Loss}{\partial b}$
이 공식을 쓰는 이유는 손실함수가 제곱오차 평균이므로 각 항에 연쇄법칙을 적용해 기울기를 구한 뒤 학습률 $\eta = 0.05$로 갱신해야 하기 때문이다.

NoteBase

계산 과정

1. W에 대한 편미분은 다음과 같다.
$\frac{\partial Loss(W,b)}{\partial W} = \frac{1}{n}\sum_{i=1}^{n} 2(Wx_i + b - y_i)x_i = \frac{2}{n}\sum_{i=1}^{n}(Wx_i + b - y_i)x_i$
2. b에 대한 편미분은 다음과 같다.
$\frac{\partial Loss(W,b)}{\partial b} = \frac{1}{n}\sum_{i=1}^{n} 2(Wx_i + b - y_i) = \frac{2}{n}\sum_{i=1}^{n}(Wx_i + b - y_i)$
3. 따라서 업데이트식은 다음과 같다.
$W_{new} = W - 0.05\frac{\partial Loss(W,b)}{\partial W}$
$b_{new} = b - 0.05\frac{\partial Loss(W,b)}{\partial b} = b - 0.05 \cdot \frac{2}{n}\sum_{i=1}^{n}(Wx_i + b - y_i) = b - 0.1 \cdot \frac{1}{n}\sum_{i=1}^{n}(Wx_i + b - y_i)$

선지별 해설

① X

W로 미분하면 내부식 $Wx_i + b - y_i$를 W에 대해 미분한 값 x_i가 반드시 곱해져야 한다. 틀린 표현은 $\frac{2}{n}\sum_{i=1}^{n}(Wx_i + b - y_i)$이고, 올바르게는 $\frac{2}{n}\sum_{i=1}^{n}(Wx_i + b - y_i)x_i$이다.

② O

경사하강법의 기본 업데이트식은 파라미터에서 학습률 곱하기 기울기를 빼는 형태이다. 따라서 $W_{new} = W - 0.05\frac{\partial Loss(W,b)}{\partial W}$는 맞는 식이다.

③ O

편향 b도 같은 방식으로 갱신하므로 $b_{new} = b - 0.05\frac{\partial Loss(W,b)}{\partial b}$는 맞는 식이다.

④ O

$\frac{\partial Loss(W,b)}{\partial b} = \frac{2}{n}\sum_{i=1}^{n}(Wx_i + b - y_i)$이므로 ③에 대입하면 $b_{new} = b - 0.05 \cdot \frac{2}{n}\sum_{i=1}^{n}(Wx_i + b - y_i)$가 되고, 이는 곧 $b_{new} = b - 0.1 \cdot \frac{1}{n}\sum_{i=1}^{n}(Wx_i + b - y_i)$와 같아서 맞는 식이다.

오답 포인트

가장 흔한 실수는 W 미분과 b 미분을 같은 형태로 적는 것이다. W 미분에는 x_i가 붙고 b 미분에는 붙지 않는다. 또 b의 기울기에 있는 계수 2와 학습률 0.05를 곱하면 0.1이 된다는 점을 놓치면 ④를 틀리게 본다.

다시 나오면 이렇게 판단

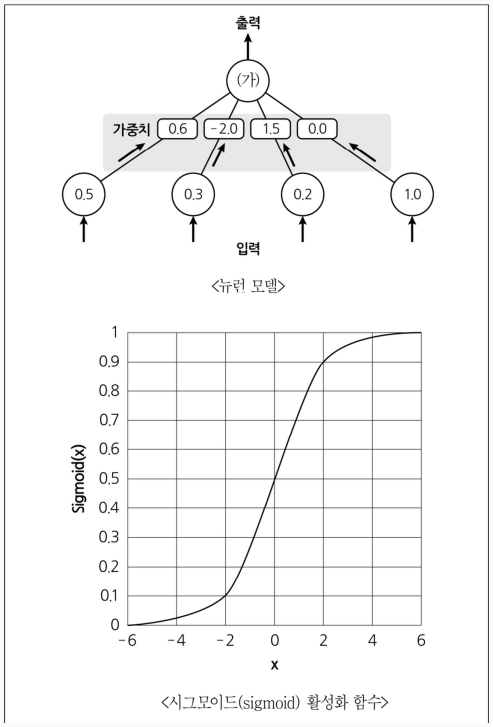
제곱오차 경사하강법 문제는 먼저 제곱항을 편미분해 내부식의 미분값이 무엇인지 확인한 뒤, 그 기울기에 학습률을 곱해 업데이트식으로 옮기면 된다. 즉 W에는 x_i가 붙는지, b에는 붙지 않는지만 먼저 점검하면 빠르게 판단할 수 있다.

암기 포인트

W 미분에는 x_i가 붙고 b 미분에는 붙지 않는다. 학습률 0.05와 미분계수 2가 만나면 0.1이 된다.

NoteBase

문 10. 다음과 같은 뉴런 모델에서 시그모이드(sigmoid) 활성화 함수가 (가)에 사용되었을 때, 출력값은? (단, 편향(bias)은 0이다)



① 0.0

② 0.5

③ 0.9

④ 1.0

정답 2번

출제 포인트

이 문제는 입력값과 가중치의 가중합을 먼저 구한 뒤, 그 값을 시그모이드 활성화 함수에 넣어 최종 출력을 판단하는 문제이다.

조건 정리

주어진 입력값은 0.5, 0.3, 0.2, 1.0이다.
주어진 가중치는 각각 0.6, -2.0, 1.5, 0.0이다.
편향은 0이므로 가중합에 추가되는 상수항은 없다.
구해야 할 값은 노드 (가)의 시그모이드 출력값이다.

NoteBase

사용할 공식

가중합은 $z = \sum x_i w_i + b$이다.
시그모이드 함수는 $\sigma(z) = \frac{1}{1+e^{-z}}$이다.
이 문제는 여러 입력과 가중치가 한 뉴런으로 들어오고, 그 결과에 시그모이드가 적용되는 전형적인 단일 뉴런 계산 구조이므로 이 공식을 사용한다.

계산 과정

1. 각 입력과 가중치를 곱하면 $0.5 \times 0.6 = 0.3$, $0.3 \times (-2.0) = -0.6$, $0.2 \times 1.5 = 0.3$, $1.0 \times 0.0 = 0.0$이다.
2. 이를 모두 더하면 $z = 0.3 - 0.6 + 0.3 + 0.0 + 0 = 0$이다.
3. 시그모이드 함수에 $z = 0$을 넣으면 $\sigma(0) = \frac{1}{1+e^{0}} = \frac{1}{2} = 0.5$이다.
4. 제시된 그래프를 보아도 $x = 0$일 때 출력은 0.5이므로 계산과 그래프 해석이 일치한다.

선지 판단

① X
0.0은 가중합 z 값일 뿐 시그모이드 출력값이 아니다. 틀린 표현은 0.0을 최종 출력으로 본 점이다. 올바르게는 가중합이 0이면 시그모이드 출력은 0.5이다.

② O
가중합이 0이고 시그모이드 함수값 $\sigma(0)$는 0.5이므로 맞는 선택지이다.

③ X
0.9는 입력값이 2 정도일 때의 시그모이드 출력에 가깝고, 이 문제의 가중합 0에 대한 값이 아니다. 틀린 표현은 출력이 0.9라는 점이다. 올바르게는 출력은 0.5이다.

④ X
1.0은 시그모이드 함수가 매우 큰 양수에서 가까워지는 값이지, $z = 0$에서의 정확한 출력이 아니다. 틀린 표현은 출력이 1.0이라는 점이다. 올바르게는 출력은 0.5이다.

오답 포인트

가중합 0과 최종 출력 0.5를 혼동하기 쉽다. 또한 음수 가중치 -2.0를 빼먹거나, 마지막 입력의 가중치가 0.0이라는 사실을 놓치면 다른 값이 나온다. 이 문항은 가중합을 먼저 구하고 그다음 활성화 함수를 적용하는 순서를 정확히 지키는지를 묻는다.

NoteBase

다시 나오면 이렇게 판단

뉴런 출력 문제는 입력×가중치를 모두 계산해 가중합을 만든 뒤, 편향을 더하고, 마지막에 활성화 함수를 적용하는 순서로 판단한다. 가중합과 활성화 함수 출력은 서로 다른 값이라는 점을 분리해서 본다.

암기 포인트

시그모이드는 $x = 0$일 때 항상 0.5이고, 뉴런의 최종 출력은 가중합 자체가 아니라 활성화 함수 통과 후의 값이다.

NoteBase

문 11. 합성곱 신경망(CNN, Convolutional Neural Network)에 대한 설명으로 옳지 않은 것은?

① 이미지 분류 업무에 CNN을 활용할 수 있다.

② CNN 내부에는 여러 개의 합성곱(convolution)층과 풀링(pooling)층을 배치할 수 있다.

③ CNN에서는 완전 연결(fully connected)층이 사용되지 않는다.

④ CNN에서는 학습 데이터에 과적합(overfitting)되는 문제를 해결하기 위해 드롭아웃(dropout) 기법이 사용될 수 있다.

정답 3번

출제 포인트

이 문제는 CNN의 기본 구성 요소와 활용, 과적합 완화 기법을 구별할 수 있는지를 묻는다.

핵심 개념

CNN은 이미지처럼 공간 구조가 있는 데이터를 처리하기 위해 합성곱층으로 특징을 추출하고, 필요하면 풀링층으로 차원을 줄이는 신경망이다.
합성곱층은 필터로 국소 특징을 뽑는 층이고, 풀링층은 중요한 특징을 유지하며 크기를 줄이는 층이다.
완전 연결층은 추출된 특징을 최종 분류로 연결하는 층이며, 전형적인 CNN 구조에서는 뒤쪽 분류기 부분에 배치될 수 있다.
드롭아웃은 일부 뉴런을 무작위로 제외해 특정 학습 데이터에만 과하게 맞춰지는 현상을 줄이는 기법이다.

선지별 해설

1번: O. CNN은 이미지의 지역적 패턴을 잘 추출하므로 이미지 분류 업무에 널리 활용된다.
2번: O. CNN은 여러 개의 합성곱층과 풀링층을 연속해서 쌓아 저수준 특징부터 고수준 특징까지 단계적으로 학습할 수 있다.
3번: X. 틀린 부분은 "완전 연결층이 사용되지 않는다"이다. 실제로는 전형적인 CNN에서 합성곱층 뒤에 완전 연결층을 두어 최종 분류를 수행할 수 있다. 최근 일부 구조는 완전 연결층을 줄이거나 생략하기도 있지만, CNN 일반에 대해 아예 사용되지 않는다고 말할 수는 없다.
4번: O. CNN도 학습 데이터에 과적합될 수 있으므로 이를 완화하기 위해 드롭아웃 기법을 사용할 수 있다.

오답 포인트

3번은 CNN을 합성곱층과 풀링층만으로 이루어진 구조처럼 단정한 것이 함정이다. CNN의 핵심은 합성곱 기반 특징 추출이지만, 전통적인 분류 CNN에서는 뒤쪽에 완전 연결층이 붙는 경우가 많다.

NoteBase

다시 나오면 이렇게 판단

CNN 문제에서는 활용 분야, 기본 층의 역할, 분류기 구성, 과적합 완화 기법을 나누어 본다. 특히 "항상", "전혀", "사용되지 않는다"처럼 구조를 절대적으로 단정하는 표현은 예외가 있는지 먼저 점검한다.

암기 포인트

CNN은 합성곱으로 특징을 뽑고, 필요하면 풀링으로 줄이며, 분류 단계에서는 완전 연결층과 드롭아웃을 함께 사용할 수 있다.

NoteBase

문 12. 생성 모델에 활용되는 자기 부호화기(auto encoder)에 대한 설명으로 옳지 않은 것은?

① 특징 추출에도 활용된다.

② 지도 학습(supervised learning)에 해당한다.

③ 인코더는 차원을 점점 줄이나 디코더는 차원을 점점 늘리기 때문에 출력층에서 원래 차원을 회복한다.

④ 변이형 자기 부호화기(variational auto encoder)는 잠재 변수의 발산을 방지하기 위한 규제화항을 오차에 포함하여 학습한다.

정답 2번

출제 포인트

이 문제는 자기 부호화기의 학습 성격, 인코더·디코더의 역할, 변이형 자기 부호화기의 규제화 목적을 구분하는 문제이다.

핵심 개념

자기 부호화기는 입력을 잠재표현으로 압축한 뒤 다시 원래 입력과 가깝게 복원하도록 학습하는 신경망이다.
일반적인 자기 부호화기는 별도의 정답 라벨 없이 입력 그 자체를 복원 목표로 삼으므로 비지도 또는 자기지도 성격으로 이해한다.
변이형 자기 부호화기는 재구성 오차에 잠재공간이 일정한 분포를 따르도록 하는 규제화항을 더해 생성에 적합한 잠재공간을 만든다.

선지별 해설

① O: 자기 부호화기의 인코더가 만든 잠재표현은 원본 데이터의 핵심 정보를 압축한 표현이므로 특징 추출에 활용된다.
② X: 틀린 표현은 '지도 학습(supervised learning)에 해당한다'이다. 자기 부호화기는 보통 입력 자체를 복원 대상으로 두므로 정답 라벨을 사용하는 지도 학습이 아니라 비지도 또는 자기지도 학습으로 본다.
③ O: 인코더는 보통 입력 차원을 줄여 잠재표현을 만들고, 디코더는 이를 다시 확장해 출력층에서 원래 입력 차원으로 복원한다. 따라서 출력층의 차원은 입력과 같아진다.
④ O: 변이형 자기 부호화기는 재구성 손실만 최소화하지 않고 잠재변수가 원하는 분포를 따르도록 규제화항을 함께 넣어 학습한다. 이 규제화가 잠재공간의 무질서한 확산을 줄이고 생성 가능성을 높인다.

오답 포인트

자기 부호화기는 입력과 출력이 모두 존재하므로 겉으로 보면 지도 학습처럼 보이지만, 복원 대상이 외부 라벨이 아니라 입력 자기 자신이라는 점이 핵심 함정이다. ②는 이 점을 놓치면 쉽게 고르기 어렵다.

NoteBase

다시 나오면 이렇게 판단

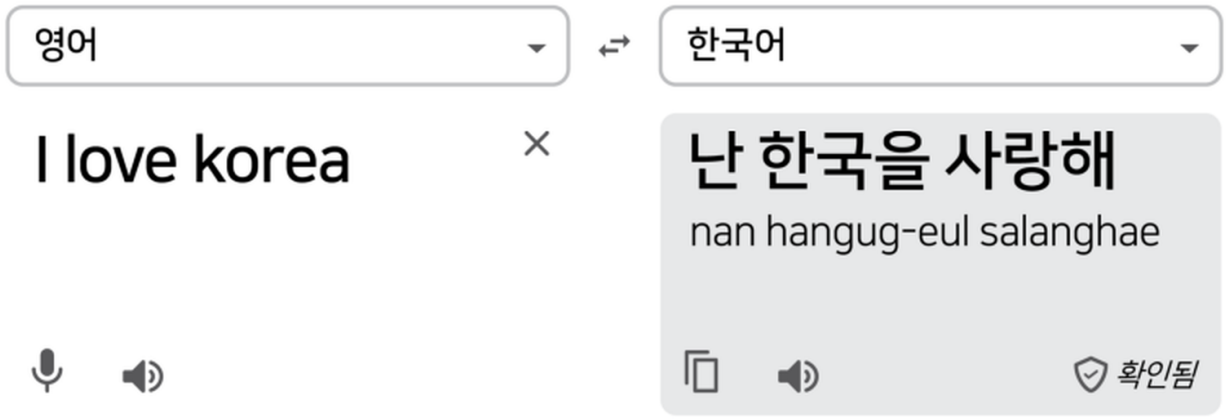
자기 부호화기 문제가 나오면 먼저 학습 목표가 외부 정답 라벨인지, 입력 자체의 복원인지부터 구분한다. 복원 대상이 입력 자기 자신이면 지도 학습이 아니라 비지도 또는 자기지도 계열로 판단한다.

암기 포인트

오토인코더는 압축 후 복원, VAE는 압축 후 분포까지 정리한다.

NoteBase

문 13. 영어로 말한 내용을 한국어로 듣게 해주는 스마트폰 기반 자동 번역 시스템에 대한 설명으로 옳지 않은 것은?



① 영어로 말한 내용을 스마트폰 마이크로 듣고, 음성 인식 기술을 사용하여 영어 문장으로 변환한다.

② 기계 번역에는 신경망을 이용하지 않는 기법도 있다.

③ 신경망을 이용하여 번역할 경우, 영어 문장을 일본어 문장으로, 일본어 문장을 한국어 문장으로 번역하는 방법을 미리 배운 신경망은 학습 데이터에 영어-한국어 문장 쌍들이 없어도 영어 문장을 한국어 문장으로 번역할 수 있다.

④ 글로 된 한국어 문장을 음성 인식 기술을 통해 한국어 음성으로 변환하고 스마트폰 스피커로 들려준다.

정답 4번

출제 포인트

이 문제는 스마트폰 기반 자동 번역 시스템의 처리 단계인 음성 인식, 기계 번역, 음성 합성의 역할 구분과 신경망 번역의 특성을 묻는다.

핵심 개념

음성 인식은 사람의 음성을 문자 형태의 텍스트로 바꾸는 기술이다. 기계 번역은 한 언어의 텍스트를 다른 언어의 텍스트로 바꾸는 기술이다. 음성 합성은 텍스트를 사람이 들을 수 있는 음성으로 바꾸는 기술이다. 신경망 기반 기계 번역은 다국어 표현을 공유하여 학습한 경우 병렬 문장쌍이 없는 언어쌍도 직접 번역할 수 있는 경우가 있다.

NoteBase

선지별 해설

1번 O: 영어 음성을 스마트폰 마이크로 입력받아 영어 문장으로 바꾸는 단계는 음성 인식이다. 말소리를 같은 언어의 텍스트로 바꾸는 설명이므로 옳다.
2번 O: 기계 번역에는 신경망 번역만 있는 것이 아니다. 규칙 기반 번역이나 통계적 기계 번역처럼 신경망을 이용하지 않는 기법도 존재하므로 옳다.
3번 O: 다국어 신경망 번역은 언어 사이의 공통 표현을 학습하여 영어-한국어 문장쌍이 없어도 영어 문장을 한국어로 번역하는 zero-shot 번역이 가능할 수 있다. 따라서 신경망 번역의 가능성을 설명한 선지로 볼 수 있다.
4번 X: 틀린 부분은 '음성 인식 기술을 통해 한국어 음성으로 변환'이다. 글로 된 문장을 소리로 바꾸는 단계는 음성 인식이 아니라 음성 합성이다. 올바르게는 글로 된 한국어 문장을 음성 합성 기술로 한국어 음성으로 변환하고 스피커로 들려준다.

오답 포인트

음성 인식과 음성 합성을 둘 다 음성과 관련된 기술로만 기억하면 4번을 맞는 설명처럼 보기 쉽다. 입력이 소리이고 출력이 글이면 음성 인식이고, 입력이 글이고 출력이 소리이면 음성 합성이라는 방향 구분이 핵심이다.

다시 나오면 이렇게 판단

통역 시스템 문제는 변환 방향을 먼저 본다. 소리→글이면 음성 인식, 글→다른 언어의 글이면 기계 번역, 글→소리이면 음성 합성으로 판단하면 된다.

암기 포인트

ASR은 소리→글, MT는 글→다른 글, TTS는 글→소리의 변환이다.

NoteBase

문 14. k-means 알고리즘의 실행 순서를 바르게 나열한 것은? (단, k는 클러스터 수이고, D는 n개의 데이터 집합이다)

(가) 각 데이터의 소속 클러스터가 바뀌지 않을 때까지 필요한 과정을 반복한다.
(나) 각 클러스터의 중심점을 다시 계산한다.
(다) 집합 D에서 k개의 데이터를 임의로 추출하고, 이 데이터들을 각 클러스터의 중심점으로 설정한다.
(라) 집합 D의 각 데이터에 대해 k개의 클러스터 중심점과의 거리를 계산하고, 그 거리가 가장 가까운 중심점의 소속 클러스터로 각 데이터를 할당한다.

① (다) - (가) - (라) - (나)
② (다) - (라) - (나) - (가)
③ (라) - (다) - (가) - (나)
④ (라) - (다) - (나) - (가)

정답 2번

출제 포인트

이 문제는 k-means 알고리즘의 기본 실행 흐름과 단계 사이의 선후관계를 정확히 구분할 수 있는지를 묻는다.

올바른 순서

1. (다) 집합 D에서 k개의 데이터를 임의로 추출해 초기 중심점으로 설정한다.
2. (라) 각 데이터를 가장 가까운 중심점의 클러스터에 할당한다.
3. (나) 현재 할당 결과를 바탕으로 각 클러스터의 중심점을 다시 계산한다.
4. (가) 소속 클러스터가 더 이상 바뀌지 않을 때까지 (라)와 (나)를 반복한다.

왜 이 순서인가: 중심점이 먼저 정해져야 거리 비교와 할당이 가능하고, 할당 결과가 있어야 새로운 중심점을 계산할 수 있기 때문이다.

선지별 해설

1번: X. 틀린 부분은 초기화 직후에 (가)를 둔 점이다. 반복 종료 조건은 할당과 중심점 재계산이 수행된 뒤에 확인할 수 있다. 올바르게는 (다) - (라) - (나) - (가)이다.
2번: O. 초기 중심점을 정한 뒤 각 데이터를 가장 가까운 중심점에 배정하고, 그 결과로 중심점을 다시 계산한 다음, 소속 변화가 없을 때까지 이를 반복하므로 올바른 순서이다.
3번: X. 틀린 부분은 (라)를 (다)보다 앞에 둔 점이다. 중심점이 정해지지 않으면 각 데이터가 어느 클러스터에 가장 가까운지 계산할 수 없다. 올바르게는 (다) - (라) - (나) - (가)이다.
4번: X. 틀린 부분은 (라)를 가장 먼저 둔 점이다. 데이터 할당은 초기 중심점 설정 이후에만 가능하다. 올바르게는 (다) - (라) - (나) - (가)이다.

NoteBase

오답 포인트

반복한다는 표현을 하나의 독립 단계처럼 초반에 배치하거나, 초기 중심점 설정 없이 바로 거리 계산을 시작하면 틀린다. k-means는 초기화 -> 할당 -> 중심점 재계산 -> 수렴 확인의 구조로 이해해야 한다.

다시 나오면 이렇게 판단

군집화 절차 문제는 먼저 초기값이 필요한지 확인하고, 그다음 현재 상태로 계산할 수 있는 단계와 이전 단계 결과를 입력으로 받는 단계를 구분하면 된다. k-means에서는 중심점 설정이 먼저, 할당이 다음, 중심점 갱신이 그다음, 반복 종료 판단이 마지막이다.

암기 포인트

k-means는 초기 중심점 설정 -> 최근접 클러스터 할당 -> 중심점 재계산 -> 수렴까지 반복의 순서이다.

NoteBase

문 15. 기계학습의 성능 향상을 위한 학습 데이터 전처리(preprocessing) 기법과 검증 기법에 대한 설명으로 옳지 않은 것은?

① 데이터 전처리의 한 예로 결손값(missing value) 처리가 있다.

② 원-핫 인코딩(one-hot encoding)에서는 해당 속성값의 특징에 의해 고윳값에 해당하는 위치에만 1을 표시하고 나머지는 0으로 표시한다.

③ k-폴드 교차검증(k-fold cross validation)은 주어진 데이터를 k개의 데이터 폴드 집합으로 만든 뒤 다양한 폴드 집합 조합에 대해 학습과 검증 과정을 반복적으로 수행하는 기법이다.

④ SMOTE(Synthetic Minority Over-sampling Technique)는 특정 분류(class)에 속하는 학습 데이터가 지나치게 많은 경우, 데이터 일부를 추출하여 제거하는데 사용된다.

정답 4번

출제 포인트

이 문제는 데이터 전처리, 범주형 인코딩, 교차검증, 데이터 불균형 처리 기법의 정의를 정확히 구별하는지를 묻는 문제이다.

핵심 개념

결손값 처리는 비어 있는 데이터를 보완하거나 제거하여 학습 가능 형태로 만드는 전처리이다.
원-핫 인코딩은 범주형 값을 각 범주별 이진 벡터로 바꾸는 표현 방법이다.
k-폴드 교차검증은 데이터를 k개 폴드로 나누고 번갈아 검증 폴드로 사용하여 모델 성능을 평가하는 검증 기법이다.
SMOTE는 소수 클래스 표본이 부족할 때 기존 소수 클래스 이웃을 바탕으로 합성 표본을 만들어 클래스 불균형을 완화하는 오버샘플링 기법이다.

선지별 해설

① O: 결손값 처리는 대표적인 데이터 전처리 작업이다. 누락된 값을 평균, 중앙값, 최빈값 등으로 대체하거나 해당 샘플을 제거하는 방식이 여기에 해당한다.
② O: 원-핫 인코딩은 해당 범주에 대응하는 위치만 1로 두고 나머지는 0으로 두는 방식이다. 범주 사이에 크기 관계가 없는 명목형 데이터를 수치화할 때 쓰인다.
③ O: k-폴드 교차검증은 데이터를 k개의 폴드로 나눈 뒤 한 폴드씩 검증에 쓰고 나머지를 학습에 사용하는 과정을 반복하는 방법이다. 모델의 일반화 성능을 한 번의 분할보다 안정적으로 평가할 수 있다.
④ X: 틀린 부분은 '특정 부류에 속하는 학습 데이터가 지나치게 많은 경우, 데이터 일부를 추출하여 제거하는데 사용된다'는 표현이다. 실제로 SMOTE는 다수 클래스를 제거하는 기법이 아니라 소수 클래스를 합성하여 늘리는 오버샘플링 기법이다.

NoteBase

오답 포인트

SMOTE를 랜덤 언더샘플링과 혼동하면 ④를 옳다고 보기 쉽다. 데이터 불균형 처리에서는 '많은 쪽을 줄이는가'와 '적은 쪽을 늘리는가'를 먼저 구분해야 한다.

다시 나오면 이렇게 판단

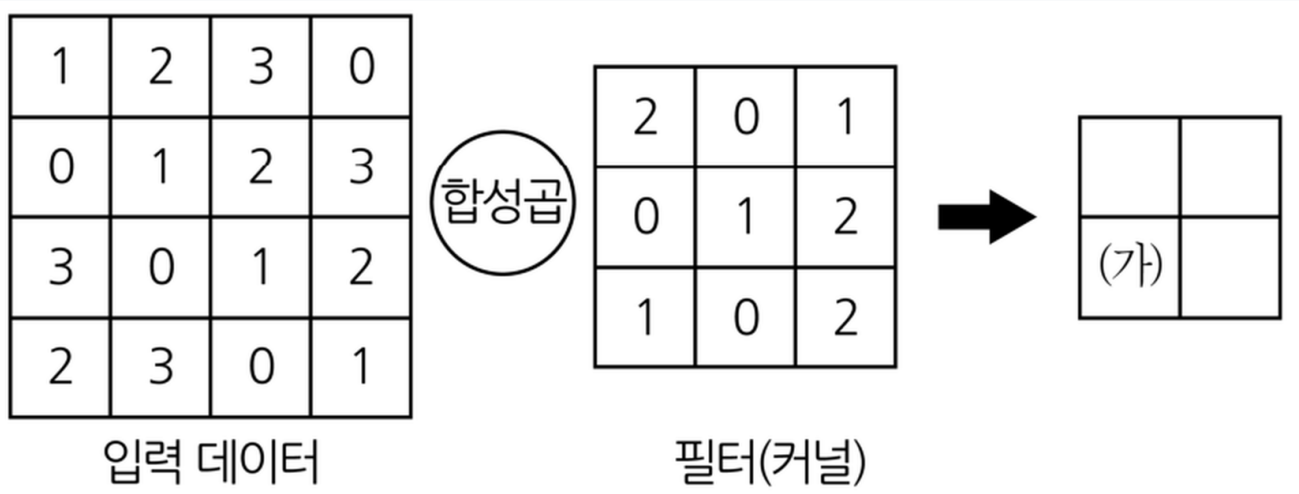
선지에 전처리, 인코딩, 검증, 불균형 처리 용어가 함께 나오면 먼저 각 기법이 무엇을 바꾸는지 판단하면 된다. 결손값 처리는 데이터 품질 보정, 원-핫 인코딩은 범주 표현 변환, k-폴드 교차검증은 성능 평가, SMOTE는 소수 클래스 증식으로 구분된다.

암기 포인트

SMOTE는 소수 클래스를 합성해 늘리고, 다수 클래스를 줄이는 것은 언더샘플링이다.

NoteBase

문 16. 합성곱 신경망(CNN, Convolutional Neural Network)에서 입력 데이터와 필터를 합성곱(convolution) 할 때, (가) 부분에 출력되는 값은? (단, 편향(bias)은 0이다)



① 5

② 6

③ 8

④ 9

정답 2번

출제 포인트

이 문제는 CNN의 valid 합성곱에서 출력의 특정 위치가 어떤 입력 부분 행렬과 필터의 원소별 곱의 합으로 계산되는지를 묻는다.

조건 정리

입력 데이터는 4행 4열이고 필터는 3행 3열이다.
편향은 0이므로 원소별 곱의 합만 계산한다.
패딩 없이 3행 3열 필터를 4행 4열 입력에 적용하므로 출력은 2행 2열이다.
(가)는 출력 2행 1열 위치이므로 입력의 2~4행, 1~3열 부분과 필터를 겹친다.

사용할 공식

출력 위치의 값은 $y_{i,j} = \sum_{m=0}^{2} \sum_{n=0}^{2} x_{i+m,j+n} k_{m,n} + b$로 계산된다.
이 공식을 쓰는 이유는 한 출력 칸이 필터와 겹친 입력 영역의 각 원소를 필터 원소와 곱한 뒤 모두 더한 값이기 때문이다.

NoteBase

계산 과정

(가)에 대응하는 입력 부분 행렬은 bmatrix0&1&2\3&0&1\2&3&0bmatrix이다.
필터는 bmatrix2&0&1\0&1&2\1&0&2bmatrix이다.
따라서 계산은
$(0 \times 2) + (1 \times 0) + (2 \times 1) + (3 \times 0) + (0 \times 1) + (1 \times 2) + (2 \times 1) + (3 \times 0) + (0 \times 2) = 6$이다.
편향이 0이므로 최종 출력값도 6이다.

선지 판단

① X: 5는 계산 결과 6과 다르다. 틀린 표현은 "5"이고, 올바르게는 "6"이다.
② O: (가)의 입력 부분 행렬과 필터의 원소별 곱을 모두 더하면 6이므로 맞다.
③ X: 8은 계산 결과 6과 다르다. 틀린 표현은 "8"이고, 올바르게는 "6"이다.
④ X: 9는 계산 결과 6과 다르다. 틀린 표현은 "9"이고, 올바르게는 "6"이다.

오답 포인트

출력 2행 1열이라는 위치를 잘못 잡아 입력의 다른 3행 3열 영역을 사용하면 오답이 나온다. 또한 CNN 문제에서는 도식에 제시된 필터를 그대로 겹쳐 원소별 곱의 합을 구하는 방식으로 처리하므로, 필터 위치와 겹치는 입력 영역을 먼저 확정해야 한다.

다시 나오면 이렇게 판단

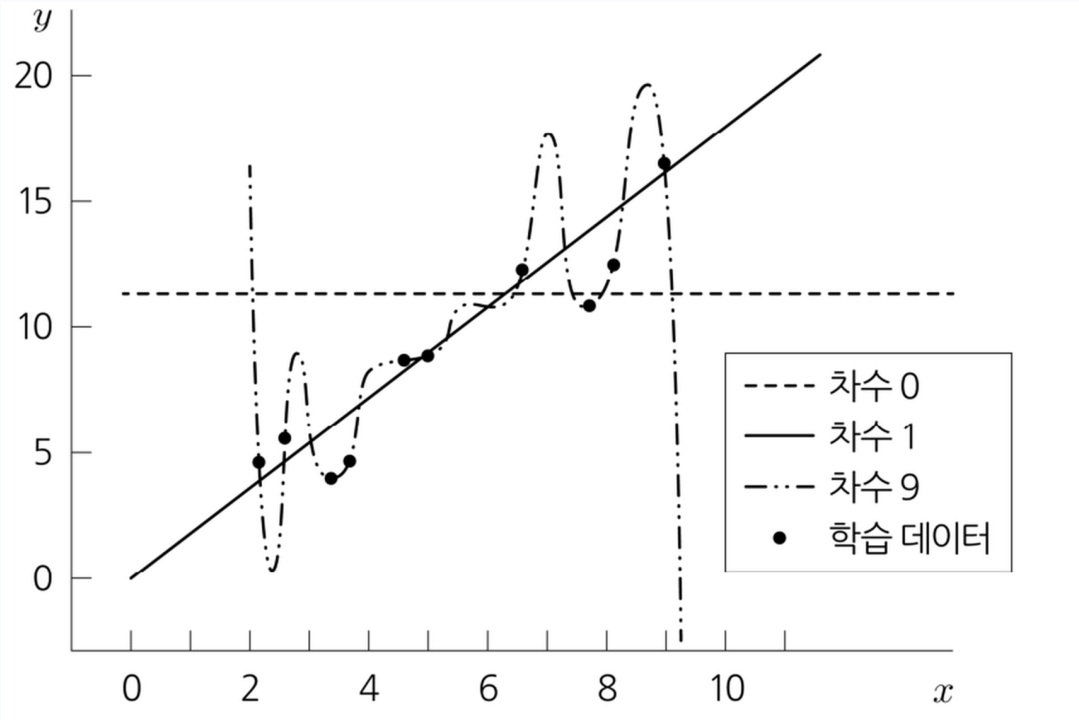
필터 크기와 출력 위치를 먼저 확인하고, 그 출력 칸에 대응하는 입력 부분 행렬을 고른 뒤 같은 위치끼리 곱해 모두 더한다. 편향이 주어지면 마지막에 더하고, 편향이 0이면 원소별 곱의 합이 그대로 출력값이 된다.

암기 포인트

CNN 합성곱 한 칸은 "겹친 입력 영역과 필터의 같은 자리 곱을 모두 더한 값"이다.

NoteBase

문 17. 다음 그래프에 대한 해석으로 옳지 않은 것은?



① 차수 9의 회귀 모델은 학습 데이터와 신규 데이터에 대해 모두 정확한 예측 결과를 나타낸다.

② 차수 1의 회귀 모델은 선형 회귀 모델로 사용할 수 있다.

③ 차수 0의 회귀 모델은 과소적합(underfitting)하기 때문에 학습 데이터와 신규 데이터에 대해 정확한 예측이 어렵다.

④ 차수 9의 회귀 모델은 과적합(overfitting)이 나타날 수 있다.

정답 1번

출제 포인트

이 문제는 그래프에서 차수 0, 1, 9 회귀 모델의 적합 상태를 비교하여 과소적합, 적절한 일반화, 과적합을 구별할 수 있는지를 묻는다.

NoteBase

선지별 해설

1번 X: 차수 9의 회귀 모델은 학습 데이터 점들을 거의 모두 지나가도록 복잡하게 맞춰져 있어 학습 데이터에는 매우 잘 맞지만 신규 데이터까지 정확하다고 볼 수는 없다. 틀린 표현은 "학습 데이터와 신규 데이터에 대해 모두 정확한 예측 결과"이며, 올바르게는 "학습 데이터에는 매우 잘 맞지만 신규 데이터에서는 일반화 성능이 떨어질 수 있다"이다.
2번 O: 차수 1의 회귀 모델은 직선으로 데이터를 설명하는 모델이므로 선형 회귀 모델로 사용할 수 있다. 그래프에서도 학습 데이터의 전체 증가 추세를 안정적으로 따른다.
3번 O: 차수 0의 회귀 모델은 상수 하나로만 예측하는 모델이어서 증가 추세를 반영하지 못한다. 그래프의 수평선처럼 데이터를 지나치게 단순화하므로 학습 데이터와 신규 데이터 모두에서 정확한 예측이 어렵다.
4번 O: 차수 9의 회귀 모델은 데이터 점마다 크게 요동하며 훈련 샘플에 과하게 맞춰진 모습이다. 이런 형태는 잡음까지 학습한 과적합의 전형적 예이다.

오답 포인트

훈련 데이터를 잘 맞춘다는 말과 새 데이터도 잘 맞춘다는 말을 같은 뜻으로 보면 1번을 놓치게 된다. 그래프에서 곡선이 데이터 점 사이에서 심하게 출렁이면 높은 정확도가 아니라 과적합 신호일 수 있다.

다시 나오면 이렇게 판단

회귀 그래프 문제는 먼저 모델의 복잡도와 데이터의 전체 추세가 얼마나 자연스럽게 맞는지를 본다. 수평선처럼 너무 단순하면 과소적합, 직선이나 완만한 곡선으로 추세를 잘 따르면 일반화가 양호한 모델, 데이터 점마다 과하게 꺾이면 과적합으로 판단한다.

암기 포인트

저차수는 과소적합 위험이 크고, 지나치게 고차수이면 학습 데이터는 잘 맞아도 신규 데이터에서는 과적합이 나타날 수 있다.

NoteBase

문 18. 기본 순환 신경망(RNN, Recurrent Neural Network)과 이의 변형 기법에 대한 설명으로 옳지 않은 것은?

① RNN은 계층이 깊어질수록 그래디언트 소멸 문제(vanishing gradient problem)가 발생하기 쉽다.

② LSTM(Long Short Term Memory)의 출력은 이전 시점의 입력값과 은닉층의 값뿐만 아니라 이후 시점의 입력값과 은닉층의 값에도 영향을 받는다.

③ RNN에서 순환 드롭아웃(recurrent dropout)은 과적합(overfitting)을 방지하기 위해 사용된다.

④ GRU(Gated Recurrent Unit)는 LSTM과 유사한 원리로 동작하지만, 게이트 수를 줄여 더 간결한 구조를 갖는다.

정답 2번

출제 포인트

이 문제는 기본 RNN, LSTM, GRU의 정보 전달 방식과 각 변형 기법의 목적을 구별할 수 있는지를 묻는다.

핵심 개념

RNN은 이전 시점의 은닉 상태를 현재 계산에 반영하는 순환 신경망이다.
LSTM은 셀 상태와 게이트를 이용해 장기 의존성을 더 안정적으로 학습하도록 만든 RNN 변형이다.
GRU는 LSTM과 같은 목적을 더 적은 게이트로 구현한 간결한 구조이다.

선지별 해설

1번: O. 기본 RNN은 시간축으로 긴 의존성을 따라 역전파할 때 기울기가 반복적으로 작아져 그래디언트 소멸이 발생하기 쉽다. 그래서 긴 시퀀스 학습에 약하다.
2번: X. 틀린 부분은 "이후 시점의 입력값과 은닉층의 값에도 영향을 받는다"는 표현이다. 기본적인 단방향 LSTM의 현재 시점 출력은 이전 시점의 상태와 현재 입력의 영향을 받아 계산되며, 아직 처리하지 않은 이후 시점 입력의 영향을 직접 받지 않는다. 올바르게는 "이전 시점의 상태와 현재 입력을 바탕으로 출력이 계산된다"이다.
3번: O. recurrent dropout은 순환 연결에 드롭아웃을 적용해 과적합을 줄이기 위한 정규화 기법이다. 적용 방식에 주의가 필요하지만 목적 자체는 과적합 방지이다.
4번: O. GRU는 LSTM과 마찬가지로 게이트를 이용해 장기 의존성 문제를 완화하지만, 보통 더 적은 수의 게이트를 사용해 구조가 단순하다.

오답 포인트

LSTM이 문맥 정보를 오래 보존한다는 말 때문에 미래 시점 정보까지 직접 사용한다고 오해하기 쉽다. 단방향 RNN 계열은 현재 출력이 과거와 현재 정보로 계산된다는 점을 놓치면 2번을 맞는 설명처럼 보기 쉽다.

NoteBase

다시 나오면 이렇게 판단

RNN 계열 문제는 먼저 정보가 흐르는 방향이 과거에서 현재인지, 미래 정보까지 쓰는 별도 구조인지부터 본다. 그다음 기본 RNN의 약점은 그래디언트 소멸, LSTM과 GRU의 목적은 장기 의존성 완화, GRU의 특징은 더 단순한 게이트 구조로 정리하면 된다.

암기 포인트

기본 RNN은 소멸에 약하고, LSTM은 셀 상태로 보완하며, GRU는 더 단순하게 줄인 구조이다.

NoteBase

문 19. 출력 노드가 3개인 신경망의 출력이 소프트맥스(softmax) 함수를 이용해 다음과 같이 계산되었을 때, 입력 샘플 3개에 대한 교차 엔트로피 오차(cross entropy error)의 평균값은? (단, 목표 출력은 원-핫 인코딩(one-hot encoding)을 사용하고, 교차 엔트로피 식에서 로그는 자연로그를 사용하며 자연로그 계산값은 아래 표를 이용한다)

<입력 샘플에 대한 실제 출력 및 목표 출력값>

입력 샘플	실제 출력			목표 출력		
샘플 #1	0.7	0.2	0.1	1	0	0
샘플 #2	0.2	0.5	0.3	0	1	0
샘플 #3	0.6	0.1	0.3	0	0	1

<자연로그 계산표>

ln(0.1)	ln(0.2)	ln(0.3)	ln(0.4)	ln(0.5)	ln(0.6)	ln(0.7)	ln(0.8)	ln(0.9)
−2.30	−1.61	−1.20	−0.92	−0.69	−0.51	−0.36	−0.22	−0.11

① 0.25

② 0.50

③ 0.75

④ 1.00

정답 3번

출제 포인트

이 문제는 소프트맥스 출력과 원-핫 타깃이 주어졌을 때 평균 교차 엔트로피 오차를 구하는 문제이다.

조건 정리

주어진 값: 샘플 #1의 정답 클래스 확률은 0.7, 샘플 #2는 0.5, 샘플 #3은 0.3이다.
구해야 할 값: 3개 샘플의 평균 교차 엔트로피 오차이다.
주의할 조건: 원-핫 인코딩이므로 각 샘플에서 정답 클래스의 확률만 골라 $-\ln p$를 계산한 뒤, 마지막에 3개 샘플 평균을 내야 한다.

NoteBase

사용할 공식

교차 엔트로피 오차는 $E = -\sum t_k \ln y_k$이다. 원-핫 인코딩에서는 정답 클래스만 1이고 나머지는 0이므로 각 샘플의 오차는 E=-(정답 클래스 확률)로 단순화된다. 여러 샘플의 평균 오차는 각 샘플 오차의 합을 샘플 수로 나눈 값이다.

계산 과정

1. 샘플 #1의 정답은 첫 번째 클래스이므로 오차는 $-\ln(0.7) = 0.36$이다.
2. 샘플 #2의 정답은 두 번째 클래스이므로 오차는 $-\ln(0.5) = 0.69$이다.
3. 샘플 #3의 정답은 세 번째 클래스이므로 오차는 $-\ln(0.3) = 1.20$이다.
4. 평균 오차는 $(0.36 + 0.69 + 1.20)/3 = 2.25/3 = 0.75$이다.

선지별 해설

① X
평균 교차 엔트로피는 0.25가 아니다. 실제 계산값은 0.75이므로 틀린 표현은 평균값이 0.25라는 부분이며, 올바르게는 0.75이다.

② X
평균 교차 엔트로피는 0.50이 아니다. 각 샘플 손실을 모두 합산해 3으로 나누면 0.75이므로 틀린 표현은 평균값이 0.50이라는 부분이며, 올바르게는 0.75이다.

③ O
샘플별 오차가 0.36, 0.69, 1.20이고 평균이 0.75이므로 맞는 선택지이다.

④ X
1.00은 평균값이 아니라 실제 평균보다 큰 값이다. 세 샘플의 교차 엔트로피 합은 2.25이고 이를 3으로 나누어야 하므로 틀린 표현은 평균값이 1.00이라는 부분이며, 올바르게는 0.75이다.

오답 포인트

교차 엔트로피는 각 출력값 전체를 다시 더하는 문제가 아니라 정답 클래스 확률만 뽑아 $-\ln p$를 계산하는 문제라는 점이 함정이다. 또 샘플별 손실을 구한 뒤 마지막에 평균을 내야 하는데, 이 평균 단계를 빼먹거나 로그 부호를 잘못 처리하면 오답이 된다.

다시 나오면 이렇게 판단

원-핫 인코딩과 소프트맥스가 함께 나오면 각 샘플에서 정답 위치의 확률만 찾고, 그 값에 $-\ln$을 취한 뒤, 여러 샘플이면 마지막에 평균을 낸다고 판단하면 된다.

NoteBase

암기 포인트

원-핫 교차 엔트로피는 정답 칸만 보고 $-\ln p$, 여러 샘플은 마지막에 평균이다.

NoteBase

문 20. 그래디언트 소멸 문제(vanishing gradient problem)에 대한 설명으로 옳지 않은 것은?

① 특정 활성화 함수를 사용하는 계층이 신경망에 많이 추가되면 학습이 되지 않는 현상을 말한다.

② 0이나 0에 가까운 수가 시그모이드(sigmoid) 함수에 들어오면 그래디언트가 0에 가까워 학습이 되지 않는다.

③ 그래디언트의 절댓값이 너무 작아지면 학습이 되지 않아 가중치(weight)와 편향(bias) 값들이 효과적으로 업데이트 되지 못한다.

④ 그래디언트 소멸 문제를 해결하기 위해 DNN(Deep Neural Network)에서는 활성화 함수로 ReLU 함수를 사용한다.

정답 2번

출제 포인트

이 문제는 그래디언트 소멸의 발생 원인과 이를 줄이는 대표적 방법을 정확히 구분하는지를 묻는 문제이다.

핵심 개념

그래디언트 소멸은 역전파 과정에서 미분값이 여러 층을 거치며 매우 작아져 앞쪽 층까지 학습 신호가 제대로 전달되지 않는 현상이다.
시그모이드는 입력이 0 부근일 때 기울기가 비교적 크고, 큰 양수나 큰 음수로 포화될 때 기울기가 0에 가까워진다.
ReLU는 양수 구간에서 기울기가 유지되어 깊은 신경망에서 그래디언트 소멸을 완화하는 데 자주 쓰인다.

NoteBase

선지별 해설

① O
그래디언트 소멸은 깊은 신경망에서 특정 활성화 함수의 미분값이 반복 곱해지면서 학습 신호가 약해지는 현상을 말하므로 옳다.

② X
틀린 부분은 "0이나 0에 가까운 수가 시그모이드 함수에 들어오면 그래디언트가 0에 가까워"라는 표현이다. 시그모이드는 입력이 0 부근일 때 오히려 기울기가 가장 크고, 입력이 매우 크거나 매우 작아 포화 구간에 들어갈 때 그래디언트가 0에 가까워진다.
틀린 표현을 고치면 "큰 양수나 큰 음수가 시그모이드 함수에 들어오면 그래디언트가 0에 가까워져 학습이 잘 되지 않는다"가 된다.

③ O
그래디언트의 절댓값이 너무 작아지면 역전파로 전달되는 갱신량도 작아져 가중치와 편향이 거의 바뀌지 않으므로 옳다.

④ O
그래디언트 소멸을 줄이기 위한 대표적 방법 중 하나가 시그모이드 대신 ReLU 계열 활성화 함수를 사용하는 것이므로 옳다.

오답 포인트

시그모이드는 "출력이 0과 1 사이"라는 이유만으로 입력 0 부근에서도 기울기가 작다고 오해하기 쉽다. 이 문제의 함정은 시그모이드의 중앙 구간과 포화 구간을 뒤바꿔 제시한 ②에 있다.

다시 나오면 이렇게 판단

활성화 함수 관련 선지가 나오면 먼저 기울기가 가장 큰 구간과 0에 가까워지는 포화 구간을 구분한다.
시그모이드는 입력 0 부근이 아니라 양끝 포화 구간에서 그래디언트 소멸이 심해진다고 판단하면 된다.

암기 포인트

시그모이드는 가운데에서 기울기가 크고 양끝에서 죽으며, ReLU는 양수 구간에서 기울기를 살린다.

NoteBase

문 21. 전문가 시스템의 데이터베이스에 4개의 사실 A, B, C, D가 있고 다음과 같은 4개의 규칙이 존재할 때, 순방향 연결 기법을 통해 데이터베이스에 새롭게 추가되는 사실의 최대 개수는?

◯ IF Y AND D AND E, THEN Z
◯ IF X AND B, THEN Y
◯ IF A, THEN X
◯ IF B, THEN W

① 1
② 2
③ 3
④ 4

정답 3번

출제 포인트

이 문제는 순방향 연결에서 현재 사실로 발동되는 규칙을 앞에서부터 추적하여 새로 추가되는 사실의 수를 세는 문제이다.

올바른 순서

1. 초기 사실 확인 - 데이터베이스에는 A, B, C, D가 들어 있다.
2. 3번 규칙 적용 - A가 있으므로 X가 추가된다.
3. 4번 규칙 적용 - B가 있으므로 W가 추가된다.
4. 2번 규칙 적용 - X와 B가 모두 있으므로 Y가 추가된다.
5. 1번 규칙 검토 - Y와 D는 있지만 E가 없으므로 Z는 추가되지 않는다.

왜 이 순서: Y는 X가 먼저 만들어져야만 생기고, Z는 E가 전제인데 E를 만드는 규칙이 없으므로 끝까지 성립하지 않는다.

NoteBase

선지별 해설

① X
이유: 1개만 추가되는 것이 아니다. A로부터 X, B로부터 W, 그리고 X와 B로부터 Y까지 총 3개가 추가된다.
틀린 표현: 1개
올바르게는: 새로 추가되는 사실은 X, W, Y의 3개이다.

② X
이유: 2개로 멈추지 않는다. X와 W를 만든 뒤 X와 B를 이용해 Y까지 더 추론된다.
틀린 표현: 2개
올바르게는: 새로 추가되는 사실은 X, W, Y의 3개이다.

③ O
이유: 초기 사실 A, B, C, D에서 출발하면 X, W, Y는 순차적으로 추가되지만, Z는 E가 없어 생성되지 않는다.
따라서 새 사실의 최대 개수는 3개이다.

④ X
이유: 4개가 되려면 Z까지 추가되어야 하는데, 1번 규칙의 전제인 E가 처음에도 없고 다른 규칙으로도 만들어지지 않는다.
틀린 표현: 4개
올바르게는: Z는 추론되지 않으므로 새로 추가되는 사실은 3개이다.

오답 포인트

1번 규칙의 Y와 D만 보고 Z까지 나온다고 착각하기 쉽다. 그러나 순방향 연결은 전제 조건이 모두 만족되어야 발동되므로, 빠진 전제 E를 끝까지 확인해야 한다.

다시 나오면 이렇게 판단

규칙기반 추론 문제는 초기 사실을 먼저 적고, 각 규칙의 전제가 모두 만족되는지 순서대로 확인한 뒤, 새 사실이 생길 때마다 다시 전제를 갱신하여 본다. 마지막에는 어떤 전제가 끝내 충족되지 않는지를 확인하면 최대 추론 개수를 안정적으로 셀 수 있다.

암기 포인트

순방향 연결은 있는 사실에서 출발해 전제가 모두 맞는 규칙만 발동한다. 전제 하나라도 비면 결론은 추가되지 않는다.

NoteBase

문 22. 학습 속도 향상을 위해 입력 데이터를 정규화하는 신경망에서 입력 데이터가 다음과 같이 나이를 나타내는 특징 데이터로 주어졌을 때, 24세에 해당하는 정규화 결과는? (단, 데이터 정규화는 정규분포를 이용하는 표준화를 의미한다)

40, 30, 22, 24

① $-\frac{5}{7}$
② $-\frac{5}{14}$
③ $-\frac{5}{\sqrt{14}}$
④ $-\frac{2}{7}$

정답 1번

출제 포인트

이 문제는 나이 데이터에 표준화 공식을 적용해 24세의 z-점수를 구하는 문제이다.

조건 정리

주어진 값: 40, 30, 22, 24
구해야 할 값: 24세의 표준화 결과
주의할 조건: 정규화는 최소-최대 정규화가 아니라 정규분포 기반 표준화이므로 $z = \frac{x-\mu}{\sigma}$를 사용한다.

사용할 공식

표준화 공식은 $z = \frac{x-\mu}{\sigma}$이다.
이 공식을 쓰는 이유: 문제에서 정규화를 정규분포를 이용하는 표준화라고 명시했으므로, 원자료 24를 평균에서 얼마나 떨어졌는지 표준편차 단위로 바꾸어야 한다.

계산 과정

1. 평균은 $\mu = \frac{40+30+22+24}{4} = 29$이다.
2. 분산은 $\frac{(40-29)^2+(30-29)^2+(22-29)^2+(24-29)^2}{4} = \frac{121+1+49+25}{4} = 49$이다.
3. 표준편차는 $\sigma = \sqrt{49} = 7$이다.
4. 24세의 표준화 값은 $z = \frac{24-29}{7} = \frac{-5}{7}$이다.

NoteBase

선지 판단

① O
- 계산 결과가 $-\frac{5}{7}$이므로 맞는 선택지이다.

② X
- $-\frac{5}{14}$는 계산 결과와 다르다. 표준화는 평균 차이 -5를 표준편차 7로 나누어야 하므로 분모가 14가 될 수 없다.
- 틀린 표현은 $-\frac{5}{14}$이고, 올바르게는 $-\frac{5}{7}$이다.

③ X
- $-\frac{5}{\sqrt{14}}$는 표준편차를 잘못 잡은 값이다. 이 데이터의 분산은 49이므로 표준편차는 $\sqrt{49}=7$이어야 한다.
- 틀린 표현은 분모 $\sqrt{14}$이고, 올바르게는 분모가 7이어서 값은 $-\frac{5}{7}$이다.

④ X
- $-\frac{2}{7}$는 평균과의 차이를 잘못 계산한 값이다. 24와 평균 29의 차이는 -2가 아니라 -5이다.
- 틀린 표현은 분자 -2이고, 올바르게는 분자가 -5여서 값은 $-\frac{5}{7}$이다.

오답 포인트

이 문항의 함정은 정규화를 최소-최대 정규화로 착각하거나, 분산과 표준편차를 혼동하거나, $24-29$를 잘못 계산하는 데 있다. 특히 표준화는 반드시 표준편차로 나눈다는 점이 핵심이다.

다시 나오면 이렇게 판단

표준화 문제는 먼저 평균을 구하고, 다음으로 분산과 표준편차를 계산한 뒤, 마지막에 $z=\frac{x-\mu}{\sigma}$에 대입하는 순서로 판단한다. 분모가 분산인지 표준편차인지부터 구분하면 오답을 줄일 수 있다.

암기 포인트

표준화는 평균을 빼고 표준편차로 나누는 값이므로 $z=\frac{x-\mu}{\sigma}$, 분산이 아니라 표준편차가 분모이다.

NoteBase

문 23. 가로, 세로 크기가 31×21 픽셀인 원본 이미지에 다음과 같은 설정으로 합성곱(convolution) 연산을 한 번 수행한 후 생성되는 이미지의 크기는?

◯ 필터 크기: 3×3
◯ 패딩 폭: 1
◯ 스트라이드 값: 2

① 15×10
② 16×11
③ 29×19
④ 31×21

정답 2번

출제 포인트

이 문제는 합성곱 출력 크기 공식을 이용해 패딩과 스트라이드가 적용된 출력 이미지 크기를 계산하는 문제이다.

조건 정리

주어진 값: 원본 크기 31×21, 필터 3×3, 패딩 1, 스트라이드 2이다.
구해야 할 값: 합성곱 1회 수행 후 출력 이미지의 가로·세로 크기이다.
주의할 조건: 패딩 1이므로 양쪽에 각각 1픽셀씩 더해지고, 스트라이드 2이므로 필터가 2칸씩 이동한다.

사용할 공식

출력 크기는 ≤ft(입력-필터+2× 패딩)/(스트라이드)+1이다.
이 공식을 쓰는 이유: 필터가 패딩이 적용된 입력 위를 스트라이드 간격으로 이동할 때 시작 가능한 위치의 개수가 곧 출력 크기가 되기 때문이다.

계산 과정

가로는 $\left\lfloor \frac{31-3+2\times1}{2} \right\rfloor + 1 = \left\lfloor \frac{30}{2} \right\rfloor + 1 = 15 + 1 = 16$이다.
세로는 $\left\lfloor \frac{21-3+2\times1}{2} \right\rfloor + 1 = \left\lfloor \frac{20}{2} \right\rfloor + 1 = 10 + 1 = 11$이다.
따라서 출력 이미지 크기는 16×11이다.

NoteBase

선지별 해설

① X: 15×10은 가로와 세로 계산에서 마지막 +1을 빠뜨린 값이다. 틀린 값 → 올바르게는 16×11이다.
② O: 공식에 값을 대입하면 가로 16, 세로 11이므로 16×11이 맞다.
③ X: 29×19는 스트라이드 1, 패딩 0처럼 생각해 입력-필터+1만 적용할 때 나오는 값이다. 이 문제는 패딩 1과 스트라이드 2가 있으므로 틀린 값 → 올바르게는 16×11이다.
④ X: 31×21은 입력과 출력 크기가 같다고 본 값이다. 그러나 이 문제는 스트라이드가 2이므로 출력 크기가 줄어든다. 틀린 값 → 올바르게는 16×11이다.
계산 결과 16×11이므로 정답은 2번이다.

오답 포인트

가장 흔한 실수는 출력 크기 공식의 마지막 +1을 빼먹거나, 패딩과 스트라이드를 무시하고 입력-필터+1만 적용하는 것이다. ③은 패딩·스트라이드 무시 함정이고, ④는 same padding이면 항상 입력 크기와 같다고 착각하는 함정이다.

다시 나오면 이렇게 판단

합성곱 크기 문제는 가로와 세로를 각각 ≤ft(입력-필터+2P)/(S)+1에 넣어 계산하면 된다. 먼저 패딩 P, 스트라이드 S를 확인한 뒤 마지막 +1까지 포함해 계산하면 된다.

암기 포인트

합성곱 출력 크기 = ≤ft(입력-필터+2P)/(S)+1이다.

NoteBase

문 24. 유전 알고리즘(genetic algorithm)에서 수행되는 과정만을 모두 고르면?

ㄱ. 가지치기(pruning)
ㄴ. 교차(crossover)
ㄷ. 군집화(clustering)
ㄹ. 변이(mutation)
ㅁ. 적합도 평가(evaluation)

① ㄱ, ㄴ, ㅁ
② ㄴ, ㄷ, ㄹ
③ ㄴ, ㄹ, ㅁ
④ ㄷ, ㄹ, ㅁ

정답 3번

출제 포인트

이 문항은 유전 알고리즘의 표준 과정과 다른 인공지능 기법을 구별하는지를 묻는 문제이다.

핵심 개념

유전 알고리즘은 적합도가 높은 해를 선택하고, 교차와 변이를 반복하여 더 나은 해를 찾는 탐색 기법이다.
가지치기는 탐색트리에서 불필요한 분기를 제거하는 기법이고, 군집화는 데이터를 비슷한 집단으로 묶는 비지도학습 기법이다.
교차는 부모 해의 일부를 섞어 새로운 해를 만들고, 변이는 일부 유전자를 무작위로 바꾸며, 적합도 평가는 각 해의 우수성을 계산한다.

선지별 해설

1번 X: ㄴ 교차와 ㅁ 적합도 평가는 맞지만 ㄱ 가지치기가 틀리다. 틀린 표현은 'ㄱ의 포함'이다. 가지치기는 유전 알고리즘의 기본 과정이 아니라 탐색 가지를 잘라 내는 기법이므로, 올바르게는 'ㄱ을 제외하고 ㄴ, ㄹ, ㅁ을 포함한다'이다.
2번 X: ㄴ 교차와 ㄹ 변이는 맞지만 ㄷ 군집화가 틀리다. 틀린 표현은 'ㄷ의 포함'이다. 군집화는 유전 알고리즘의 과정이 아니라 비슷한 데이터들을 묶는 학습 기법이므로, 올바르게는 'ㄷ 대신 ㅁ 적합도 평가가 포함된다'이다.
3번 O: ㄴ 교차, ㄹ 변이, ㅁ 적합도 평가는 모두 유전 알고리즘에서 수행되는 대표 과정이다. 교차는 해의 조합, 변이는 해의 다양성 확보, 적합도 평가는 해의 우수성 판단에 해당한다.
4번 X: ㄹ 변이와 ㅁ 적합도 평가는 맞지만 ㄷ 군집화가 틀리다. 틀린 표현은 'ㄷ의 포함'이다. 올바르게는 '군집화는 제외되고 교차가 포함되어 ㄴ, ㄹ, ㅁ이 된다'이다.

NoteBase

오답 포인트

가지치기와 군집화도 인공지능 용어이므로 유전 알고리즘 과정처럼 보이기 쉽다는 점이 함정이다. 특히 '군집화'를 집단을 만든다는 말만 보고 개체 집단 진화 과정과 혼동하면 2번이나 4번으로 흔들리기 쉽다.

다시 나오면 이렇게 판단

유전 알고리즘 문제는 각 용어가 '해를 평가하거나 새 해를 만드는 진화 연산인지'를 기준으로 판단한다. 적합도 평가, 선택, 교차, 변이는 유전 알고리즘 쪽이고, 가지치기나 군집화처럼 다른 목적의 기법은 제외하면 된다.

암기 포인트

유전 알고리즘 핵심은 평가-선택-교차-변이이고, 가지치기와 군집화는 별개 기법이다.

NoteBase

문 25. 다음 설명에 해당하는 자연어처리 기법은?

주변에 있는 단어를 이용해 중간에 있는 특정 단어를 추론하는 방법으로, 큰 차원으로 표현된 특정 단어의 원-핫 인코딩(one-hot encoding) 벡터를 작은 차원의 실수 벡터로 바꾸어 줄 수 있다.

① CBOW(Continuous Bag of Words)

② n-Gram

③ Skip-Gram

④ TF-IDF(Term Frequency-Inverse Document Frequency) vector

정답 1번

출제 포인트

이 문제는 CBOW와 Skip-Gram, n-Gram, TF-IDF의 역할 차이를 구별하는 문제이다.

핵심 개념

CBOW는 주변 단어들을 입력으로 사용해 가운데의 목표 단어를 예측하는 Word2Vec 방식이며, 학습 과정에서 원-핫 벡터가 저차원의 밀집 임베딩으로 변환된다.
n-Gram은 연속한 n개의 단어 묶음을 다루는 표현 방식이다.
Skip-Gram은 중심 단어를 이용해 주변 단어를 예측하는 방식으로 CBOW와 예측 방향이 반대이다.
TF-IDF는 문서 안팎의 단어 빈도를 이용해 단어의 중요도를 수치화한 가중치 벡터이다.

NoteBase

선지별 해설

① O
주변에 있는 단어들로 가운데 특정 단어를 맞히는 설명은 CBOW의 정의와 일치한다. Word2Vec의 CBOW는 입력 단어를 원-핫으로 두고 은닉층의 가중치를 통해 저차원 임베딩을 학습한다.

② X
틀린 부분은 주변 단어로 중간 단어를 추론한다는 점을 n-Gram에 연결한 표현이다.
n-Gram은 인접한 단어를 n개씩 잘라서 연속 패턴을 보는 방식이며, 가운데 단어를 예측해 임베딩을 학습하는 모델이 아니다. 올바르게는 연속한 단어 묶음을 표현하는 통계적 언어 모델의 기본 단위이다.

③ X
틀린 부분은 주변 단어를 이용해 중심 단어를 예측한다는 설명을 Skip-Gram에 대응시킨 표현이다.
Skip-Gram은 반대로 중심 단어 하나로 주변 단어들을 예측한다. 올바르게는 예측 방향이 CBOW와 반대인 Word2Vec 방식이다.

④ X
틀린 부분은 원-핫 벡터를 작은 차원의 실수 벡터로 바꾸는 임베딩 학습을 TF-IDF로 본 표현이다.
TF-IDF는 단어의 중요도를 계산하는 가중치 방식일 뿐, 주변 문맥으로 중심 단어를 예측하거나 저차원 임베딩을 학습하는 모델이 아니다. 올바르게는 문서 내 빈도와 역문서 빈도를 결합한 벡터 표현이다.

오답 포인트

CBOW와 Skip-Gram은 둘 다 Word2Vec 계열이라 혼동하기 쉽지만 예측 방향이 반대이다. 또한 n-Gram과 TF-IDF는 단어 패턴이나 중요도를 표현하는 방법이지, 문맥 예측으로 임베딩을 학습하는 신경망 방식이 아니다.

다시 나오면 이렇게 판단

설명에 주변 단어로 중심 단어를 맞힌다고 나오면 CBOW로 판단한다. 중심 단어로 주변 단어를 맞히면 Skip-Gram이고, 연속한 단어 묶음이면 n-Gram이며, 단어 중요도 가중치이면 TF-IDF로 구분된다.

암기 포인트

CBOW는 주변→중심, Skip-Gram은 중심→주변, n-Gram은 연속 묶음, TF-IDF는 중요도 가중치이다.

NoteBase