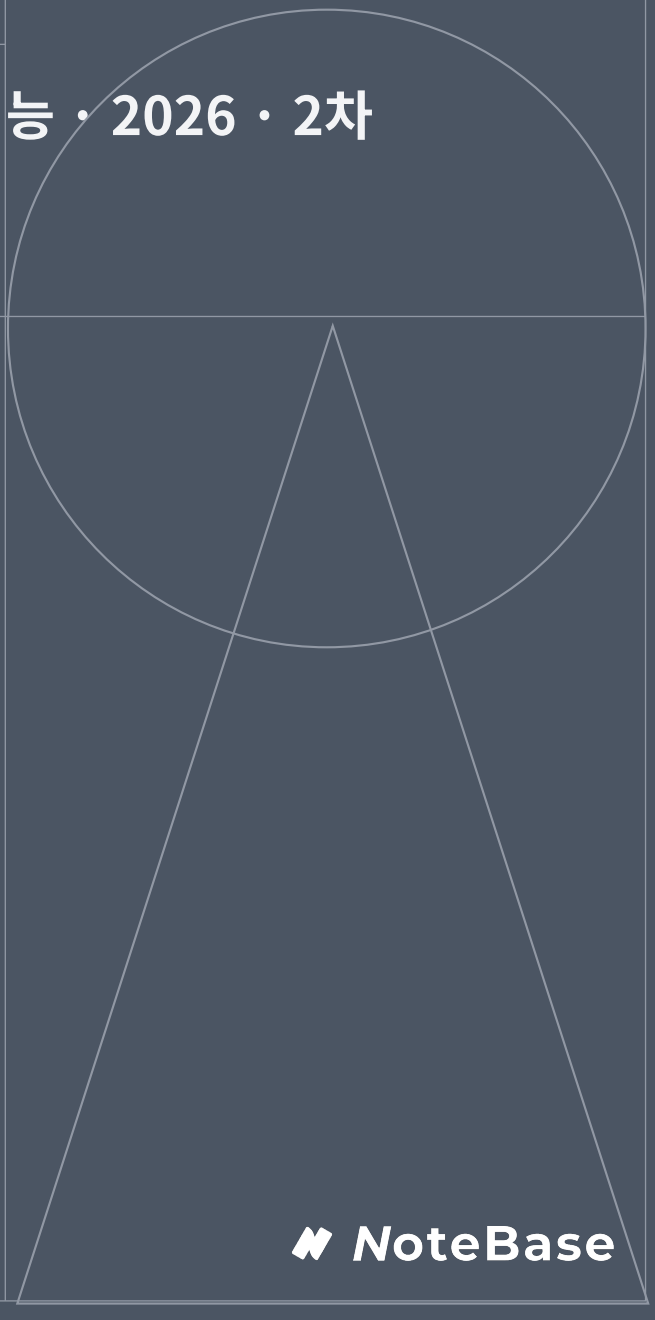
기출OX 해설집

공무원/국가직/7급 · 인공지능 · 2026 · 2차



NoteBase

기출OX 해설집

공무원/국가직/7급 · 인공지능 · 2026 · 2차

공무원/국가직/7급 · 인공지능 · 2026 · 2차

문 1. (가)에 해당하는 인공지능 개념은?

◯ 인공지능이 인간의 지능을 초월하는 시점을 '인공지능의 (가)'이라 한다.
◯ (가)에 도달하면 인공지능은 자신보다 더 좋은 지능을 가진 또 다른 인공지능을 설계하고 생산할 수 있게 된다.
◯ 인공지능 과학자 겸 미래학자인 레이 커즈와일(Ray Kurzweil)은 그의 저서 『(가)이 온다』에서 2045년 그 시점이 도래할 것으로 예측했다.

① 분기점(branch point)
② 임계점(critical point)
③ 전환점(turning point)
④ 특이점(singularity)

정답 4번

간단한 해설

인공지능이 인간의 지능을 넘어 스스로 더 뛰어난 인공지능을 만들어 내는 시점을 **특이점(singularity)**이라고 한다. 커즈와일의 『특이점이 온다』와 2045년이라는 예측도 이 개념을 가리킨다.

빈칸별 해설

(가) 특이점(singularity): 인공지능이 인간의 지능을 초월하고, 자신보다 더 뛰어난 인공지능을 설계할 수 있게 되는 시점을 뜻한다. 더 뛰어난 인공지능이 다시 그보다 뛰어난 인공지능을 만드는 과정이 이어지면 지능의 발전이 가속될 수 있다. 따라서 지문의 핵심은 단순히 기술이 변하는 시점이 아니라 **인간 지능의 초월과 인공지능의 자기 개선**이다.

커즈와일의 저서명은 『특이점이 온다』이므로 책 제목의 (가)에도 같은 용어가 들어간다. 2045년은 그가 특이점의 도래 시기로 **예측한 연도**이며, 확정된 도래 시점을 뜻하지 않는다.

세 설명의 (가)에 공통으로 들어가는 개념은 특이점이며, 이는 ④의 용어와 일치한다.

NoteBase

문 2. 비지도 학습으로 분류되는 기법은?

① 랜덤 포레스트(random forest)

② 서포트 벡터 머신(support vector machine)

③ 의사결정 트리(decision tree)

④ 주성분 분석(principle component analysis)

정답 4번

간단한 해설

주성분 분석은 **정답 레이블 없이** 입력 데이터의 분산이 큰 방향을 찾아 차원을 줄이는 기법이므로 비지도 학습에 속한다.

개념 구분

비지도 학습은 정답 레이블 없이 데이터 자체의 구조나 특징을 찾는 방식이다. 주성분 분석(PCA)은 **데이터의 분산을 최대한 보존하는 방향**을 주성분으로 찾고, 일부 주성분에 데이터를 투영하여 차원을 줄인다. 이 과정은 입력 데이터만으로 수행할 수 있으며, 예측할 범주나 목표값을 주어 학습시키지 않는다.

랜덤 포레스트·서포트 벡터 머신·의사결정 트리는 **입력과 목표값의 관계**를 학습하는 대표적인 지도 학습 기법이다. 예를 들어 분류 문제에서는 학습 데이터에 붙은 범주 레이블을 이용해 새로운 데이터의 범주를 예측한다. 이처럼 목표값을 학습에 이용한다는 점에서 주성분 분석과 구별된다.

NoteBase

문 3. 퍼셉트론(perceptron)에 대한 설명으로 옳지 않은 것은?

① 퍼셉트론의 하이퍼 매개 변수(hyper parameter)를 정하는 것은 사람이 해야 한다.

② 단층 퍼셉트론이 해결할 수 있는 문제는 비선형 영역을 포함한다.

③ 퍼셉트론은 다수의 신호를 입력으로 받아 하나의 신호를 출력한다.

④ 단층 퍼셉트론의 구성 요소 중 임곗값은 바이어스(bias)로 대체할 수 있다.

정답 2번

간단한 해설

단층 퍼셉트론의 결정 경계는 직선이나 초평면이므로 **선형 분리가 가능한 문제**를 해결한다. XOR처럼 하나의 직선으로 두 부류를 나눌 수 없는 문제는 해결하지 못한다.

선지별 해설

① O: 하이퍼 매개 변수는 학습률이나 학습 반복 횟수처럼 학습 방법을 정하는 설정값이다. 기본적인 퍼셉트론 학습에서는 사람이 이를 정하고, 학습 알고리즘은 데이터에 따라 가중치와 바이어스를 갱신한다. 즉, **사람이 정하는 학습 설정**과 데이터로 학습하는 매개 변수를 구분한 설명이다.

② X: 단층 퍼셉트론은 입력의 가중합을 임곗값과 비교하므로, 두 부류를 나누는 결정 경계가 2차원에서는 직선, 더 높은 차원에서는 초평면이다. 따라서 '비선형 영역을 포함한다'는 표현은 **'선형 분리가 가능한 문제로 한정된다'**로 고쳐야 한다. 예를 들어 XOR은 두 입력이 다를 때만 1을 출력하는데, 입력 조합을 평면에 표시하면 1인 두 점과 0인 두 점을 하나의 직선으로 분리할 수 없어 단층 퍼셉트론으로 구현할 수 없다.

③ O: 하나의 퍼셉트론은 여러 입력에 각각 가중치를 곱해 더한 뒤, 그 합이 임곗값 이상인지에 따라 **하나의 출력 신호**를 낸다. 기본적인 이진 퍼셉트론에서는 이 출력이 0 또는 1이다.

④ O: 입력의 가중합이 임곗값 이상일 때 1을 출력한다면 그 조건은 $\sum_i w_i x_i \geq \theta$이다. 여기서 입력은 x_i, 가중치는 w_i, 임곗값은 θ이다. **바이어스를 임곗값의 음수**인 $b = -\theta$로 두면 같은 조건을 $\sum_i w_i x_i + b \geq 0$으로 나타낼 수 있다. 즉, 임곗값을 따로 비교하는 방식과 바이어스를 더해 0과 비교하는 방식은 같은 판단을 한다.

NoteBase

문 4. 다음 설명에 해당하는 딥러닝 알고리즘은?

◯ 마이크로소프트 연구팀에서 개발한 이미지 분류를 위한 딥러닝 모델이다.
◯ 레지듀얼 블록(residual block)을 도입하여 스킵 연결(skip connection)을 사용한다.
◯ 레지듀얼 블록의 출력은 입력값과 레지듀얼 함수(residual function)의 결과를 더한 형태로 표현할 수 있다.
◯ 이러한 구조를 통해 심층 신경망의 학습 효율을 높일 수 있다.

① ResNet
② DenseNet
③ MobileNet
④ RCNN

정답 1번

간단한 해설

입력과 잔차 함수의 출력을 **더하는 스킵 연결**은 ResNet의 핵심 구조이다. 원래 입력을 전달하면서 필요한 변화량을 학습하므로 깊은 신경망의 학습을 돕는다.

핵심 풀이

ResNet은 마이크로소프트 연구팀이 개발한 이미지 분류 모델로, 레지듀얼 블록을 사용한다. 입력을 x, 블록 내부에서 학습하는 잔차 함수를 $F(x)$라 하면, 입력과 잔차 출력의 차원이 같을 때 블록의 덧셈 결과는 $y = x + F(x)$로 나타낸다. 핵심은 여러 층을 거친 결과에 **원래 입력을 더한다**는 것이다.

잔차는 입력에 추가로 반영할 변화량이다. 예를 들어 입력을 그대로 유지하는 것이 유리하다면, 잔차 함수가 0에 가까운 값을 내도록 학습하면 된다. 스킵 연결은 역전파 때 기울기가 흐르는 경로도 제공하므로, 층이 깊어질 때의 최적화 어려움을 줄여 학습을 돕는다.

DenseNet도 앞선 층의 정보를 뒤의 층에 전달하지만, 기본적인 밀집 연결에서는 특징을 **더하는 대신 이어 붙인다**. 따라서 지문의 '입력값과 레지듀얼 함수의 결과를 더한다'는 단서는 ResNet을 가리킨다.

NoteBase

문 5. 심층 신경망에서 역전파 알고리즘을 사용할 때 발생할 수 있는 기울기 소실 문제(gradient vanishing problem)를 해결하기 위해 개발된 활성화 함수(activation function)는?

① 시그모이드(sigmoid)

② 계단 함수(step function)

③ 지수 선형 유닛(exponential linear unit)

④ 하이퍼볼릭 탄젠트(hyperbolic tangent)

정답 3번

간단한 해설

지수 선형 유닛(ELU)은 **양수 구간의 기울기가 1**이어서, 활성화 함수의 작은 미분값이 반복해서 곱해지며 기울기가 줄어드는 현상을 완화한다.

핵심 풀이

기울기 소실은 역전파 과정에서 앞쪽 층으로 전달되는 기울기가 매우 작아져 가중치가 거의 갱신되지 않는 현상이다. 역전파에서는 연쇄법칙에 따라 각 층의 가중치와 활성화 함수의 미분값이 곱해지므로, 작은 미분값이 여러 층에 걸쳐 곱해지면 학습에 필요한 신호가 약해질 수 있다.

지수 선형 유닛(ELU): 양수 입력에서는 $f(x) = x$이므로 $f'(x) = 1$이다. 따라서 **양수 구간에서 활성화 함수의 미분으로 기울기가 축소되지 않는다**는 점이 기울기 소실을 완화하는 핵심이다. 음수 구간에는 지수형 곡선을 사용한다. 다만 큰 음수에서는 포화하므로 모든 구간에서 기울기 소실을 완전히 없애는 것은 아니다.

시그모이드·하이퍼볼릭 탄젠트: 입력이 양수나 음수 방향으로 매우 커지면 출력이 일정한 값에 가까워지는 포화 현상이 나타난다. 이때 **미분값이 0에 가까워져** 여러 층을 거치는 역전파에서 기울기 소실을 일으킬 수 있다.

계단 함수: 문턱값을 제외한 구간에서는 **미분값이 0**이고 문턱값에서는 미분할 수 없다. 따라서 활성화 함수의 미분을 이용해 가중치를 갱신하는 일반적인 역전파 학습에 적합하지 않다.

NoteBase

문 6. (가)~(다)의 데이터 종류와 이에 대한 예를 바르게 연결한 것은?

(가) 정형 데이터(structured data)
(나) 비정형 데이터(unstructured data)
(다) 반정형 데이터(semi-structured data)

A. SNS B. HTML C. VIDEO
D. EXCEL E. JSON F. RDB
G. XML

① (가): D, F, (나): A, C, (다): B, E, G
② (가): D, F, G, (나): A, C, (다): B, E
③ (가): D, F, (나): A, C, E, (다): B, G
④ (가): F, (나): B, E, G, (다): A, C, D

정답 1번

간단한 해설

HTML·JSON·XML은 **태그나 키로 데이터의 구조를 나타내는 반정형 데이터**이다. 구조를 표현하는 규칙은 있지만, 관계형 데이터베이스처럼 모든 자료를 고정된 표 구조에 맞추는 것은 아니다.

NoteBase

구성 요소별 해설

A. SNS → (나) 비정형 데이터: SNS의 게시글·사진 같은 콘텐츠는 **내용이 일정한 행·열이나 고정된 필드로 정리되어 있지 않다**. 이 문항에서는 이러한 콘텐츠를 기준으로 비정형 데이터로 분류한다.

B. HTML → (다) 반정형 데이터: 태그로 제목·문단 등의 요소와 그 관계를 나타낸다. **태그가 내부 구조를 표시**하지만, 문서마다 요소의 종류와 배치가 달라질 수 있어 고정된 표 구조의 정형 데이터와 구별된다.

C. VIDEO → (나) 비정형 데이터: 영상에 담긴 장면이나 움직임의 내용은 일정한 필드의 표로 바로 표현되지 않는다. 영상 파일에 저장 형식이 있다는 사실과 **내용이 정형화되어 있다는 것은 다르다**.

D. EXCEL → (가) 정형 데이터: 엑셀에서 행·열의 표로 정리한 데이터는 각 열이 이름·점수 같은 정해진 항목을 나타내고, 각 행에 그 항목에 맞는 값을 기록한다. 이처럼 **일정한 표 구조에 따라 정리한 데이터**를 정형 데이터로 분류한다.

E. JSON → (다) 반정형 데이터: 키와 값의 쌍, 객체와 배열로 데이터를 표현한다. **키와 중첩 관계로 구조를 나타내면서도** 객체마다 포함하는 항목이 달라질 수 있으므로 반정형 데이터에 해당한다.

F. RDB → (가) 정형 데이터: 관계형 데이터베이스는 데이터를 행과 열로 이루어진 테이블에 저장한다. 각 열의 이름과 자료형 등 **미리 정의한 스키마**에 맞추어 데이터를 관리하므로 정형 데이터에 해당한다.

G. XML → (다) 반정형 데이터: 태그로 데이터의 의미와 계층 관계를 표현한다. 표의 고정된 열에 맞추기보다 **태그와 계층 구조로 항목을 구분**하므로 반정형 데이터로 분류한다.

분류 결과는 (가) D·F, (나) A·C, (다) B·E·G이며, ①의 연결과 일치한다.

NoteBase

문 7. 데이터 1,000개를 대상으로 5-폴드 교차검증을 수행할 때, 한 번의 반복에서 사용되는 훈련 데이터와 검증 데이터의 개수는? (단, 각 폴드의 크기는 같다고 가정한다)

① 훈련 데이터 개수: 200, 검증 데이터 개수: 800

② 훈련 데이터 개수: 500, 검증 데이터 개수: 500

③ 훈련 데이터 개수: 800, 검증 데이터 개수: 200

④ 훈련 데이터 개수: 1,000, 검증 데이터 개수: 1,000

정답 3번

간단한 해설

5-폴드 교차검증은 한 번의 반복에서 **1개 폴드를 검증에, 나머지 4개 폴드를 훈련에** 사용한다. 1,000개를 같은 크기로 나누면 폴드당 200개이므로 훈련 데이터는 800개, 검증 데이터는 200개이다.

계산 과정

데이터 분할: 5-폴드 교차검증에서는 전체 데이터를 크기가 같은 5개 폴드로 나눈다. 한 번의 반복에서는 **1개 폴드로 검증하고 나머지 4개 폴드로 훈련**하며, 다음 반복에서는 검증에 사용할 폴드를 바꾼다.

검증 데이터 개수: 전체 데이터가 1,000개이므로 한 폴드의 데이터 개수는 $\frac{1,000}{5} = 200$개이다. 검증에는 한 폴드만 사용하므로 검증 데이터는 200개이다.

훈련 데이터 개수: 검증에 사용한 한 폴드를 제외한 네 폴드로 모델을 학습한다. 따라서 훈련 데이터는 1,000 − 200 = 800개이다. 이는 **한 번의 반복에서 사용하는 개수**이며, 여러 반복의 데이터 개수를 합산하지 않는다.

훈련 데이터 800개, 검증 데이터 200개이므로 ③과 일치한다.

NoteBase

문 8. 인공지능 생성모델에 해당하지 않는 것은?

① Diffusion Model

② DQN(Deep Q-Network)

③ GANs(Generative Adversarial Networks)

④ VAEs(Variational Autoencoders)

정답 2번

간단한 해설

DQN은 행동의 기대 누적 보상을 학습해 **행동을 선택하는 강화학습 알고리즘**이다. 학습한 데이터 분포를 바탕으로 새로운 표본을 만드는 생성모델에 해당하지 않는다.

개념 구분

생성모델은 학습 데이터의 분포를 학습하여 **새로운 표본을 생성**하는 모델이다. Diffusion Model, GANs, VAEs가 이에 해당하며, 학습한 이미지의 특성을 반영한 새 이미지를 만드는 것이 대표적인 활용 예이다.

DQN(Deep Q-Network)은 신경망으로 행동가치를 근사하는 강화학습 알고리즘이다. 행동가치는 주어진 상태에서 특정 행동을 한 뒤 기대할 수 있는 누적 보상을 나타낸다. 이를 이용해 **높은 보상을 얻을 행동을 선택**하므로, 새로운 데이터 표본을 생성하는 생성모델과 구별된다.

NoteBase

문 9. 딥러닝 기반의 인공지능 모델에서 설명 가능성(explainability)이 부족한 이유로 옳지 않은 것은?

① 학습 데이터의 부족

② 비선형적인 데이터 처리

③ 추상적 특징의 학습

④ 모델의 복잡성

정답 1번

간단한 해설

학습 데이터의 부족은 주로 **예측 성능과 일반화**에 영향을 미치는 문제이다. 딥러닝의 판단 근거를 설명하기 어려운 핵심 이유는 여러 층의 복잡한 비선형 연산과 사람이 곧바로 의미를 파악하기 어려운 내부 특징에 있다.

개념 구분

설명 가능성은 모델이 어떤 근거와 과정을 통해 특정 출력을 냈는지 사람이 이해할 수 있는 정도이다. **예측이 정확한가**와 **판단 근거를 이해할 수 있는가**는 구별해야 한다.

학습 데이터의 부족: 데이터가 부족하면 학습한 사례에 치우치는 과적합이 발생하거나 새로운 데이터에 대한 예측 성능이 낮아질 수 있다. 이는 학습과 일반화의 문제이며, 딥러닝의 내부 판단 과정을 설명하기 어려운 구조적 이유와는 다르다. 데이터를 많이 확보해도 복잡한 모델의 판단 근거가 저절로 이해되는 것은 아니다.

비선형적인 데이터 처리와 모델의 복잡성: 딥러닝은 여러 층에서 가중치와 비선형 변환을 결합한다. 따라서 입력의 변화가 출력에 미치는 영향을 하나의 단순한 규칙으로 표현하기 어렵고, 많은 연결과 매개변수가 함께 작용하여 **개별 요소의 기여를 추적하기 어렵다**.

추상적 특징의 학습: 내부에서 학습한 특징은 사람이 사용하는 개념과 일대일로 대응하지 않을 수 있다. 예를 들어 이미지 분류에서 여러 내부 활성값이 함께 판단에 기여하면, 이를 사람이 이해하는 하나의 특징으로 이름 붙이기 어렵다. 이러한 **내부 표현과 사람의 개념 사이의 간극**도 설명을 어렵게 한다.

NoteBase

문 10. 2017년 발표된 아실로마 인공지능 원칙(Asilomar AI principles)에 대한 설명으로 옳지 않은 것은?

① 인공지능 개발의 목적, 윤리, 가치 등에 대해 23가지 원칙을 제시했다.

② 인공지능 시스템은 반드시 인간의 직접적인 통제 없이 자율적으로 작동해야 한다.

③ 연구 목표는 방향성이 없는 지능이 아니라 인간에게 유용한 지능 개발이어야 한다.

④ 인공지능 시스템의 실패 원인을 이해할 수 있도록 투명성과 설명 가능성이 확보되어야 한다.

정답 2번

간단한 해설

②의 '반드시 인간의 직접적인 통제 없이'라는 표현이 잘못되었다. 아실로마 원칙은 **AI에 의사결정을 위임할지와 그 방식을 인간이 선택**하여, 인간이 정한 목표를 달성하도록 해야 한다고 본다.

선지별 해설

① O: 2017년 아실로마 회의에서 마련된 원칙은 **총 23개**이며, 연구 관련 쟁점, 윤리와 가치, 장기적 쟁점의 세 영역으로 구성된다. 인공지능 연구의 목표뿐 아니라 안전성, 투명성, 인간의 가치와 통제 등 개발과 활용의 기준을 제시한다.

② X: 제16원칙인 '인간의 통제'는 인간이 정한 목표를 달성하기 위해 **의사결정의 위임 여부와 방식을 인간이 선택**해야 한다고 규정한다. AI에 자율적인 판단을 맡길 수는 있지만, 인간의 직접적인 통제를 반드시 배제해야 한다는 뜻은 아니다. 선지의 표현은 인간이 위임 여부와 방식을 결정해야 한다는 내용으로 고쳐야 한다.

③ O: 제1원칙인 '연구 목표'는 방향성이 없는 지능이 아니라 **인간에게 유익한 지능**을 만드는 것을 목표로 삼는다. 지능의 능력을 높이는 것만으로는 충분하지 않으며, 그 능력이 인간에게 도움이 되도록 연구 방향을 정해야 한다는 의미이다.

④ O: 제7원칙인 '실패의 투명성'은 AI 시스템이 해를 끼쳤을 때 **그 원인을 확인할 수 있어야 한다**고 규정한다. 피해가 발생했다는 결과만 확인하는 데 그치지 않고, 왜 그런 결과가 나왔는지 이해할 수 있어야 한다는 점에서 투명성과 설명 가능성을 요구한다.

NoteBase

문 11. Q−러닝 알고리즘에서 다음 조건을 적용한 Q의 업데이트 값은?

◯ 현재 Q−값: 5
◯ 보상: 10
◯ 학습률: 0.5
◯ 할인계수: 0.9
◯ 다음 상태의 Q−값 중 최댓값: 8

① 6.0
② 8.6
③ 11.1
④ 17.2

정답 3번

간단한 해설

Q값은 보상과 할인된 다음 상태의 최댓값으로 만든 **목표값에 학습률만큼 가까워지도록** 갱신한다. 목표값 17.2와 현재 값 5의 차이 중 0.5만 반영하므로 업데이트 값은 11.1이다.

계산 과정

Q값은 현재 상태에서 특정 행동을 했을 때 얻을 것으로 예상하는 누적 할인 보상의 값이다. 현재 Q값을 Q, 보상을 r, 학습률을 α, 할인계수를 γ, 다음 상태의 Q값 중 최댓값을 M이라 두면, 주어진 값은 각각 5, 10, 0.5, 0.9, 8이다. Q−러닝은 **다음 상태의 최대 Q값**을 이용해 목표값 $T = r + \gamma M$을 만들고, Q_새 값=Q+α(T-Q)로 갱신한다.

목표값 계산: 다음 상태의 최대 Q값 8에 할인계수 0.9를 적용하면 $0.9 \times 8 = 7.2$이다. 여기에 즉시 받는 보상 10을 더하므로 목표값은 $10 + 7.2 = 17.2$이다. 이는 즉시 보상과 할인된 미래 가치의 합이다.

갱신량 계산: 목표값과 현재 Q값의 차이는 $17.2 - 5 = 12.2$이다. 학습률 0.5는 **이 차이의 절반만 반영**한다는 뜻이므로, 현재 Q값에 더할 양은 $0.5 \times 12.2 = 6.1$이다.

최종 업데이트: 현재 Q값 5에 갱신량 6.1을 더하면 $5 + 6.1 = 11.1$이다. 이 값은 ③의 11.1과 일치한다.

NoteBase

문 12. 입력 채널 수가 c, 출력 채널 수가 k, 커널 크기가 p × p인 2D 합성곱층(convolutional layer)을 사용할 때, 학습해야 할 파라미터인 가중치의 총 개수는? (단, 바이어스 값은 없다고 가정한다)

① $c \times k$

② $p^2 \times k$

③ $p^2 \times c + k$

④ $p^2 \times c \times k$

정답 4번

간단한 해설

출력 채널 하나를 만드는 필터는 **입력 채널 전체**를 처리하므로 가중치가 $p^2 \cdot c$개 필요하다. 이러한 필터가 k개이므로 바이어스를 제외한 전체 가중치는 $p^2 \cdot c \cdot k$개이다.

계산 과정

필터 하나의 가중치 수: 입력 채널 수는 c, 커널의 공간 크기는 p × p이다. 출력 채널 하나를 만드는 필터는 **c개의 입력 채널 모두에 대응하는 가중치**를 갖는다. 입력 채널마다 p^2개의 가중치가 필요하므로 필터 하나의 가중치는 $p^2 \cdot c$개이다.

전체 필터의 가중치 수: 출력 채널 수가 k이므로 서로 다른 필터가 k개 필요하다. 따라서 전체 가중치 수는 $(p^2 \cdot c) \cdot k = p^2 \cdot c \cdot k$개이다.

계수 조건: 합성곱에서는 필터가 입력의 여러 위치를 이동하면서 **같은 가중치를 공유**하므로 위치마다 가중치를 새로 세지 않는다. 또한 문제에서 바이어스가 없다고 했으므로 추가로 더할 파라미터도 없다.

따라서 학습할 가중치의 총 개수는 $p^2 \cdot c \cdot k$개이며, ④의 식에 해당한다.

NoteBase

문 13. 기호주의(symbolic) AI와 연결주의(connectionism) AI에 대한 설명으로 옳지 않은 것은?

① 기호주의 AI는 이미지, 음성 인식을 위한 딥러닝 모델을 개발하기 위해 고안되었고, 연결주의 AI는 전문가 시스템과 같은 지식 기반 시스템을 주로 구축한다.

② 기호주의 AI는 해석 가능성이 크다는 장점이 있으나 환경 노이즈에 약하며, 연결주의 AI는 해석 가능성은 작지만 노이즈 강인성과 범용 근사 능력이 높다.

③ 기호주의 AI는 논리 연산에 기반해 이산적(discrete) 문제를 해결하며, 연결주의 AI는 주로 미분 가능 함수를 이용한 학습을 통해 문제를 해결한다.

④ 기호주의 AI는 지식을 명시적인 생성 규칙, 논리식 등 규칙으로 표현하고, 연결주의 AI는 분산된 가중치로 지식을 암묵적으로 저장한다.

정답 1번

간단한 해설

①은 두 접근법의 대표적인 응용을 서로 바꾸어 설명했다. 명시적인 지식과 규칙으로 추론하는 **전문가 시스템은 기호주의 AI**에, 신경망을 학습시키는 **이미지·음성 인식 딥러닝은 연결주의 AI**에 해당한다.

NoteBase

선지별 해설

① X: 기호주의 AI는 기호로 표현한 지식과 규칙을 조작하여 추론하는 접근법이고, 연결주의 AI는 연결된 인공 뉴런의 가중치를 학습하는 접근법이다. 따라서 **전문가 시스템은 기호주의 AI의 대표적인 응용**이며, 이미지·음성 인식에 쓰이는 딥러닝은 연결주의 AI의 대표적인 응용이다. 선지는 두 접근법의 응용 분야를 서로 바꾸어 설명했다.

② O: 기호주의 AI는 적용한 규칙과 추론 과정을 확인할 수 있어 판단 이유를 설명하기 쉽다. 반면 입력이 불완전하거나 잡음 때문에 규칙의 조건을 충족하지 못하면 추론에 어려움을 겪을 수 있다. 연결주의 AI는 여러 뉴런과 가중치에 걸쳐 패턴을 학습하므로 내부 판단 과정을 직접 해석하기 어렵지만, 학습한 범위의 입력 변동이나 잡음에는 상대적으로 강인할 수 있다. 또한 적절한 활성화 함수와 충분한 은닉 뉴런을 갖춘 신경망은 유계의 닫힌 입력 영역에서 연속함수를 원하는 정확도로 근사할 수 있다. 이 **범용 근사 능력은 함수의 표현 능력**을 뜻하며, 모든 잡음에 강하거나 학습이 언제나 성공한다는 보장은 아니다.

③ O: 기호주의 AI는 명제·술어 등의 기호와 논리 규칙을 이용해 사실의 참·거짓이나 규칙의 적용 여부 같은 이산적인 관계를 다룬다. 연결주의 AI는 주로 신경망의 출력과 목표 사이의 오차를 줄이도록 가중치를 조정한다. 이때 **미분을 이용해 오차가 줄어드는 방향을 구하는 학습**이 핵심이며, 역전파와 경사하강법이 대표적으로 사용된다.

④ O: 기호주의 AI에서는 '조건이 성립하면 결론을 도출한다'는 생성 규칙이나 논리식으로 지식을 명시한다. 예를 들어 'A이면 B이다'라는 규칙과 A라는 사실로부터 B를 추론할 수 있다. 연결주의 AI에서는 학습으로 조정된 **여러 연결 가중치에 지식이 분산되어 저장**된다. 따라서 개별 가중치 하나가 사람이 읽을 수 있는 규칙 하나와 일대일로 대응하는 것은 아니다.

NoteBase

문 14. 트랜스포머(transformer) 모델에 대한 설명으로 옳지 않은 것은?

① 서열 데이터(sequence data)를 순차적으로 처리하는 RNN 모델과는 다르게 병렬처리가 가능하다.

② 위치 임베딩(positional embedding)은 입력 토큰들의 위치 정보를 모델에 제공하기 위해 사용된다.

③ 자기 주목(self-attention) 메커니즘은 특정 위치가 입력의 어떤 위치에 주목해야 하는지를 가중치 형태로 수치화하여 표현한다.

④ 주목 점수(attention score)는 query 벡터와 value 벡터의 내적을 계산한 후 소프트맥스(softmax) 함수를 적용하여 계산된다.

정답 4번

간단한 해설

④는 주목의 관련도를 구할 때 사용하는 벡터를 잘못 제시했다. **query와 key의 내적**을 바탕으로 소프트맥스 가중치를 구하며, value는 그 가중치로 합산할 정보이다.

선지별 해설

① O: RNN은 이전 위치의 은닉 상태를 계산해야 다음 위치를 처리할 수 있다. 트랜스포머는 이러한 순환 의존성이 없어, 학습 시 **여러 토큰 위치의 연산을 병렬로 수행**할 수 있다. 다만 자기회귀 방식으로 문장을 생성할 때는 앞서 생성한 토큰을 다음 입력에 사용하므로 생성 단계까지 모두 병렬화되는 것은 아니다.

② O: 자기 주목 연산만으로는 토큰의 순서를 구별할 정보가 충분하지 않다. 위치 임베딩은 **각 토큰이 놓인 위치 정보**를 입력 표현에 반영하여, 같은 토큰이라도 문장 안의 위치에 따라 다르게 처리할 수 있게 한다.

③ O: 자기 주목은 같은 입력 서열에서 query, key, value를 만든다. 특정 위치의 query와 각 위치의 key를 비교하고 소프트맥스를 적용하여, **각 위치에 얼마나 주목할지를 나타내는 가중치**를 구한다. 이 가중치로 각 위치의 value를 가중합하여 문맥을 반영한 표현을 만든다.

④ X: 관련도를 계산하는 내적의 대상은 **query와 value가 아니라 query와 key**이다. 트랜스포머의 스케일 내적 주목에서는 이 내적을 key 벡터 차원의 제곱근으로 나누어 점수를 조정한 뒤, 소프트맥스를 적용해 주목 가중치를 구한다. value는 관련도를 비교하는 대상이 아니라, 구한 가중치에 따라 합산되는 정보이다.

NoteBase

문 15. 연합학습(federated learning)의 학습 과정을 순서대로 바르게 나열한 것은?

(가) 각 디바이스는 로컬 데이터로 모델을 학습시켜 파라미터를 업데이트한다.
(나) 중앙 서버는 여러 디바이스로부터 받은 파라미터를 집계하여 글로벌 모델을 갱신한다.
(다) 중앙 서버는 초기 글로벌 모델을 각 디바이스로 배포한다.
(라) 각 디바이스는 업데이트된 파라미터만 중앙 서버로 전송한다.

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

정답 3번

간단한 해설

각 디바이스가 학습하려면 먼저 **초기 글로벌 모델을 받아야** 한다. 중앙 서버가 학습 결과를 집계하려면 디바이스의 **로컬 학습과 파라미터 전송이 먼저** 이루어져야 한다.

올바른 순서

(다) → (가) → (라) → (나)

(다) 초기 글로벌 모델 배포: 중앙 서버가 각 디바이스에 초기 모델을 보낸다. 각 디바이스가 **같은 글로벌 모델을 출발점으로** 로컬 학습을 수행할 수 있도록 모델을 먼저 제공하는 단계이다.

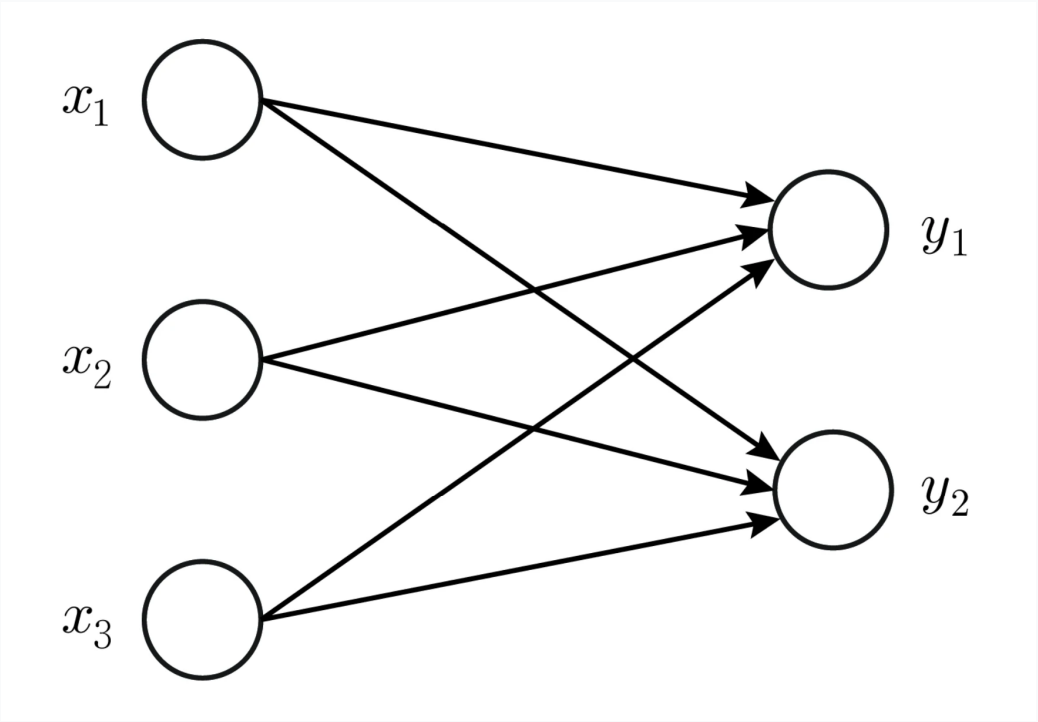
(가) 로컬 학습: 각 디바이스는 전달받은 모델을 자신의 로컬 데이터로 학습한다. 이 과정에서 가중치처럼 학습에 따라 바뀌는 모델 내부 값인 파라미터를 업데이트하므로, 서버로 보낼 학습 결과는 이 단계 이후에 마련된다.

(라) 업데이트된 파라미터 전송: 각 디바이스는 로컬 학습으로 바뀐 파라미터를 중앙 서버에 보낸다. 이때 전송 대상은 **로컬 데이터 자체가 아니라 업데이트된 파라미터**이며, 서버가 여러 디바이스의 학습 결과를 모으기 위해 필요한 단계이다.

(나) 글로벌 모델 갱신: 중앙 서버는 여러 디바이스에서 받은 파라미터를 집계하여 글로벌 모델을 갱신한다. 집계할 학습 결과가 서버에 도착해야 하므로, 파라미터 전송이 글로벌 모델 갱신보다 앞선다.

NoteBase

문 16. 다음 퍼셉트론 신경망에서 학습이 가능한 데이터 세트는?



①

x_1	x_2	x_3	y_1	y_2
0	0	0	0	0
0	0	1	0	1
1	1	0	1	0
1	1	1	1	1

②

x_1	x_2	x_3	y_1	y_2
0	0	0	0	0
0	0	1	0	1
1	1	0	1	1
1	1	1	1	0

③

x_1	x_2	x_3	y_1	y_2
0	0	0	0	0
0	0	1	1	0
1	1	0	1	1
1	1	1	0	1

NoteBase

④

x_1	x_2	x_3	y_1	y_2
0	0	0	0	1
0	1	0	1	1
1	0	1	1	0
1	1	1	0	0

정답 1번

간단한 해설

은닉층이 없는 퍼셉트론은 각 출력의 0과 1을 **하나의 선형 경계로 분리**할 수 있어야 한다. 제시된 데이터에서 $y_1 = x_1, y_2 = x_3$인 관계는 각 입력값에 임계값을 적용하는 것만으로 구현할 수 있다.

공통 해설

그림은 세 입력이 두 출력에 직접 연결된 단층 퍼셉트론이다. 각 출력 뉴런은 입력의 가중합을 임계값과 비교하여 0 또는 1을 내므로, 출력별로 0인 입력들과 1인 입력들이 하나의 평면으로 나뉘는 '선형 분리'가 가능해야 한다. 두 출력은 서로 다른 가중치를 사용할 수 있지만, **두 출력 모두 선형 분리가 가능**해야 데이터 세트 전체를 학습할 수 있다.

출력이 $y_1 = x_1, y_2 = x_3$인 데이터에서는 첫 번째 출력은 첫 번째 입력을, 두 번째 출력은 세 번째 입력을 그대로 따르면 된다. 예를 들어 각 출력에서 해당 입력의 가중치를 1, 나머지 가중치를 0으로 두고 임계값을 0.5로 잡으면 표의 네 행을 모두 만족한다. 즉, 각 출력에 필요한 선형 경계가 존재한다.

나머지 데이터에는 한 출력이 XOR 관계인 경우가 포함되어 있다. ②와 ③에서는 항상 $x_2 = x_1$이므로 서로 독립적으로 변하는 입력은 x_1, x_3이고, ④에서는 항상 $x_3 = x_1$이므로 x_1, x_2를 보면 된다. 이 두 입력의 조합이 (0, 0), (0, 1), (1, 0), (1, 1)로 바뀔 때, ②의 y_2, ③과 ④의 y_1은 모두 0, 1, 1, 0이다. 이는 두 입력이 다를 때만 1인 XOR로, 정사각형의 서로 마주 보는 꼭짓점끼리 같은 출력이 된다. 두 부류를 잇는 대각선이 중앙에서 교차하므로 **하나의 직선으로 두 부류를 분리할 수 없다**. 중복된 입력을 하나 더 사용해도 이 관계는 선형 분리되지 않으므로 해당 데이터 세트는 그림의 신경망으로 학습할 수 없다.

문 17. 어떤 다중 분류 신경망의 출력층 로짓(logit) 벡터가 $[2, 1, 0]$이다. 소프트맥스 함수를 적용하여 입력값들에 대한 확률분포를 구할 때, 가장 큰 확률값은? (단, $e \approx 2.7$, $e^2 \approx 7.4$이고, 소수점 셋째 자리에서 반올림한다)

① 0.25

② 0.33

③ 0.67

④ 0.89

정답 3번

간단한 해설

소프트맥스의 최대 확률은 가장 큰 로짓 2의 지수값 7.4를 **모든 지수값의 합 11.1**로 나누어 구한다. 소수점 셋째 자리에서 반올림하면 $\frac{7.4}{11.1} \approx 0.67$이다.

계산 과정

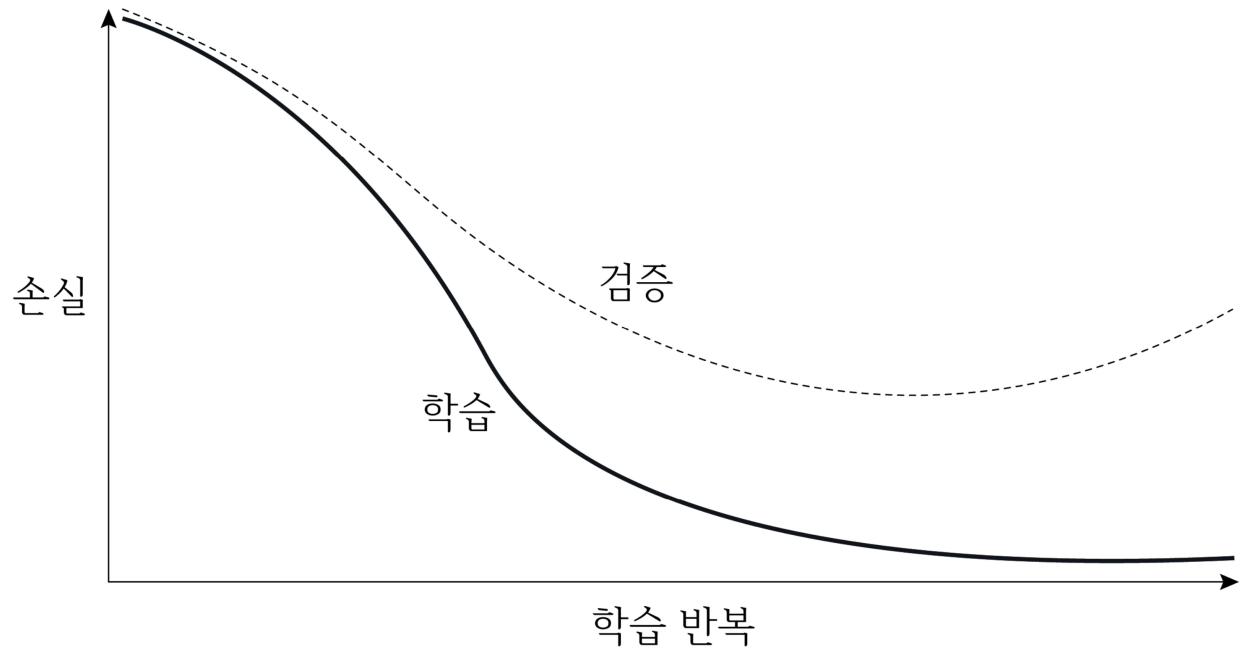
소프트맥스는 로짓의 지수값을 전체 지수값의 합으로 나누어, 확률의 합이 1이 되도록 변환한다. 로짓을 z_i, 해당 클래스의 확률을 p_i라 하면 $p_i = \frac{e^{z_i}}{\sum_{j=1}^{3} e^{z_j}}$이다. 모든 클래스의 분모가 같고 지수함수는 증가하므로, **가장 큰 로짓 2에 대응하는 확률**이 가장 크다.

주어진 로짓 벡터는 $[2, 1, 0]$이다. 각 지수값은 $e^2 \approx 7.4$, $e^1 \approx 2.7$, $e^0 = 1$이므로 분모는 $7.4 + 2.7 + 1 = 11.1$이다. 이 값은 **세 클래스의 지수값을 모두 더한 값**이다.

따라서 최대 확률은 $\frac{e^2}{e^2 + e^1 + e^0} \approx \frac{7.4}{11.1} \approx 0.6667$이다. 소수점 셋째 자리의 숫자가 6이므로 둘째 자리까지 반올림하면 0.67이며, ③의 값과 일치한다.

NoteBase

문 18. 신경망에서 학습의 진행에 따른 학습 데이터의 손실값과 검증 데이터의 손실값이 다음과 같을 때, 이에 대한 설명으로 옳지 않은 것은?



① 학습 데이터를 추가하여 검증 손실값을 학습 손실값에 근접시킬 수 있다.

② 신경망 학습 모델이 학습 데이터 세트의 패턴을 적절히 학습하지 못했다는 의미이다.

③ 신경망의 은닉층 개수, 뉴런의 개수 등 학습 매개 변수가 과도하게 많으면 발생할 수 있다.

④ 신경망에서 검증 손실값이 다시 상승하는 현상을 완화하기 위하여 가중치 규제 방법이 사용될 수 있다.

정답 2번

간단한 해설

학습 손실은 낮지만 검증 손실이 다시 커지는 모습은 **과대적합**을 나타낸다. ②처럼 학습 데이터의 패턴조차 충분히 익히지 못한 상태는 **과소적합**이며, 일반적으로 학습 손실도 높게 나타난다.

NoteBase

선지별 해설

가로축은 학습 반복, 세로축은 손실이다. 실선인 학습 손실은 낮아져 낮은 수준을 유지하지만, 점선인 검증 손실은 감소하다가 다시 상승한다. 이는 학습 데이터에는 잘 맞지만 학습에 직접 사용하지 않은 데이터에는 예측 성능이 나빠지는 과대적합의 전형적인 모습이다.

① O: 모집단의 특성을 잘 반영하는 **학습 데이터를 충분히 추가**하면 일부 표본의 우연한 특징에 치우쳐 학습하는 것을 줄일 수 있다. 이에 따라 새로운 데이터에 대한 일반화 성능이 좋아져 검증 손실과 학습 손실의 차이가 줄어들 수 있다. 단순히 데이터 수를 늘리는 것만으로 항상 개선되는 것은 아니다.

② X: 학습 데이터의 패턴을 충분히 익히지 못한 상태는 **과소적합**이다. 이때는 학습 데이터에서도 오차가 커 학습 손실이 높게 나타난다. 그림은 학습 손실은 낮지만 검증 손실이 다시 증가하므로, 학습 데이터의 세부 특징이나 잡음에까지 지나치게 맞춘 과대적합으로 설명해야 한다.

③ O: 은닉층이나 뉴런을 늘리면 보통 학습할 가중치 등의 수가 증가하여 더 복잡한 관계를 표현할 수 있다. 모델의 **표현 능력이 학습 데이터의 양과 다양성에 비해 지나치게 크면** 일반적인 패턴뿐 아니라 개별 표본의 잡음까지 학습하여 과대적합이 발생할 수 있다.

④ O: 가중치 규제는 학습 목적함수에 **큰 가중치에 대한 벌점**을 더하여 가중치가 과도하게 커지는 것을 억제하는 방법이다. 학습 표본에 지나치게 민감한 모델이 되는 것을 줄여 일반화 성능을 높일 수 있으므로, 과대적합에 따른 검증 손실의 재상승을 완화하는 데 사용할 수 있다.

NoteBase

문 19. 활성화 함수에 대한 설명으로 옳지 않은 것은?

① 활성화 함수는 인공신경망에서 각 뉴런이 입력을 받아 출력을 계산하는 함수이다.

② 시그모이드 함수는 S자형 함수로 로지스틱(logistic) 함수가 해당된다.

③ 소프트맥스 함수는 입력받은 값이 0⊠1 사이의 값으로 출력되도록 정규화하여 출력의 총합이 1이 되는 특성을 갖는다.

④ 리키렐루(Leaky ReLU) 함수는 입력값이 음수일 경우에 0을 출력하고, 그 외의 경우는 입력값을 그대로 출력한다.

정답 4번

간단한 해설

④의 설명은 **일반 ReLU**의 특성이다. Leaky ReLU는 음수 입력을 0으로 만들지 않고 **작은 양의 기울기를 적용한 음수 값**으로 출력한다.

선지별 해설

① O: 인공신경망의 뉴런은 입력값에 가중치를 곱해 합산하고 편향을 더한 값을 활성화 함수에 넣어 출력을 구한다. 활성화 함수는 이 값을 다음 층으로 전달할 출력으로 변환하며, 비선형 활성화 함수를 사용하면 신경망이 **비선형 관계**를 학습할 수 있다.

② O: 시그모이드 함수는 S자형 곡선을 갖는 함수를 말하며, 로지스틱 함수가 대표적인 예이다. 로지스틱 함수는 입력이 작아질수록 출력이 0에 가까워지고, 커질수록 1에 가까워지며, 입력이 0일 때는 0.5를 출력한다.

③ O: 소프트맥스 함수는 여러 입력값 각각에 지수함수를 적용한 뒤, 그 값을 전체 지수값의 합으로 나눈다. 지수값은 모두 양수이므로 각 출력은 0과 1 사이가 되고, 같은 합으로 나누어 정규화하므로 **전체 출력의 합은 1**이 된다. 이 특성 때문에 다중 클래스 분류에서 클래스별 확률을 나타내는 데 사용한다.

④ X: 음수 입력에서 0을 출력하고 0 이상의 입력을 그대로 출력하는 것은 일반 ReLU이다. Leaky ReLU는 음수 입력에 작은 양의 계수 α를 곱한 αx를 출력하고, 0 이상의 입력은 그대로 출력한다. 보통 $0 < \alpha < 1$로 두어 **음수 구간에서도 기울기를 0이 아닌 작은 값으로 유지**한다. 예를 들어 계수가 0.01이고 입력이 -2이면, 일반 ReLU의 출력은 0이지만 Leaky ReLU의 출력은 -0.02이다.

NoteBase

문 20. 다음 조건에서 어떤 검사자가 진단키트를 한 번 사용하여 양성 반응을 보였을 때, 그 사람이 실제로 A 암에 걸려 있을 확률은?

◯ 10,000명 중 5명의 확률로 A 암이 발생한다.
◯ A 암이 있는 경우 0.99의 확률로 진단키트에 양성 반응이 나오고, A 암이 없는 경우에는 0.95의 확률로 진단키트에 음성 반응이 나온다.

① 약 0.01
② 약 0.05
③ 약 0.95
④ 약 0.99

정답 1번

간단한 해설

구해야 하는 값은 **양성자 중 실제 암 환자의 비율**이다. 암 발생 확률이 매우 낮아 암이 없는데도 양성으로 나오는 경우가 더 많으므로, 이 비율은 약 0.0098, 즉 약 1%이다.

NoteBase

계산 과정

1. 사건과 주어진 확률: A를 암에 걸린 사건, B를 검사에서 양성 반응이 나온 사건이라 하자. 구할 값은 $P(A \mid B)$이다. 암 발생 확률은 $P(A) = \frac{5}{10{,}000} = 0.0005$이고, 암이 없을 확률은 $P(A^c) = 1 - 0.0005 = 0.9995$이다.

2. 실제 환자가 양성일 확률: 암 환자에게서 양성이 나올 확률은 $P(B \mid A) = 0.99$이다. 따라서 암에 걸려 있으면서 양성인 확률은 $P(A \cap B) = 0.0005 \cdot 0.99 = 0.000495$이다. 0.99는 **암 환자 안에서의 양성 확률**이므로, 양성이 나온 사람의 암 확률과 구별해야 한다.

3. 암이 없는 사람이 양성일 확률: 암이 없을 때 음성 확률이 0.95이므로, 양성 확률은 $P(B \mid A^c) = 1 - 0.95 = 0.05$이다. 따라서 암이 없으면서 양성인 확률은 $P(A^c \cap B) = 0.9995 \cdot 0.05 = 0.049975$이다. 이는 **암이 없는데도 양성으로 판정되는 경우**이며, 실제 환자가 양성인 확률보다 크다.

4. 전체 양성 확률: 양성 반응은 암이 있는 경우와 없는 경우로 나뉘고 두 경우는 겹치지 않는다. 따라서 두 확률을 더하면 $P(B) = 0.000495 + 0.049975 = 0.05047$이다. 양성이 나왔다는 조건에서 확률을 구하므로, **전체 양성 확률을 분모**로 삼는다.

5. 양성일 때 실제 암에 걸려 있을 확률: 베이즈 정리에 따라 $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$이다. 앞에서 구한 값을 대입하면 $P(A \mid B) = \frac{0.000495}{0.05047} \approx 0.0098$이다. 따라서 실제 암에 걸려 있을 확률은 약 0.01이며, ①의 값과 일치한다.

NoteBase

문 21. (가) ~ (다)의 효율적 파인튜닝(Parameter-Efficient Fine-tuning) 기법과 이에 대한 설명을 바르게 연결한 것은? (단, 각 기법에서는 사전 학습된 모델의 기존 가중치를 고정(freeze)한다)

(가) LoRA
(나) Adapter
(다) Prefix Tuning
A. 가중치의 변화량을 저랭크 행렬로 표현하여 이를 별도로 학습하는 기법
B. 각 층의 어텐션 계산에 사용할 학습 가능한 접두 벡터를 추가하여 새로운 작업에 적응시키는 기법
C. 모델 내부에 작은 추가 모듈을 삽입하여 새로운 작업에 필요한 변화만 별도로 학습하는 기법

① (가): A, (나): B, (다): C
② (가): A, (나): C, (다): B
③ (가): B, (나): A, (다): C
④ (가): C, (나): A, (다): B

정답 2번

간단한 해설

세 기법은 모두 기존 가중치를 고정하지만, **새로 학습하는 대상**이 다르다. Adapter는 모델 안에 삽입한 작은 모듈을 학습하고, Prefix Tuning은 어텐션이 참조할 접두 벡터를 학습한다.

NoteBase

대응 관계 해설

문제의 조건에서는 세 기법 모두 사전 학습된 모델의 기존 가중치를 갱신하지 않는다. 대신 작업에 필요한 소수의 매개변수를 별도로 학습하므로, 무엇을 추가하고 학습하는지로 구별한다.

(가) LoRA → A: 기존 가중치 자체를 다시 학습하는 대신, **가중치 변화량을 저랭크 행렬의 곱으로 표현**하고 그 행렬들을 학습한다. 큰 가중치 행렬 전체보다 적은 매개변수로 변화량을 나타내어 학습 부담을 줄이는 방식이므로 A에 대응한다.

(나) Adapter → C: 모델 내부에 **작은 추가 모듈을 삽입**하고, 그 모듈의 매개변수를 학습하여 새로운 작업에 적응시킨다. 기존 모델이 계산한 표현을 추가 모듈이 작업에 맞게 조정하는 방식이므로 C에 대응한다. 어텐션에 접두 벡터를 제공하는 B와는 추가하는 대상이 다르다.

(다) Prefix Tuning → B: 각 층의 어텐션 계산에 사용할 **학습 가능한 접두 벡터**를 추가한다. 어텐션이 이 벡터를 추가 문맥처럼 참조하게 하여 작업에 맞는 출력을 유도하므로 B에 대응한다. 작은 처리 모듈을 삽입하는 Adapter와 달리, 어텐션이 참조할 벡터를 학습한다.

따라서 (가)–A, (나)–C, (다)–B로 연결되며, ②의 조합과 일치한다.

NoteBase

문 22. 다음 문장에서 바이그램(bigram) 언어 모델을 사용할 때, ‘나는 학교에’ 다음에 ‘간다’가 나올 확률을 계산하는 올바른 식은?

나는 학교에 간다.

① P(간다)

② P(간다 | 나는)

③ P(간다 | 학교에)

④ P(간다 | 나는, 학교에)

정답 3번

간단한 해설

바이그램은 **바로 앞 한 단어**를 조건으로 다음 단어의 확률을 구한다. '간다' 바로 앞은 '학교에'이므로, '학교에'가 주어졌을 때 '간다'가 나올 조건부확률을 사용한다.

공통 해설

바이그램 언어 모델은 앞선 문맥 전체 대신 바로 앞 한 단어만으로 다음 단어의 확률을 근사한다. 일반적으로 $P(w_i \mid w_{i-1})$로 나타내며, 바이그램의 '두 단어'는 **직전 단어와 예측할 단어를 합친 것**이다. 앞의 두 단어를 모두 조건으로 삼는다는 뜻이 아니다.

'나는 학교에 간다'에서 예측할 단어는 '간다', 직전 단어는 '학교에'이다. 따라서 사용할 식은 P(간다 학교에)이다. 조건을 두지 않거나, 직전 단어가 아닌 '나는'만 조건으로 삼거나, '나는'과 '학교에'를 모두 조건으로 삼는 식은 이 바이그램 규칙에 맞지 않는다.

NoteBase

문 23. 오토인코더(autoencoder)에 대한 설명으로 옳지 않은 것은?

① 인코더를 통해 입력 데이터의 특징을 추출하고 디코더를 통해 원본 데이터를 재구성하는 학습을 한다.

② 잡음제거 오토인코더는 입력데이터와 동일한 데이터가 출력되도록 학습한다.

③ 오토인코더는 출력의 결과가 입력값과 동일하도록 비지도 학습을 수행한다.

④ 희소 오토인코더는 코딩층에서 출력값이 0이 아닌 노드의 개수를 최소화하는 오토인코더이다.

정답 2번

간단한 해설

잡음제거 오토인코더는 잡음이 섞인 데이터를 입력받아 **잡음을 넣기 전의 원본 데이터**를 복원하도록 학습한다. 따라서 출력의 목표는 잡음이 포함된 입력 데이터 자체가 아니다.

선지별 해설

① O: 인코더는 입력 데이터를 특징을 담은 잠재 표현으로 변환하고, 디코더는 그 표현을 이용해 원본 데이터를 재구성한다. **재구성 결과와 원본 사이의 오차**를 줄이면서 입력의 구조와 특징을 학습한다.

② X: 잡음제거 오토인코더의 학습 입력은 원본에 잡음을 가한 데이터이고, 출력의 목표는 **잡음이 없는 원본 데이터**이다. 예를 들어 깨끗한 이미지에 잡음을 섞어 입력하더라도, 출력은 잡음이 섞인 이미지가 아니라 깨끗한 이미지에 가까워지도록 학습한다. 따라서 '입력데이터와 동일한 데이터'를 출력하도록 학습한다는 설명은 잡음제거 학습의 목표와 다르다.

③ O: 기본적인 오토인코더는 입력 데이터를 재구성 목표로 삼아, 출력이 입력과 최대한 같아지도록 학습한다. 별도의 사람이 부여한 분류 레이블 없이 **데이터 자체를 학습 목표로 사용**하므로 비지도 학습에 해당한다. 여기서 '동일하도록'은 학습 목표를 뜻하며, 학습 후 모든 입력을 오차 없이 복원한다는 보장은 아니다.

④ O: 희소 오토인코더는 재구성 오차를 줄이는 동시에 코딩층에 희소성 제약을 가하여 **소수의 노드만 활성화**되도록 유도한다. 선지의 '0이 아닌 노드의 개수를 최소화'한다는 표현은 이러한 희소 표현을 뜻한다. 노드 자체의 총개수를 줄이는 것이 아니라 활성화를 제한하는 것이며, 구현에 따라 비활성 노드의 출력은 정확히 0이거나 0에 가까운 값이 될 수 있다.

NoteBase

문 24. CLIP(Contrastive Language-Image Pretraining)에 대한 설명으로 옳지 않은 것은?

① 사전훈련을 통해 텍스트 데이터를 학습해 놓은 모델인 트랜스포머 언어 모델을 기반으로 한다.

② 이미지와 텍스트를 각각 임베딩 공간에 표현하고, 서로 대응하는 이미지와 텍스트의 유사도가 높아지도록 대조 학습(contrastive learning)을 수행한다.

③ 제로 샷(zero-shot) 예측과는 달리, 클래스를 추가하고 모델을 재학습하여 새로운 입력 이미지를 분류할 수 있다.

④ 시각 정보와 언어 정보 사이의 대응 관계를 학습한 멀티 모달 모델로 이미지 검색, 이미지 설명 생성, 시각 질의응답 등 다양한 작업으로 확장할 수 있다.

정답 3번

간단한 해설

CLIP은 새 클래스의 이름이나 설명을 텍스트로 제시하고 이미지와의 유사도를 비교하는 **제로샷 분류**를 수행한다. 새 클래스를 분류하기 위해 **반드시 모델을 재학습할 필요는 없다**.

선지별 해설

① O: CLIP의 텍스트 인코더는 **트랜스포머 언어모델 계열의 구조**를 사용한다. 사전훈련으로 얻은 텍스트 표현을 이미지 표현과 비교한다. 다만 이는 기존 GPT 등의 사전학습 가중치를 그대로 가져왔다는 뜻이 아니다. 원래 CLIP은 이미지·텍스트 쌍을 이용하여 두 인코더를 처음부터 함께 사전훈련하였다. 구조를 바탕으로 한다는 점과 기존 모델의 가중치를 재사용한다는 점을 구별해야 한다.

② O: 이미지 인코더와 텍스트 인코더가 각각 만든 표현을 **공통 임베딩 공간**으로 투영한다. 올바른 이미지·텍스트 쌍의 유사도는 높이고, 대응하지 않는 쌍의 유사도는 낮추는 대조학습을 수행한다.

③ X: CLIP의 특징은 새 클래스를 추가할 때마다 재학습하는 방식이 아니라, 후보 클래스의 이름이나 설명을 텍스트로 만들고 입력 이미지와의 유사도를 비교하는 **제로샷 예측**을 할 수 있다는 것이다. 따라서 새 클래스 분류에 재학습이 필요하다는 설명으로 이해하면 잘못이다. 별도의 목적에 따라 미세조정을 수행할 수 있다는 사실과 제로샷 분류에 재학습이 필수라는 주장은 다르다.

④ O: 이미지와 언어의 의미적 대응 관계를 학습하므로 이미지·텍스트 간 검색 등에 활용할 수 있다. 또한 다른 생성모델이나 질의응답 모듈과 결합하여 이미지 설명 생성이나 시각 질의응답으로 **확장**할 수 있다. CLIP 자체가 독립적인 문장 생성기라는 뜻은 아니다.

NoteBase

문 25. 다음 사례에서 과적합 위험을 최소화하면서 성능을 확보하기 위한 초기 절차로 옳은 것은?

대규모 ImageNet으로 사전 학습된 CNN(Convolutional Neural Network) 모델을 샘플 수가 수천 장 수준인 소규모 의료 영상 분류 과제에 적용하려고 한다.

① 사전 학습된 백본(backbone)을 고정하고, 새롭게 초기화한 분류기 레이어만 학습한다.

② 모든 레이어를 동일 학습률로 한 번에 동시 미세조정(fine-tune)한다.

③ 사전학습 가중치를 전부 고정한 채 드롭아웃(dropout) 비율만 조정한다.

④ 모든 레이어를 랜덤값으로 초기화하고 네트워크 전체를 새로 훈련한다.

정답 1번

간단한 해설

소규모 데이터에서는 **사전 학습된 특징은 보존하고 새 분류기만 학습**하면 학습할 매개변수를 줄여 과적합 위험을 낮출 수 있다. 새 분류기는 기존 특징을 의료 영상의 분류 결과에 연결하도록 학습한다.

공통 해설

전이학습은 다른 과제에서 학습한 특징을 새 과제에 재사용하는 방법이다. CNN의 백본은 영상에서 특징을 추출하는 부분이고, 분류기는 그 특징을 각 범주의 예측으로 연결하는 부분이다. 수천 장 수준의 의료 영상으로 학습을 시작할 때는 **백본의 가중치를 고정하고 새 분류기만 학습**하여, 사전 학습된 특징을 보존하면서 의료 영상의 분류 기준을 익히게 한다. 전체 네트워크를 학습할 때보다 조정하는 매개변수가 적으므로 제한된 학습 자료에 과도하게 맞춰질 위험을 줄일 수 있다.

이처럼 특징의 재사용과 새 분류 기준의 학습을 함께 충족해야 한다. 처음부터 모든 레이어를 같은 학습률로 미세조정하면 학습 대상이 많아지고 기존 특징도 바뀌므로 소규모 자료에서 과적합하거나 유용한 특징을 훼손할 위험이 있다. 가중치를 전부 고정한 채 드롭아웃 비율만 바꾸면 새 분류 기준에 맞게 가중치를 학습하지 못한다. 모든 레이어를 무작위로 초기화하면 사전 학습된 특징까지 소규모 자료로 다시 배워야 하므로 특징 재사용의 이점이 사라진다.

NoteBase