

## AI&기계학습 기초

1 다음 중 AI, ML, DL의 관계에 대한 설명으로 옳은 것은?

- ① ML은 AI과 DL의 상위 개념이다.
- ② AI는 반드시 ML을 사용해야 한다.
- ③ ML은 AI의 한 부분으로, 데이터로부터 학습을 수행한다.
- ④ 규칙 기반 시스템은 ML의 한 분야이다.
- ⑤ ML은 사람이 직접 코딩한 규칙을 그대로 사용하는 방식이다.

---

## 확인문제 정답 및 해설

1 다음 중 AI, ML, DL의 관계에 대한 설명으로 옳은 것은?

- ① ML은 AI과 DL의 상위 개념이다.
- ② AI는 반드시 ML을 사용해야 한다.
- ③ ML은 AI의 한 부분으로, 데이터로부터 학습을 수행한다.
- ④ 규칙 기반 시스템은 ML의 한 분야이다.
- ⑤ ML은 사람이 직접 코딩한 규칙을 그대로 사용하는 방식이다.

정답 및 해설

정답: ③

해설: ML은 AI의 한 분야로, 데이터로부터 규칙을 학습해 목표 달성을 돕는다.

- ① ML은 AI의 하위 개념이다.
- ② AI는 규칙/휴리스틱 기반만으로도 구현 가능하다.
- ④ 규칙 기반 시스템은 AI의 대표적인 예시이다.
- ⑤ ML은 데이터에서 추론한다.

---

## 확인문제

2 다음 중 Feature와 Label에 대한 설명으로 옳지 않은 것은?

- ① Feature는 모델이 예측에 사용하는 입력 정보이다.
- ② Label은 모델이 예측하려는 정답, 즉 학습의 목표이다.
- ③ Feature와 Label의 분포와 관계가 학습 결과를 결정한다.
- ④ Feature는 반드시 수치형 데이터만 될 수 있다.

⑤ Label은 분류 문제에서는 카테고리, 회귀 문제에서는 수치값이 될 수 있다.

---

#### 확인문제 정답 및 해설

2 다음 중 Feature와 Label에 대한 설명으로 옳지 않은 것은?

- ① Feature는 모델이 예측에 사용하는 입력 정보이다.
- ② Label은 모델이 예측하려는 정답, 즉 학습의 목표이다.
- ③ Feature와 Label의 분포와 관계가 학습 결과를 결정한다.
- ④ Feature는 반드시 수치형 데이터만 될 수 있다.
- ⑤ Label은 분류 문제에서는 카테고리, 회귀 문제에서는 수치값이 될 수 있다.

#### 정답 및 해설

정답: ④

해설: Feature는 수치형뿐 아니라 범주형, 텍스트, 이미지/음성 등도 가능하다.

---

#### 확인문제

3 다음 중  $f(X)$ , 즉 미지의 참 함수를 학습하는 이유로 적절하지 않은 것은?

- ① Feature(X)와 label(Y)사이의 관계를 알기 위해
- ② 모든 데이터의 측정오차( $\epsilon$ )를 제거하기 위해
- ③ 새로운 입력 X에 대해 목표값 Y를 예측하기 위해
- ④ 여러 feature(X)들 중 label(Y)와 관계를 갖고 있는 feature를 알기 위해
- ⑤ Feature(X) 값의 변화에 따라서 label(Y)이 어떻게 변화하는지 알기 위해

---

### 확인문제 정답 및 해설

3 다음 중  $f(X)$ , 즉 미지의 참 함수를 학습하는 이유로 적절하지 않은 것은?

- ① Feature(X)와 label(Y)사이의 관계를 알기 위해
- ② 모든 데이터의 측정오차( $\epsilon$ )를 제거하기 위해
- ③ 새로운 입력  $X$ 에 대해 목표값  $Y$ 를 예측하기 위해
- ④ 여러 feature(X)들 중 label(Y)와 관계를 갖고 있는 feature를 알기 위해
- ⑤ Feature(X) 값의 변화에 따라서 label(Y)이 어떻게 변화하는지 알기 위해

### 정답 및 해설

정답: ②

해설: 학습을 잘 한다고 하여서 측정오차를 줄이거나 제거할 수 없다.

---

### 확인문제

1 다음 중 지도학습(supervised learning)에 대한 설명으로 옳은 것은?

- ① 입력 데이터만 있고 정답(라벨)이 없는 데이터로 학습한다.
- ② 새로운 입력에 대해 정답을 잘 맞추는 함수를 학습한다.
- ③ 오버피팅은 훈련데이터에서의 오류를 최대한 줄이는 것으로 학습이 목표하는 방향이다.
- ④ 지도학습 모델의 출력은 반드시 범주(카테고리)이다.
- ⑤ 지도학습은 측정오차를 최소화하는 것이 목적이다.

---

### 확인문제 정답 및 해설

1 다음 중 지도학습(supervised learning)에 대한 설명으로 옳은 것은?

- ① 입력 데이터만 있고 정답(라벨)이 없는 데이터로 학습한다.
- ② 새로운 입력에 대해 정답을 잘 맞추는 함수를 학습한다.
- ③ 오버피팅은 훈련데이터에서의 오류를 최대한 줄이는 것으로 학습이

목표하는 방향이다.

④ 지도학습 모델의 출력은 반드시 범주(카테고리)이다.

⑤ 지도학습은 측정오차를 최소화하는 것이 목적이다.

정답 및 해설

정답: ②

해설: 지도학습은 입력(특성)과 정답(라벨)이 쌍으로 있는 데이터를 학습하여 새로운 입력의 정답을 맞추는 규칙을 찾는다.

① 입력만 있고 정답이 없는 경우는 비지도 학습이다.

③ 일반화 성능을 높이는 것이 궁극적인 목적이기에 오버피팅은 피해야 한다.

④ 지도학습 모델 출력은 수치(회귀문제)와 는 범주(분류문제) 모두 가능하다.

⑤ 지도학습에서는 예측값과 정답의 차이(오차)를 줄이는 것이 핵심이다.

---

확인문제

2 다음 중 회귀(Regression) 문제의 특징으로 옳은 것은?

① 스팸/정상 분류와 같은 범주 예측 문제를 다룬다.

② 예측값은 연속적인 숫자(예: 가격, 점수, 온도)로 나온다.

③ 대표적인 성능 지표는 정확도(Accuracy)이다.

④ 예측 결과는 이산적인 값으로 한정된다.

⑤ 훈련 데이터와 테스트 데이터의 분포가 다르면 반드시 발생한다.

---

확인문제 정답 및 해설

2 다음 중 회귀(Regression) 문제의 특징으로 옳은 것은?

① 스팸/정상 분류와 같은 범주 예측 문제를 다룬다.

② 예측값은 연속적인 숫자(예: 가격, 점수, 온도)로 나온다.

③ 대표적인 성능 지표는 정확도(Accuracy)이다.

④ 예측 결과는 이산적인 값으로 한정된다.

⑤ 훈련 데이터와 테스트 데이터의 분포가 다르면 반드시 발생한다.

정답 및 해설

정답: ②

해설: 회귀 문제는 연속적인 수치를 예측하는 문제다. 예를 들어 집값, 매출액, 온도 등이 있다.

① 범주를 예측하는 것은 분류(Classification) 문제다.

③ 회귀에서는 MSE, RMSE,  $R^2$  등의 지표를 사용한다.

④ 이산 값은 분류 문제의 특징이다.

⑤ 훈련/테스트 분포 차이는 데이터 쉬프트 개념이지 회귀의 특징은 아니다.

---

#### 확인문제

3 다음 중 혼동행렬에 대한 설명으로 옳지 않은 것은?

① TP (True Positive): 실제 양성을 양성으로 예측한 경우

② TN(True Negative): 실제 음성을 음성으로 예측한 경우

FP(False Positive): 실제 음성인데 양성으로 잘못 예측한 경우

④ FN(False Negative): 실제 양성인데 음성으로 잘못 예측한 경우

⑤ 혼동행렬은 회귀 문제의 성능 평가에서 사용된다.

---

#### 확인문제 정답 및 해설

3 다음 중 혼동행렬에 대한 설명으로 옳지 않은 것은?

① TP(True Positive): 실제 양성을 양성으로 예측한 경우

② TN(True Negative): 실제 음성을 음성으로 예측한 경우

FP(False Positive): 실제 음성인데 양성으로 잘못 예측한 경우

④ FN(False Negative): 실제 양성인데 음성으로 잘못 예측한 경우

⑤ 혼동행렬은 회귀 문제의 성능 평가에서 사용된다.

정답 및 해설

정답: ⑤

해설: 혼동행렬은 분류 문제의 성능 평가 지표이다. 회귀 문제에서는 혼

동행렬 대신 MSE, RMSE,  $R^2$  등의 지표를 사용한다.

### 확인문제

훈련 오류와 테스트 오류에 대한 설명으로 옳은 것은?

- ① 훈련 오류는 새로운 데이터에 대한 예측 정확도를 측정한다.
- ② 테스트 오류는 학습에 사용된 데이터에 대해 계산한 평균 오차이다.
- ③ 궁극적으로 테스트 오류를 최소화하는 것이 학습의 목적이다.
- ④ 일반적으로 훈련 오류는 테스트 오류보다 크게 나타난다.
- ⑤ 훈련 오류는 항상 0에 수렴해야 좋은 모델이다.

### 확인문제 정답 및 해설

훈련 오류와 테스트 오류에 대한 설명으로 옳은 것은?

- ① 훈련 오류는 새로운 데이터에 대한 예측 정확도를 측정한다.
- ② 테스트 오류는 학습에 사용된 데이터에 대해 계산한 평균 오차이다.
- ③ 궁극적으로 테스트 오류를 최소화하는 것이 학습의 목적이다.
- ④ 일반적으로 훈련 오류는 테스트 오류보다 크게 나타난다.
- ⑤ 훈련 오류는 항상 0에 수렴해야 좋은 모델이다.

정답: ③

해설:

학습의 목적은 일반화 성능을 높이는 것이며, 이는 테스트 오류를 최소화하는 것이다.

- ① 새로운 데이터는 테스트 오류가 담당한다.
- ② 이는 훈련 오류에 대한 설명이다.
- ④ 일반적으로 훈련 오류 < 테스트 오류이다.
- ⑤ 훈련 오류가 0에 가까우면 오버피팅 가능성이 높다.

### 확인문제

다음 중 언더피팅과 오버피팅의 특징을 올바르게 설명한 것은?

- ① 언더피팅은 훈련 오류와 테스트 오류 모두 크게 나타난다.
- ② 언더피팅은 모델이 지나치게 복잡해 데이터 잡음까지 학습하는 현상이다.

- ③ 오버피팅은 모델이 단순해 중요한 패턴을 놓치는 현상이다.
- ④ 오버피팅은 훈련 오류가 크고 테스트 오류가 작게 나타난다.
- ⑤ 두 경우 모두 모델 복잡도를 높이면 해결된다.

## 확인문제 정답 및 해설

다음 중 언더피팅과 오버피팅의 특징을 올바르게 설명한 것은?

- ① 언더피팅은 훈련 오류와 테스트 오류 모두 크게 나타난다.
- ② 언더피팅은 모델이 지나치게 복잡해 데이터 잡음까지 학습하는 현상이다.
- ③ 오버피팅은 모델이 단순해 중요한 패턴을 놓치는 현상이다.
- ④ 오버피팅은 훈련 오류가 크고 테스트 오류가 작게 나타난다.
- ⑤ 두 경우 모두 모델 복잡도를 높이면 해결된다.

정답: ①

해설:

언더피팅은 모델이 너무 단순하여 데이터 패턴을 제대로 학습하지 못해 훈련 오류와 테스트 오류 모두 크게 나타난다.

②는 오버피팅 설명이다.

③은 반대 현상이다.

④ 오버피팅은 훈련 오류 ↓, 테스트 오류 ↑ 이다.

⑤ 언더피팅은 복잡도를 높여 개선 가능하지만, 오버피팅은 오히려 악화된다.

## 확인문제

K-겹 교차검증(K-fold CV)에 대한 설명으로 옳지 않은 것은?

- ① 데이터를 K개 그룹(폴드)으로 나누어 각 그룹을 한 번씩 검증셋으로 사용한다.
- ② K번 반복한 후 각 폴드별 평균 오차를 계산해 일반화 성능을 추정한다.
- ③  $K=n$ 일 경우 Leave-One-Out 교차검증(LOOCV)이다.
- ④ 교차검증은 데이터 활용성을 높이면서 테스트 오류에 대한 추정이 가능하다.
- ⑤ 교차검증 시 각 폴드에 속한 데이터가 겹쳐도 무방하다.

## 확인문제 정답 및 해설

K-겹 교차검증(K-fold CV)에 대한 설명으로 옳지 않은 것은?

- ① 데이터를 K개 그룹(폴드)으로 나누어 각 그룹을 한 번씩 검증셋으로 사용한다.
- ② K번 반복한 후 각 폴드별 평균 오차를 계산해 일반화 성능을 추정한다.
- ③  $K=n$ 일 경우 Leave-One-Out 교차검증(LOOCV)이다.
- ④ 교차검증은 데이터 활용성을 높이면서 테스트 오류에 대한 추정이 가능하다.
- ⑤ 교차검증 시 각 폴드에 속한 데이터가 겹쳐도 무방하다.

정답: ⑤

해설:

교차검증에서는 각 폴드별 데이터가 겹치지 않도록 명확히 분리해야 한다.

## 확인문제

다음 중 비지도 학습의 예로 해당하는 것은?

- A. 스팸/이메일 분류
- B. 집값 변화 예측
- C. (라벨이 없는 데이터에서) 구매 성향별 고객 묶음
- D. 주간 날씨 예측

## 확인문제 정답 및 해설

다음 중 비지도 학습의 예로 해당하는 것은?

- B. 집값 변화 예측
- C. (라벨이 없는 데이터에서) 구매 성향별 고객 묶음
- D. 주간 날씨 예측

정답: C. 구매 성향별 고객 묶음

해설: 정답이 없는 묶음 문제로 전형적인 클러스터링 문제



## 확인문제

다음은 클러스터링 기법에 대한 설명이다. 각 설명이 K-means의 특징인지, 계층적 군집의 특징인지 알맞게 분류하시오. (AI 기계학습 기초)

※ 단, 두 방법 모두에 해당하지 않을 수 있다.

- A. 클러스터 수 K를 미리 지정해야 한다.
- B. 덴드로그램(Dendrogram)을 통해 전체 구조를 시각적으로 제공한다.
- C. 데이터 간 병합(상향식) 과정을 통해 클러스터를 형성한다.
- D. 클러스터 내부 변동(Within-cluster variation)의 합이 최소가 되도록 분할한다.
- E. 매우 큰 데이터셋에도 계산 복잡도가 크게 늘지 않아 효율적이다.

## 확인문제 정답 및 해설

다음은 클러스터링 기법에 대한 설명이다. 각 설명이 K-means의 특징인지, 계층적 군집의 특징인지 알맞게 분류하시오. (AI 기계학습 기초)

※ 단, 두 방법 모두에 해당하지 않을 수 있다.

- A. 클러스터 수 K를 미리 지정해야 한다.
- B. 덴드로그램(Dendrogram)을 통해 전체 구조를 시각적으로 제공한다.
- C. 데이터 간 병합(상향식) 과정을 통해 클러스터를 형성한다.
- D. 클러스터 내부 변동(Within-cluster variation)의 합이 최소가 되도록 분할한다.
- E. 매우 큰 데이터셋에도 계산 복잡도가 크게 늘지 않아 효율적이다.

정답: K-means: A, D, E / 계층적 군집: B, C

해설:

K-means는 데이터를 K개의 그룹으로 나누는 대표적 기법으로, 군집 내 분산을 최소화하도록 반복 계산한다. 빠르고 대규모 데이터에 효율적이지만 클러스터 수를 미리 지정해야 하고 복잡한 모양의 데이터에는 한계가 있다.

계층적 군집은 데이터를 병합-분할하는 트리(덴드로그램) 구조를 만드는 방식으로, 클러스터 수를 미리 정하지 않아도 되고 데이터 구조를 시각적으로 파악하기 쉽다. 그러나 데이터가 많을수록 계산 비용이 커져 대규모 데이터에는 비효율적이다.

## AI&기계학습 방법론1 선형회귀

1 다음 질문에 대해 서술하시오.

- ① 선형회귀의 기본 목적은 무엇인가?
- ② 단순선형회귀 모형의 수식과 각 계수의 의미는?
- ③ 잔차(Residual)의 정의와 역할은?
- ④ 최소제곱법(Least Squares)의 핵심 아이디어는?

---

### 확인문제 정답 및 해설

1 다음 질문에 대해 서술하시오.

- ① 선형회귀의 기본 목적은 무엇인가?
- ② 단순선형회귀 모형의 수식과 각 계수의 의미는?
- ③ 잔차(Residual)의 정의와 역할은?
- ④ 최소제곱법(Least Squares)의 핵심 아이디어는?

정답 및 해설

해설: ① 입력 변수와 출력 변수 간의 선형 관계를 찾고, 이를 통해 예측 및 해석을 수행하는 것

해설: ② 모형:  $Y = \beta_0 + \beta_1 X + \varepsilon$ ,  $\beta_0$ ,  $\beta_1$ 은 절편과 기울기 모수, 또는 관측오차

해설: ③ 잔차 = 실제값 - 예측값. 모델이 설명하지 못한 부분으로, 모델 적합도를 판단하는 지표

해설: 잔차제곱합(RSS)을 최소화하는 계수를 찾는 방법으로, 데이터와 회귀선 간의 차이를 최소화하는 원리

---

## 확인문제

1 다음 질문에 대해 서술하시오.

- ① 다중선형회귀 모형의 수식을 쓰고, 각 항의 의미를 설명하시오.
- ② 다중선형회귀에서 변수를 추가할 때 생길 수 있는 장점과 단점은 무엇인가?
- ③ 다중공선성이 발생하면 어떤 문제가 생기는가?
- ④ 다중선형회귀 모형의 성능을 평가하는 대표적인 지표 두가지는 무엇인가?

---

## 확인문제 정답 및 해설

1 다음 질문에 대해 서술하시오.

- ① 다중선형회귀 모형의 수식을 쓰고, 각 항의 의미를 설명하시오.
- ② 다중선형회귀에서 변수를 추가할 때 생길 수 있는 장점과 단점은 무엇인가?
- ③ 다중공선성이 발생하면 어떤 문제가 생기는가?
- ④ 다중선형회귀 모형의 성능을 평가하는 대표적인 지표 두가지는 무엇인가?

### 정답 및 해설

해설:  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$ , Y: 종속변수, X: 설명변수,  $\beta_0$ :

절편,  $\beta_1, \dots, \beta_p$ : 설명변수의 효과,  $\varepsilon$ : 오차항

해설: ② 변수를 추가하면 더 많은 정보를 고려할 수 있어 설명력이 향상될 수 있다.

하지만 불필요한 변수를 넣으면 과적합 위험이 커지고 해석이 복잡해진다.

해설: ③ 다중공선성이 심하면 계수의 분산이 커져 추정이 불안정해지고, 해석이 어렵거나 예측력이 떨어진다.

해설: 대표적인 지표로 결정계수  $R^2$ 과 잔차표준오차 RSE가 있다.

---

### 확인문제

1 다음 중 틀린 설명은?

- A. 다중범주 문제에는 선형회귀보다 (Multiclass) Logistic Regression 또는 다른 분류 모델이 적절하다.
- B. 선형 회귀 예측 값을 확률로 해석하면 0미만 또는 1초과가 나올 수 있어 확률로서 타당하지 않다.
- C. 기상 상태 예측 모델을 만들 때,  $Y \in \{1: 맑음, 2: 눈, 3: 비\}$ 와 같이 정수형 코딩을 진행해야만 적절한 예측이 가능하다.
- D. 범주형(질적) 변수는 순서가 없는 집합에서 값을 가진다.

---

### 확인문제 정답 및 해설

1 다음 중 틀린 설명은?

- A. 다중범주 문제에는 선형회귀보다 (Multiclass) Logistic Regression 또는 다른 분류 모델이 적절하다.
- B. 선형 회귀 예측 값을 확률로 해석하면 0미만 또는 1초과가 나올 수 있어 확률로서 타당하지 않다.
- C. 기상 상태 예측 모델을 만들 때,  $Y \in \{1: 맑음, 2: 눈, 3: 비\}$ 와 같이 정수형 코딩을 진행해야만 적절한 예측이 가능하다.
- D. 범주형(질적) 변수는 순서가 없는 집합에서 값을 가진다.

정답 및 해설

정답: C

해설

- 범주형 변수를 정수 코딩했을 때, 인위적 순서를 암시하고 12와 23의 거리 동일성까지 암묵적으로 가정하게 됨.

---

### 확인문제

2 다음 중 로지스틱 함수의 출력으로 옳은 것은?

- A.  $[0, 1]$
- B.  $(0,1)$
- C.  $(-\infty, +\infty)$

D.  $\{0, 1\}$

---

확인문제 정답 및 해설

2 다음 중 로지스틱 함수의 출력으로 옳은 것은?

A.  $[0, 1]$

B.  $(0, 1)$

C.  $(-\infty, +\infty)$

D.  $\{0, 1\}$

정답 및 해설

정답: B

해설: 로지스틱 회귀의 출력은 어떤  $x, \beta$  에 대해서도  $p(x; \beta) \in (0, 1)$ : 예측 확률은 항상 0과 1 사이의 값이 보장됩니다. A가 답이 아닌 이유는 끝점 0과 1을 포함하기 때문입니다.

로지스틱 회귀의 출력 중 0과 1은 극한으로만 접근합니다.

---

확인문제

3 다음 문장이 옳은가? (O/X)

로지스틱 회귀에서 "현재 확률 함수가 데이터를 얼마나 잘 설명하는지"를

나타낸 지표로 사용하는 것은 평균 제곱 오차(MSE; Mean Squared Error) 이다.

---

확인문제 정답 및 해설

3 다음 문장이 옳은가? (O/X)

로지스틱 회귀에서 "현재 확률 함수가 데이터를 얼마나 잘 설명하는지"를

나타낸 지표로 사용하는 것은 평균 제곱 오차(MSE; Mean Squared Error) 이다.

정답 및 해설

정답: X

해설

로지스틱 회귀에서 모수(가중치)를 추정할 때 사용하는 표준 지표는 평균제곱오차(MSE)가 아니라, 우도(likelihood)를 최대화하는 방법 (MLE; Maximum Likelihood Estimation)이다.

## 확인문제

1. 손실함수(Loss Function)의 역할로 가장 알맞은 것은?

- a. 입력 데이터를 무작위로 섞는다.
- b. 모델의 예측이 얼마나 잘못되었는지 측정한다.
- c. 파라미터를 업데이트하는 규칙을 제공한다.
- d. 학습률(learning rate)을 설정한다.

2. 경사하강법(Gradient Descent)에서 'Step 1'에 해당하는 단계는?

- a. 파라미터를 무작위로 초기화
- b. 손실함수를 최소화
- c. 파라미터에 대한 기울기(미분값)를 계산
- d. 데이터를 섞어 미니배치를 생성

3. 확률적 경사 하강법(SGD)의 장점은?

- a. 항상 전역 최적해(global optimum)에 도달한다.
- b. 계산 비용이 커서 정밀하다.
- c. 일부 데이터만으로 업데이트해 연산 비용이 작다.
- d. 항상 같은 결과로 수렴한다.

4. 다음 중 Convex 문제의 특징으로 맞는 것은?

- a. 2차 미분이 항상 0이다.
- b. 2차 미분이 모든 구간에서 음수이다.
- c. 2차 미분이 모든 구간에서 양수이다.
- d. 미분값이 존재하지 않는다.

5. 학습 과정에서 손실함수를 최소화하기 위해 조정하는 값은 무엇인가요?
6. 경사하강법에서 파라미터를 업데이트 시 사용하는 고정된 비율은?
7. SGD가 경사하강법(full-batch GD)보다 항상 수렴 속도가 빠른가?
8. SGD는 국소 최소값(local minima)에서 벗어날 수 있는 장점이 있다. 그 이유는?

## 확인문제 정답 및 해설

정답:

- b. 모델의 예측이 얼마나 잘못되었는지 측정한다
- c. 파라미터에 대한 기울기(미분값)를 계산
- c. 일부 데이터만으로 업데이트해 연산 비용이 작다
- c. 2차 미분이 모든 구간에서 양수이다

1. 손실함수는 모델 예측값과 실제 정답 사이의 차이를 수치화한 지표다.

이 값이 작아질수록 모델의 예측 성능이 좋아졌음을 의미하며, 학습은 손실을 최소화하는 방향으로 진행된다.

2. 경사하강법의 첫 단계는 현재 파라미터에서 손실함수의 기울기를 계산하는 것이다.

이 기울기는 손실이 가장 빨리 줄어드는 방향을 알려주며, 이후 파라미터를 이 방향으로 이동시킨다.

3. SGD는 전체 데이터가 아니라 일부 샘플(미니배치)을 사용해 기울기를 계산한다.

덕분에 계산량이 적어 속도가 빠르고, 대규모 데이터셋에도 효율적으로 적용할 수 있다.

4. Convex 문제는 손실함수가 아래로 볼록한 형태로, 2차 미분이 항상 양수이거나 0이다.

따라서 어떤 국소 최소점(local minimum)도 전역 최소점(global

minimum)과 동일하다.

정답:

- 5. 파라미터(Parameters)
- 6. 학습률(Learning Rate,  $\alpha$ )
- 7. 에폭(Epoch)
- 8. 노이즈 섞인 업데이트로 탐색 가능

5. 학습 과정에서 실제로 조정되는 것은 모델의 파라미터(가중치, 편향)이다.

이 값들을 반복적으로 갱신하여 손실이 최소화되도록 한다.

6. 학습률(learning rate)은 파라미터를 한 번 업데이트할 때 이동하는 거리의 비율이다.

값이 너무 크면 발산할 수 있고, 너무 작으면 학습이 느려진다.

7. 항상 그렇지 않다. 예를 들어, convex 문제(또는 convex에 가까운 경우)에서는

오히려 GD의 수렴 속도가 더 빠를 수 있다.

SGD가 수렴 속도가 빠르다고 단정지으면 안 된다.

8. SGD는 미니배치를 무작위로 선택하기 때문에 기울기에 노이즈가 포함된다.

이 특성 덕분에 국소 최소값이나 안장점(saddle point)에 머무르지 않고 더 넓은 영역을 탐색할 수 있다.

확인문제-딥러닝 및 이미지 모델

확인문제

1 CNN 기반 모델 변천사를 순서대로 나열하면?

- A. MobileNet
- B. VGG
- C. AlexNet
- D. GoogLeNet
- E. ResNet



---

### 확인문제 정답 및 해설

1 CNN 기반 모델 변천사를 순서대로 나열하면?

- A. MobileNet
- B. VGG
- C. AlexNet
- D. GoogLeNet
- E. ResNet

정답 및 해설

정답:

C→B→D→E→A

해설:

- AlexNet (2012) – ILSVRC 대회 우승, 딥러닝 CNN의 본격적인 시작점임.
  - VGG (2014) – 3x3 작은 필터를 깊게 쌓은 단순하면서도 강력한 구조임.
  - GoogLeNet / Inception (2014) – Inception 모듈 도입, 효율적인 연산 적용함.
  - ResNet (2015) – Residual Connection(잔차 연결) 도입, 기울기 소실 문제 해결함.
  - MobileNet (2017) – 모바일, 임베디드 환경 최적화, Depthwise Separable Convolution 도입함.
- 각 모델은 깊이, 효율성, 연결 구조, 경량화 등 CNN의 한계를 단계적으로 극복하며 등장했음.

---

### 확인문제

2 CNN 모델의 특징과 가장 먼 것은?

- ① 이미지의 모든 화소를 나열하여 1차원 벡터로 만들어 입력 데이터를 활용한다

- ② 기본 구조는 합성곱 → 활성화 → 풀링 → 완전연결 등으로 이루어져 있다.
- ③ 지역적인 특징을 먼저 학습하고, 층이 깊어질수록 추상적 특징을 학습한다.
- ④ 위치 변화나 노이즈에 대해 강건하다.
- ⑤ 층이 깊어질수록 지역 수용 영역(Receptive Field)이 커져, 더 넓은 맥락을 이해할 수 있다

---

#### 확인문제 정답 및 해설

2 CNN 모델의 특징과 가장 먼 것은?

- ① 이미지의 모든 화소를 나열하여 1차원 벡터로 만들어 입력 데이터를 활용한다
- ② 기본 구조는 합성곱 → 활성화 → 풀링 → 완전연결 등으로 이루어져 있다.
- ③ 지역적인 특징을 먼저 학습하고, 층이 깊어질수록 추상적 특징을 학습한다.
- ④ 위치 변화나 노이즈에 대해 강건하다.
- ⑤ 층이 깊어질수록 지역 수용 영역(Receptive Field)이 커져, 더 넓은 맥락을 이해할 수 있다

정답 및 해설

정답: ①

해설:

CNN도 마지막의 분류 단계에서 FC Layer(완전연결층)를 쓰기 때문에 feature map을 1차원 벡터로 변환하긴 하지만 이것은 출력 단계에서만 일어나는 과정이고, CNN의 핵심은 입력 이미지의 2차원 구조를 보존한 채 합성곱과 풀링으로 지역적 특징을 추출하는 것이다. 따라서 " 이미지의 모든 화소를 나열하여 1차원 벡터로 입력한다"는 설명은 CNN 특징과 거리가 멀다.

---

#### 확인문제

3 CNN에서 풀링(Pooling)의 주요 목적이 아닌 것은?

- ① 차원 축소 및 연산량 감소
- ② 위치 변화에 대한 불변성 확보
- ③ 네트워크의 파라미터 수 증가
- ④ 지역적 특징을 요약하여 다음 층에 전달.

---

확인문제 정답 및 해설

3 CNN에서 풀링(Pooling)의 주요 목적이 아닌 것은?

- ① 차원 축소 및 연산량 감소
- ② 위치 변화에 대한 불변성 확보
- ③ 네트워크의 파라미터 수 증가
- ④ 지역적 특징을 요약하여 다음 층에 전달.

정답 및 해설

정답: ③

해설:

- ① Pooling은 보통 Max Pooling이나 Average Pooling을 사용하여 입력의 크기를 줄입니다. 이렇게 하면 연산해야 할 데이터의 양이 줄어들어 학습과 추론 속도가 빨라지고, 메모리 사용량도 절감됩니다.
- ② Max Pooling은 특정 맵의 가장 두드러진 값을 취함으로써, 입력 이미지가 약간 이동하거나 노이즈가 추가되어도 중요한 특징은 유지됩니다. 이 덕분에 풀링은 위치 변화에 대한 불변성을 확보합니다.
- ④ Pooling은 인접한 픽셀 정보를 요약하여 지역적인 특징을 더 압축된 형태로 전달합니다. 이는 이후 합성곱 층이 더 넓은 맥락(Context)을 이해하는 데 기여합니다.
- ③ Pooling 연산 자체는 학습해야 할 파라미터가 존재하지 않으며(비선형적 연산), 오히려 feature map의 크기를 줄여서 다음 층에서 필요한 연산량과 파라미터 수를 감소시킵니다. 따라서 "네트워크의 파라미터 수 증가"는 pooling의 주요 목적과는 거리가 멉니다.

확인문제

CNN 구조의 장점이 아닌 것은? (딥러닝 및 이미지 모델)

- ① 지역적인 특징 학습에 특화되어 작은 필터로 패턴이나 경계선 특징

을 잘 포착한다.

- ② 파라미터 효율성이 뛰어나 이미지 블록별 가중치를 공유함으로써 연산량을 줄인다.
- ③ Pooling과 Stride 연산으로 인해 위치 변화와 노이즈에도 견고하다.
- ④ 이미지 분류, 탐지 등 다양한 이미지 기반 과업에서 표준 모델로 활용된다.
- ⑤ 텍스트나 음성과 같은 순차적 데이터의 순서 정보를 효과적으로 반영한다.

## 확인문제 정답 및 해설

CNN 구조의 장점이 아닌 것은? (딥러닝 및 이미지 모델)

- ① 지역적인 특징 학습에 특화되어 작은 필터로 패턴이나 경계선 특징을 잘 포착한다.
- ② 파라미터 효율성이 뛰어나 이미지 블록별 가중치를 공유함으로써 연산량을 줄인다.
- ③ Pooling과 Stride 연산으로 인해 위치 변화와 노이즈에도 견고하다.
- ④ 이미지 분류, 탐지 등 다양한 이미지 기반 과업에서 표준 모델로 활용된다.
- ⑤ 텍스트나 음성과 같은 순차적 데이터의 순서 정보를 효과적으로 반영한다.

정답: ⑤

해설:

CNN은 이미지 처리에 최적화된 구조로, 지역적 특징 학습과 파라미터 효율성, 노이즈에 대한 견고함 등의 장점을 가진다.

그러나 텍스트나 음성과 같이 순차적(Sequential) 데이터의 순서 정보를 반영하는 데는 한계가 있다.

이러한 데이터는 주로 RNN(Recurrent Neural Network)이나 Transformer 계열 모델에 더 적합하다.

## 확인문제

다음 중 RNN의 특징으로 옳지 않은 것은? (딥러닝 및 이미지 모델)

- ① 은닉 상태(hidden state)를 통해 이전 시점의 정보를 다음 시점으로

전달한다.

- ② 순차적 데이터를 처리하는 데 적합하다.
- ③ 입력 데이터의 시계열 의존성을 반영할 수 있다.
- ④ 병렬 연산이 쉬워 대규모 데이터 처리에서 속도가 빠르다.
- ⑤ 언어 모델링, 기계 번역, 음성 인식 등에 활용된다.

확인문제 정답 및 해설

다음 중 RNN의 특징으로 옳지 않은 것은? (딥러닝 및 이미지 모델)

- ① 은닉 상태(hidden state)를 통해 이전 시점의 정보를 다음 시점으로 전달한다.
- ② 순차적 데이터를 처리하는 데 적합하다.
- ③ 입력 데이터의 시계열 의존성을 반영할 수 있다.
- ④ 병렬 연산이 쉬워 대규모 데이터 처리에서 속도가 빠르다.
- ⑤ 언어 모델링, 기계 번역, 음성 인식 등에 활용된다.

정답: ④

해설:

RNN은 입력 데이터를 순차적으로 처리하기 때문에 병렬 연산이 어렵고,

대규모 데이터 학습 시 속도가 느려지는 단점이 있다.

이 문제는 Transformer 계열 모델이 등장하면서 크게 개선되었다.

확인문제

다음 중 빈칸에 들어갈 단어로 알맞은 것을 모두 고르시오. (딥러닝 및 이미지 모델)

RNN의 대표적인 한계는 \_\_\_\_\_ 문제와 장기 의존성(Long-term Dependency) 학습의 어려움이다.

보기:

A. 기울기 소실 B. 기울기 폭발 C. 과적합 D. 병렬화 어려움

확인문제 정답 및 해설

다음 중 빈칸에 들어갈 단어로 알맞은 것을 모두 고르시오. (딥러닝 및 이미지 모델)

RNN의 대표적인 한계는 \_\_\_\_\_ 문제와 장기 의존성(Long-term Dependency) 학습의 어려움이다.

보기:

A. 기울기 소실 B. 기울기 폭발 C. 과적합 D. 병렬화 어려움

정답: A, B

해설:

RNN은 시간 역전파(BPTT) 시 동일 가중치가 반복 곱해지면서 기울기 소실/폭발 문제가 자주 발생한다.

기울기 소실: 시간이 멀어질수록 기울기가 0에 가까워져 장기 의존성을 학습하기 어렵다.

기울기 폭발: 기울기가 비정상적으로 커져 학습이 불안정해진다.

과적합: RNN만의 고유 한계라기보다는, 모든 머신러닝 모델이 겪을 수 있는 일반적인 문제에 해당된다.

병렬화 어려움: 시퀀스를 순서대로 처리해야 해서 GPU 병렬화가 어렵지만, 대표적인 한계로는 거리가 멀다.

확인문제

다음 중 Vision Transformer(ViT)의 특징에 대한 설명으로 옳지 않은 것은? (딥러닝 및 이미지 모델)

- ① 패치를 나눈 후 위치 인코딩을 통해 순서(위치) 정보를 추가한다.
- ② 인코더는 정규화, 다중헤드 어텐션, 연결층(MLP)으로 반복적으로 구성된다.
- ③ CNN과 달리 수용영역을 넓히지 않아도 전역 맥락을 학습할 수 있다.
- ④ 데이터가 작을 때도 CNN보다 항상 높은 성능을 보장한다.
- ⑤ 대규모 학습 데이터와 GPU 자원이 필요하다.

확인문제 정답 및 해설

다음 중 Vision Transformer(ViT)의 특징에 대한 설명으로 옳지 않은 것은? (딥러닝 및 이미지 모델)

- ① 패치를 나눈 후 위치 인코딩을 통해 순서(위치) 정보를 추가한다.
- ② 인코더는 정규화, 다중헤드 어텐션, 연결층(MLP)으로 반복적으로 구성된다.
- ③ CNN과 달리 수용영역을 넓히지 않아도 전역 맥락을 학습할 수 있다.
- ④ 데이터가 작을 때도 CNN보다 항상 높은 성능을 보장한다.
- ⑤ 대규모 학습 데이터와 GPU 자원이 필요하다.

정답: ④

해설:

ViT는 대규모 학습 데이터가 있을 때 강력한 성능을 발휘하지만, 데이터가 적은 경우에는 오히려 CNN보다 성능이 떨어질 수 있다. 따라서 "항상 높은 성능을 보장한다"는 설명은 잘못되었다.

다음 질문에 대해 서술하시오. (딥러닝 및 이미지 모델)

- ① 좋은 모델 구조를 골랐는데도 학습이 실패할 수 있는 이유는?
- ② Sigmoid 함수의 주요 문제는?
- ③ Leaky ReLU가 ReLU와 다른 점은 무엇인가?
- ④ ResNet에서 사용하는 전처리 방식은?

정답 및 해설:

- ① 구조가 좋아도 학습률, 정규화, 초기화 등 훈련 전략이 적절하지 않으면 성능이 떨어진다.
- ② 출력이 0이나 1에 너무 가까우면 기울기가 0이 되어 역전파가 끊긴다.
- ③ ReLU는 음수 입력에서 기울기가 0이 되지만, Leaky ReLU는 이를 방지하기 위해 작은 기울기를 둔다.
- ④ 단순 평균 Pooling보다 한 단계 더 나아가 스킵연결(Skip Connection)을 적용하여 학습 효율성을 높인다.

확인문제

다음 질문에 대해 서술하시오.

- ① 드롭아웃의 문제점을 설명하시오.
- ② 계단식 학습 비율 감소 방식의 단점은?
- ③ 학습 비율 선정의 기본 전략은?
- ④ 하이퍼파라미터를 선정하는 두 가지 방식은?

#### 확인문제 정답 및 해설

다음 질문에 대해 서술하시오.

- ① 드롭아웃 활용 시 유의할 점을 설명하시오.
- ② 계단식 학습 비율 감소 방식의 단점은?
- ③ 학습 비율 선정의 기본 전략은?
- ④ 하이퍼파라미터를 선정하는 두 가지 방식은?

#### 정답 및 해설

- ① 학습 시 실질적 출력 값이 줄어드나 테스트 시에는 뉴런을 전부 사용하기 때문에 학습-추론 간 분포 차이가 생겨 성능이 저하됨
- ② 변경 지점을 여러 개 선정하여 복잡하며 학습 곡선이 불안정할 수 있음
- ③ 학습 초반에는 큰 학습 비율로 손실 함수 표면의 넓은 영역을 빠르게 탐색, 이후 안정적 수렴을 위해 학습 비율을 줄임
- ④ 그리드 탐색은 하이퍼파라미터 후보들을 격자처럼 전부 조합하는 방식이고, 랜덤 탐색은 무작위로 값을 선택해 성능에 크게 영향을 주지 않는 축의 값만 선택해 시간을 줄이는 방식

#### 확인문제 - 자연어처리 및 텍스트 파운데이션 거대 모델

1 다음 중 거대 언어 모델과 BERT, GPT-1과 같은 기존 언어 모델의 차이점으로 올바른 것은?

- a) 다음 토큰 예측에 기반한 자기 지도 학습 방법
- b) 트랜스포머에 기반한 모델 구조
- c) 모델 및 학습 규모가 커지면서 나타난 창발성
- d) 이미지와 같은 도메인을 포함한 새로운 멀티 모달 학습 데이터



정답 및 해설

정답: c) 모델 및 학습 규모가 커지면서 나타난 창발성

해설: a), b)의 경우 기존 언어 모델들에서도 적용되고 있으며, 거대 언어 모델은 많은 양의 텍스트 데이터에 대해 다음 토큰 예측을 통한 사전 학습으로 학습 되므로 d)와 같은 이미지 데이터를 활용하지 않음

---

2 다음 중 지시 학습과 선호 학습에 대해 올바르지 않은 것은?

- a) 효과적인 지시 학습을 위해서는 다양한 지시 데이터가 필요하다
- b) 크기가 충분하지 않은 거대 언어 모델은 지시 학습 후에 성능이 떨어질 수 있다
- c) 보상 모델은 사람의 선호를 모방하도록 학습 된다
- d) 강화 학습 기반 선호 학습의 모든 과정에는 사람의 개입이 필요하다

정답 및 해설

정답: d) 강화 학습 기반 선호 학습의 모든 과정에는 사람의 지도가 필요하다

해설: 강화 학습 기반 선호 학습의 첫번째 과정 (지시 학습), 두번째 과정 (보상 모델 학습)에는 사람의 지도를 통한 데이터 제공이 필요하지만, 세번째 과정 (강화 학습)에서는 모델 스스로 응답을 생성하고 보상 모델을 통해 피드백을 받아 학습이 되므로 사람의 개입이 필요하지 않다.

---

3 다음 중 거대 언어 모델의 다양한 디코딩 알고리즘에 대한 설명으로 올바르지 않은 것은?

- a) Greedy Decoding은 가장 확률이 높은 토큰을 다음 토큰으로 생성한다
- b) Beam Search는 확률이 높은 응답 후보를 동시에 고려하기 때문에 더 많은 생성 비용을 필요로 한다
- c) 1 이상의 Temperature를 사용할 경우 더 다양한 응답이 생성될 확률이 감소 한다

d) Top-K Sampling은 생성 확률이 높은 K개의 토큰만을 생성 후보로 두고 확률 값에 따라 무작위로 다음 토큰을 생성한다

정답 및 해설

정답: c) 1 이상의 Temperature를 사용할 경우 더 다양한 응답이 생성될 확률이 감소 한다

해설: 1 이상의 temperature를 사용할 경우 생성 확률 분포가 기존에 비해 smooth 해지므로 (uniform에 가까워지므로) 더 다양한 토큰을 생성할 확률이 증가하고, 따라서 다양한 응답이 생성될 확률 또한 증가한다

확인문제-파인튜닝 및 AI모델 활용

1. 이진수  $1110_2$ 와  $0011_2$ 를 덧셈했을 때의 결과는?

- a.  $1110_2$
- b.  $0001_2$
- c.  $10000_2$
- d.  $10001_2$

2. 음수 표현 시 최상위 비트(MSB)를 부호 비트로 사용, 나머지 비트는 절댓값(크기)을 표현하는 방식은?

- a. 부호-크기 표현 방식
- b. 2의 보수 표현 방식
- c. 고정소수점 수 체계
- d. 부동소수점 수 체계

3. 인공지능 모델을 가볍게 만들어, 스마트폰 같은 작은 기기에서 효율적으로 사용할 수 있도록 하는 과정은?

- a. 모델 학습 (Model Training)
- b. 데이터 증강 (Data Augmentation)
- c. 모델 검증 (Model Validation)
- d. 모델 경량화 (Model Lightweighting)

4. 양자화(Quantization) 기법 중 모델 학습이 완료된 후에 양자화를 적용하는 방식은?

- a. QAT (Quantization Aware Training)
- b. PTQ (Post-Training Quantization)

- c. Weight-only Quantization
- d. Mixed-precision Quantization

---

#### 확인문제

- 5. 모델의 성능에 크게 기여하지 않는 불필요한 부분(가중치)을 제거하여 모델을 가볍게 만드는 기법은?
  - a. 양자화 (Quantization)
  - b. 가지치기 (Pruning)
  - c. 지식 증류 (Distillation)
  - d. LoRA
- 6. 크고 똑똑한 모델(선생님)의 지식을 가벼운 모델(학생)에게 가르쳐 주는 방식의 경량화 기법은?
  - a. 양자화 (Quantization)
  - b. 가지치기 (Pruning)
  - c. 지식 증류 (Distillation)
  - d. QLoRA
- 7. 대규모 언어 모델(LLM) 전체를 바꾸지 않고 필요한 부분만 '포스트잇'처럼 작은 변화를 주어 효율적으로 파인튜닝하는 기법은?
- 8. LoRA (Low-Rank Adaptation)가 파라미터 효율적 파인튜닝에 유리한 이유는?

---

#### 확인문제 정답 및 해설

##### 정답 및 해설

- 1. d.  $10001_2$
- 2. a. 부호-크기 표현 방식
- 3. d. 모델 경량화
- 4. b. PTQ (Post-Training Quantization)

5. b. 가지치기 (Pruning)
6. c. 지식 증류 (Distillation)
7. LoRA, QLoRA

(8번 정답) LoRA가 PEFT에 유리한 이유는 파라미터 업데이트에 저비용(랭크가 낮은 A, B만 학습)이기 때문임.

1.  $10001_2$ : 두 수  $1110_2(14)$ 와  $0011_2(3)$ 를 더하면  $10001_2(17)$ 이 됨.
2. 부호-크기 표현 방식: 최상위 비트(MSB)를 부호(0: 양수, 1: 음수)로 사용하고 나머지 비트들에는 절댓값을 그대로 표현. 예) 5의 4비트 표현  $0101 \rightarrow -5$ 는  $1101$ 로 표현.
3. 모델 경량화(Model Lightweighting): 가중치 수 축소, 정밀도(예: 32→8비트) 하향, 구조 단순화 등으로 파라미터 메모리와 연산량을 줄여 모바일/엣지 하드웨어에서도 지연과 전력 소모를 낮추면서 성능을 보존하는 과정.
4. PTQ(Post-Training Quantization): 전체 학습이 끝난 뒤(post-training) 모델의 가중치/활성값(캘리브레이션 데이터) 분포를 분석해 정밀도를 낮추는(양자화) 방식으로, 재학습 없이도 모델 크기와 추론 연산을 즉시 줄일 수 있다.
5. Pruning: 중요도가 낮은(가중치/채널/필터) 연결/뉴런을 제거해 파라미터 수와 FLOPs를 감소시켜, 성능 저하가 최소가 되도록 선택적으로 제거하는 기법.
6. 지식증류(Distillation): 큰 모델(Teacher)의 복잡한 정답('확신도' 정보) 혹은 분포 자체를 작은 모델(Student)이 모방해 학습하도록 보조함으로써, 단순 정답(hard label)로만 학습할 때보다 적은 파라미터로도 일반화 정보를 내재화하도록 돕는 학습 방법이다.
7. LoRA/QLoRA: 기본 가중치(W)는 고정된 채, 랭크가 낮은 두 행렬(A, B)의 곱을 더하여( $W+AB$ ) 그 부분만 학습. 기본 가중치를 4비트로 양자화해 메모리 효율을 높인 상태에서 어댑터만 16비트로 학습하는 방식이 QLoRA.

확인문제-Mastering AI Agent

1 다음 중 에이전트 기반 심층 연구의 핵심 특징으로 올바른 것은?

- a) 단순히 데이터를 수집하는 수동형 모델
- b) 검색 결과를 토대로 추론, 재조정하는 능동형 구조
- c) 학습 데이터를 분류하는 지도학습 모델
- d) 정적인 FAQ형 챗봇

2 에이전트 기반 심층 연구에서 LLM이 수행하는 역할로 가장 적절한 것은?

- a) 검색 결과의 단순 요약 및 번역
- b) 훈련 데이터를 수집하는 수동적인 도구
- c) 간단한 응답만을 생성하는 생성기
- d) 능동적인 정보 탐색 에이전트 역할

3 기존 RAG(검색 증강 생성)와 비교했을 때, AI 에이전트가 '오류 전파'를 줄일 수 있는 핵심적인 이유는?

- a) 추론을 먼저 수행한 후 검색을 한 번만 실행
- b) 검색 단계에서 더 많은 문서를 한 번에 가져옴
- c) 강화학습을 통해 항상 정답만 검색하도록 훈련
- d) 추론 과정이 언제, 무엇을 검색할지 동적으로 결정

4 에이전트의 문제 해결 사이클에서 <탐색> 단계의 주요 역할은?

- a) 질의에 대한 답변을 최종적으로 생성한다.
- b) 모델의 세부 목표를 설정한다.
- c) 웹페이지의 내용을 읽고 유용한 정보를 단기 저장소에 모은다.
- d) 에이전트 모델의 정책을 업데이트한다.

---

확인문제

5 멀티 에이전트 시스템의 효율성을 높이기 위한 구조적 장치로 옳은 것은?

- a) 한 번에 하나의 에이전트만 실행
- b) 리드 에이전트가 하위 에이전트를 병렬로 실행하고 결과를 조율
- c) 에이전트 간 통신을 차단하여 오류 감소
- d) 사용자의 직접 명령에만 반응

6 다중 에이전트 시스템의 한계로 제시된 '평가 및 디버깅의 어려움'의 주된 원인은?

- a) LLM의 토큰 제한 때문에
- b) 매 실행마다 에이전트의 경로와 답변이 달라 일관된 평가가 어렵기 때문
- c) 데이터셋이 부족하기 때문
- d) 강화학습 알고리즘이 적용되지 않기 때문

7 리드 에이전트가 수행하는 역할로 옳은 것을 모두 고르시오.

- a) 사용자의 질문을 분석한다.
- b) 연구 계획을 수립한다.
- c) 하위 에이전트의 실행 순서를 무작위로 변경한다.
- d) 하위 에이전트의 결과를 종합하여 추가 조사를 지시한다.

8 Test-Time Scaling(TTS)의 적용 예시로 올바른 것은?

- a) 모델의 크기를 줄여서 처리 속도를 높이는 방법
- b) 추론 단계수를 늘려 정확도를 향상시키는 Self-Refinement 과정
- c) 데이터셋을 나누어 병렬 처리하는 알고리즘
- d) 모델이 자동으로 사용자 피드백을 수집하는 과정

---

확인문제 정답 및 해설

정답 및 해설

- 1. b) 검색 결과를 토대로 추론, 재조정하는 능동형 구조
- 2. d) 능동적인 정보 탐색 에이전트 역할
- 3. d) 추론 과정이 언제, 무엇을 검색할지 동적으로 결정
- 4. c) 웹페이지의 내용을 읽고 유용한 정보를 단기 저장소에 모은다.

- 1. 에이전트 기반 심층 연구는 단순 응답 생성이 아닌 검색과 추론을 결합한 능동적 구조입니다. 검색 결과를 바탕으로 스스로 추론하고, 부족한 정보를 다시 검색하여 재조정합니다.
- 2. LLM은 단순 텍스트 생성기가 아니라 스스로 탐색하며 정보를 보강하

는 능동적 에이전트로서 작동합니다. 즉, 필요 시 외부 정보를 찾고 활용  
해 더 정확한 결과를 냅니다.

3. 기존 RAG는 "검색 후 추론"의 고정 순서를 따르지만, AI 에이전트는  
추론이 '언제, 무엇을 검색할지' 동적으로 결정합니다. 이로써 오류 전파  
를 줄이고 효율적인 검색이 가능해집니다.

4. 에이전트의 <탐색> 단계는 웹페이지를 읽고, 유용한 정보를 단기 저  
장소(memory)에 모으는 역할을 합니다. 필요 없는 페이지는 빠르게 건  
너뛰어 효율을 높입니다.

---

## 확인문제 정답 및 해설

### 정답 및 해설

5. b) 리드 에이전트가 하위 에이전트를 병렬로 실행하고 결과를 조율

6. b) 매 실행마다 에이전트의 경로와 답변이 달라 일관된 평가가 어렵  
기 때문

7. a,b,d

8. b) 추론 단계수를 늘려 정확도를 향상시키는 Self-Refinement 과정

5. 멀티 에이전트 시스템은 리드 에이전트가 여러 하위 에이전트를 '병  
렬' 실행(최대 90%)하고 결과를 종합합니다. 이는 협업 구조의 핵심 효  
율성 포인트입니다.

6. 멀티 에이전트의 평가가 어려운 이유는 실행할 때마다 경로와 답변이  
달라 일관된 평가가 불가능하기 때문입니다. 같은 입력에도 매번 다른  
대화 흐름이 생길 수 있습니다.

7. 리드 에이전트는 단순 명령 중개자가 아니라 (a)사용자의 질문을 분석  
하고 (b)연구 계획을 세우며 (d)하위 에이전트의 결과를 종합/조율하는  
핵심 역할을 합니다

8. Test-Time Scaling(TTS)은 추론 단계수를 늘려(Self-Refinement)성능  
을 향상시키는 접근입니다. 즉, 더 많은 사고 단계를 거쳐 더 정교한 결  
과를 내는 방식입니다.

## 확인문제-가속기 구동을 위한 100% 정수 연산 양자화

1 다음 설명이 맞으면 O, 틀리면 X를 선택하세요.

AI 가속기(NPU/TPU)를 효과적으로 사용하기 위해서는

모델의 가중치와 활성화 값만 정수로 양자화하면 충분하다.

확인문제 정답 및 해설

1 다음 설명이 맞으면 O, 틀리면 X를 선택하세요.

AI 가속기(NPU/TPU)를 효과적으로 사용하기 위해서는  
모델의 가중치와 활성화 값만 정수로 양자화하면 충분하다.

정답 및 해설

정답: X

해설

AI 가속기를 효과적으로 활용하려면 모델의 값 뿐 아니라 연산 과정도  
100% 정수 연산이어야 함.

---

확인문제

2 AI 모델의 값들이 모두 양자화 되었음에도 연산 과정이 양자화 되지  
않는 이유는 무엇인가?

- A. 오프셋 파라미터
- B. 활성화 함수 파라미터.
- C. 정규화 파라미터
- D. 스케일링 파라미터.

---

확인문제 정답 및 해설

2 AI 모델의 값들이 모두 양자화 되었음에도 연산 과정이 양자화 되지  
않는 이유는 무엇인가?

- A. 오프셋 파라미터
- B. 활성화 함수 파라미터.
- C. 정규화 파라미터
- D. 스케일링 파라미터.



정답 및 해설

정답: D

해설

양자화된 값들을 다시 실제 연산 범위로 복원하기 위해 사용되는 실수 파라미터로, 연산과정의 비양자화 원인.

---

확인문제

3 다음 중 연산 과정을 양자화 할 때 올바르지 않는 것은?

- A. 행렬 곱에서 스케일링 파라미터  $M = (S1 \times S2) \div S3$  를 그대로 사용하면 부동 소수점 연산이 발생한다.
- B. 실수 스케일링 파라미터를  $M = (2^{-n}) \times M0$  형태로 분해하여 비트 쉬프팅을 활용한다.
- C. 모든 실수 파라미터( $S1, S2, S3$ )를 하나의 값  $M$ 으로 합치기만 하면 정수 연산이 가능하다.
- D. 비트 쉬프팅( $\gg n$ )은 2의  $n$ 제곱 으로 나누는 연산을 빠른 정수 연산으로 대체한다.

---

확인문제 정답 및 해설

3 다음 중 연산 과정을 양자화 할 때 올바르지 않는 것은?

- A. 행렬 곱에서 스케일링 파라미터  $M = (S1 \times S2) \div S3$  를 그대로 사용하면 부동 소수점 연산이 발생한다.
- B. 실수 스케일링 파라미터를  $M = (2^{\text{의 } -n\text{제곱}}) \times M0$  형태로 분해하여 비트 쉬프팅을 활용한다.
- C. 모든 실수 파라미터( $S1, S2, S3$ )를 하나의 값  $M$ 으로 합치기만 하면 정수 연산이 가능하다.
- D. 비트 쉬프팅( $\gg n$ )은 2의  $n$ 제곱 으로 나누는 연산을 빠른 정수 연산으로 대체한다.

정답 및 해설

정답: C

## 해설

실수들을 계산해도 결과는 여전히 실수임. 반드시 비트 쉬프팅 같은 변환 작업이 필요함.

## 확인문제-테스트 타임 도메인 적용

1 TTA(Test-Time Adaptation)의 주요 목적은 무엇인가?

- A. 테스트 데이터의 라벨을 예측한 후 모델을 재학습하기 위함이다.
- B. 테스트 데이터의 분포를 파악하여, 모델이 새로운 환경에 적응하도록 하기 위함이다.
- C. 학습 데이터의 평균과 분산을 고정하여 과적합을 방지하기 위함이다.
- D. 테스트 데이터를 이용해 모델의 모든 가중치를 다시 초기화하기 위함이다.
- E. 학습 단계에서 모델의 복잡도를 줄이기 위해 사용된다.

---

## 확인문제 정답 및 해설

1 TTA(Test-Time Adaptation)의 주요 목적은 무엇인가?

- A. 테스트 데이터의 라벨을 예측한 후 모델을 재학습하기 위함이다.
- B. 테스트 데이터의 분포를 파악하여, 모델이 새로운 환경에 적응하도록 하기 위함이다.
- C. 학습 데이터의 평균과 분산을 고정하여 과적합을 방지하기 위함이다.
- D. 테스트 데이터를 이용해 모델의 모든 가중치를 다시 초기화하기 위함이다.
- E. 학습 단계에서 모델의 복잡도를 줄이기 위해 사용된다.

정답: B

## 해설

## 정답 및 해설

- TTA는 Test-Time Adaptation, 즉 "테스트 시점 적응" 기법이다.
- 모델이 학습 단계에서 본 데이터와 실제 환경의 데이터 분포가 다를 때, 테스트 데이터의 통계(분포)를 반영해 모델을 조정함으로써 성능 저하를

막는 것이 목적이다.

- 학습을 새로 하는 것이 아니라, 테스트 중에 일부 계층(주로 BatchNorm)을 테스트 데이터에 맞게 조정한다.

→ 즉, "테스트 환경 변화에 모델이 유연하게 적응하도록 만드는 기법"이다.

---

확인문제

1 다음 중 TENT와 SAR의 주요 차이점으로 옳은 것은?

- A. TENT는 학습 데이터를 이용해 BatchNorm을 업데이트하고, SAR은 테스트 데이터를 이용한다.
- B. SAR은 불확실성이 높은 샘플을 더 적극적으로 업데이트에 반영한다.
- C. TENT는 테스트 데이터의 엔트로피를 최소화하며, SAR은 그 과정을 더 안정적으로 수행하도록 개선한다.
- D. TENT는 파라미터 업데이트를 하지 않지만, SAR은 전체 네트워크를 다시 학습한다.
- E. 두 기법 모두 테스트 시점에 모델을 변경하지 않는다.

---

확인문제 정답 및 해설

1 다음 중 TENT와 SAR의 주요 차이점으로 옳은 것은?

- A. TENT는 학습 데이터를 이용해 BatchNorm을 업데이트하고, SAR은 테스트 데이터를 이용한다.
- B. SAR은 불확실성이 높은 샘플을 더 적극적으로 업데이트에 반영한다.
- C. TENT는 테스트 데이터의 엔트로피를 최소화하며, SAR은 그 과정을 더 안정적으로 수행하도록 개선한다.
- D. TENT는 파라미터 업데이트를 하지 않지만, SAR은 전체 네트워크를 다시 학습한다.
- E. 두 기법 모두 테스트 시점에 모델을 변경하지 않는다.

정답 및 해설

정답: C

## 해설

- TENT(Test-time Entropy Minimization)는 테스트 데이터를 입력받을 때, BatchNorm의 평균·분산(통계)과, (파라미터)를 업데이트하여 엔트로피를 최소화하는 방식으로 모델을 적응시킨다.

하지만 이 과정에서 노이즈가 큰 테스트 샘플이 들어오면 모델이 불안정해질 수 있다.

- SAR(Stable Adaptation Refinement)은 엔트로피가 낮은(확실한) 샘플만 선택해 업데이트하고, 엔트로피 변화가 너무 크지 않게 평탄화(loss smoothing)를 적용하여 적응 과정을 더 안정적으로 만드는 기법이다.  
→ 따라서 TENT는 적응의 기본형, SAR은 안정성 강화형이라고 할 수 있다.

---

## 확인문제

1 다음 중 TPT(Test-time Prompt Tuning) 기법의 핵심 아이디어로 가장 적절한 것은?

- A. 테스트 이미지 대신 학습 이미지를 사용해 프롬프트를 미리 튜닝한다.
- B. 테스트 이미지의 여러 증강본을 활용해 예측의 일관성을 높인다.
- C. 이미지 인코더의 통계값을 고정하여 모델을 안정화한다.
- D. 테스트 시점에서 모델 전체 파라미터를 다시 학습한다.
- E. 프롬프트 대신 모델의 가중치를 직접 업데이트한다.

---

## 확인문제 정답 및 해설

1 다음 중 TPT(Test-time Prompt Tuning) 기법의 핵심 아이디어로 가장 적절한 것은?

- A. 테스트 이미지 대신 학습 이미지를 사용해 프롬프트를 미리 튜닝한다.
- B. 테스트 이미지의 여러 증강본을 활용해 예측의 일관성을 높인다.
- C. 이미지 인코더의 통계값을 고정하여 모델을 안정화한다.
- D. 테스트 시점에서 모델 전체 파라미터를 다시 학습한다.
- E. 프롬프트 대신 모델의 가중치를 직접 업데이트한다.

정답 및 해설

정답: B

해설

- TPT는 테스트 이미지의 증강본 여러 개를 생성하고, 그중 엔트로피가 가장 낮은 샘플들을 이용해 텍스트 프롬프트를 조정하는 방식으로 예측 일관성을 향상시키는 기법이다.  
→ 즉, "증강을 통해 모델이 다양한 시점에서도 같은 답을 하도록 유도하는 것이 핵심이다."

---

확인문제

1 PromptAlign 기법의 주요 목표는 무엇인가?

- A. 테스트 이미지에 맞는 새로운 언어 모델을 학습한다.
- B. 프롬프트를 무작위로 변경해 테스트 데이터를 다양화한다.
- C. 학습 데이터와 테스트 데이터의 분포 차이를 줄이기 위해 통계적 정렬을 수행한다.
- D. 테스트 데이터의 레이블을 예측한 뒤 프롬프트를 재훈련한다.
- E. 학습 데이터만을 이용해 테스트 적응 없이 프롬프트를 미리 고정한다.

---

확인문제 정답 및 해설

1 PromptAlign 기법의 주요 목표는 무엇인가?

- A. 테스트 이미지에 맞는 새로운 언어 모델을 학습한다.
- B. 프롬프트를 무작위로 변경해 테스트 데이터를 다양화한다.
- C. 학습 데이터와 테스트 데이터의 분포 차이를 줄이기 위해 통계적 정렬을 수행한다.
- D. 테스트 데이터의 레이블을 예측한 뒤 프롬프트를 재훈련한다.
- E. 학습 데이터만을 이용해 테스트 적응 없이 프롬프트를 미리 고정한다.

정답 및 해설

정답: C

해설

- PromptAlign은 학습 데이터와 테스트 데이터의 분포 차이를 줄이기 위해

각 계층별 통계(평균, 분산)를 정렬하는 방식으로 프롬프트를 정렬한다.

- 학습 단계에서 저장한 통계와 테스트 단계에서 계산한 통계의 차이를 최소화하여

모델이 두 분포를 자연스럽게 일치시키도록 적응하게 만든다.

→라벨이 없어도 학습-테스트 간 분포를 정렬할 수 있는 비지도 TTA 기법이다.

1 초거대 AI의 스케일링 전략에 대한 설명으로 가장 옳지 않은 것은?

- a) AI 모델의 성능은 모델 크기, 학습 데이터 양, 컴퓨팅 자원이 증가함에 따라 향상되는 경향이 있다
- b) 분포 이동 문제를 해결하기 위해 모든 경우의 수를 학습 데이터로 준비하는 것은 현실적으로 가능하다
- c) 초거대 AI 전략은 "데이터 더! 모델 더 크게!"라는 방식으로 규모로 보임을 드는 접근이다
- d) 스케일링 법칙의 한계로는 비용 폭증, 환경 문제, 데이터 고갈 등이 있다

정답 및 해설

정답: b)

해설: 분포 이동 문제의 핵심은 세상의 모든 경우의 수를 학습 데이터로 미리 준비하는 것이 불가능하다는 점임. 이것이 바로 초거대 AI가 막대한 비용에도 불구하고 한계를 가지는 이유이며, 적응적 센싱과 같은 새로운 접근법이 필요한 이유

---

2 적응적 센싱(Adaptive Sensing)에 대한 설명으로 가장 올바른 것은?

- a) 모델의 파라미터 수를 증가시켜 더 많은 데이터 분포를 학습하는 기법이다
- b) 센서 파라미터를 조절하여 다양한 데이터를 모델이 학습한 영역(ID)안으로 이동시키는 기법이다
- c) 학습 데이터의 양을 늘려서 분포 이동 문제를 근본적으로 해결하는

기법이다

d) 모델 일반화와 동일한 개념으로, 학습 단계에서 적용되는 기법이다

정답 및 해설

정답: b)

해설: 적응적 센싱은 모델을 키우는 대신 센서(카메라 등)의 파라미터를 똑똑하게 조절하여, 데이터를 모델이 잘 인식할 수 있는 영역으로 이동시키는 기법. 이는 인간이 안경을 쓰는 것과 유사한 접근으로, 추론 단계에서 실시간으로 적용됨.

---

3 Lens(적응적 센싱 기법)와 초거대 모델의 성능 비교에 대한 내용으로 옳은 것은? (2-5 성능 효과 그림 2-8 참고)

- a) EfficientNet-B3에 Lens를 적용한 것이 OpenCLIP-h(초거대 모델)보다 50배 많은 파라미터를 사용하지만 더 효율적이다
- b) 적응적 센싱은 초거대 모델보다 항상 많은 학습 데이터가 필요하다
- c) 50배 작은 모델과 2,000,000배 적은 학습 데이터로도 적응적 센싱을 사용하면 초거대 모델보다 4~10% 더 높은 정확도를 달성할 수 있다
- d) 적응적 센싱은 모델 지식 영역(ID)을 확장하는 방식이고, 모델 일반화는 데이터를 ID 영역으로 이동시키는 방식이다

정답 및 해설

정답: c)

해설: EfficientNet-B3(Lens 적용)는 OpenCLIP-h보다 50배 작은 모델 파라미터와 2,000,000배 적은 학습 데이터를 사용했지만, ImageNet-ES(Luminous) 벤치마크에서 4~10%의 정확도 향상을 보임. 이는 적응적 센싱이 모델 크기와 데이터 양에 의존하는 기존 방식보다 훨씬 효율적임을 보여줌.

확인문제-응용 분야 전문 지식

1 수면다원검사의 근본적인 문제점과 이를 해결하기 위한 AI 개발의 필

요성에 대한 설명 중 옳바르지 않은 것은?

- a) 전문 수면기사가 AASM 가이드라인에 따라 시간 소모가 크고 노동 집약적으로 채점한다
- b) 서로 다른 채점자 간 동의율이 82%로, 첫날 밤 효과로 인한 측정 부정확성이 주요 문제이다
- c) 의료진의 진단결과를 신뢰하려면 AI는 자신의 판단에 대한 설명력 (explainability)을 제공해야 한다
- d) 멀티모달 신호의 종합적 분석을 위해서는 복잡한 멀티모달 시계열 신호 분석 AI 개발이 필요하다

---

확인문제 정답 및 해설

1 수면다원검사의 근본적인 문제점과 이를 해결하기 위한 AI 개발의 필요성에 대한 설명 중 옳바르지 않은 것은?

- a) 전문 수면기사가 AASM 가이드라인에 따라 시간 소모가 크고 노동 집약적으로 채점한다
- b) 서로 다른 채점자 간 동의율이 82%로, 첫날 밤 효과로 인한 측정 부정확성이 주요 문제이다
- c) 의료진의 진단결과를 신뢰하려면 AI는 자신의 판단에 대한 설명력 (explainability)을 제공해야 한다
- d) 멀티모달 신호의 종합적 분석을 위해서는 복잡한 멀티모달 시계열 신호 분석 AI 개발이 필요하다

정답 및 해설

정답: b)

해설: 첫날 밤 효과: 낮선 장소에서, 센서를 여러 개 부착하고 자면 평소와 수면 상태가 달라짐 (실제로 매우 불편함)

하룻 밤 문제: 평소에 편안하게 잘 때에도 수면 상태는 매일 밤 변함. 하룻 밤 결과는 평소의 평균적 수면 패턴을 놓칠 가능성이 있음

---

확인문제



2 수면다원검사 자동 채점 AI 개발 시 도메인 전문지식을 주입한 핵심 아이디어에 대한 설명 중 가장 적절하지 않은 것은?

- a) 시계열 신호 데이터를 이미지 포맷으로 변환하여 컴퓨터 비전 기법을 적용했다
- b) 비전 트랜스포머의 Positional embedding을 활용하여 30초 에폭 내 생체신호 위치를 모델링했다
- c) 에폭 간 맥락을 파악하기 위해 RNN(순환 신경망)을 사용하여 시간적 의존성을 학습했다
- d) 전문 수면기사가 여러 에폭의 수면단계를 종합적으로 판단하는 방식을 모방했다

---

확인문제 정답 및 해설

2 수면다원검사 자동 채점 AI 개발 시 도메인 전문지식을 주입한 핵심 아이디어에 대한 설명 중 가장 적절하지 않은 것은?

- a) 시계열 신호 데이터를 이미지 포맷으로 변환하여 컴퓨터 비전 기법을 적용했다
- b) 비전 트랜스포머의 Positional embedding을 활용하여 30초 에폭 내 생체신호 위치를 모델링했다
- c) 에폭 간 맥락을 파악하기 위해 RNN(순환 신경망)을 사용하여 시간적 의존성을 학습했다
- d) 전문 수면기사가 여러 에폭의 수면단계를 종합적으로 판단하는 방식을 모방했다

정답 및 해설

정답: c)

해설: 수면다원검사 자동 채점 AI는 에폭 간 맥락 파악을 위해 트랜스포머 인코더를 사용

---

## 확인문제

3 비접촉식 수면 무호흡증 진단 AI의 온디바이스 구현에 대한 설명입니다. 다음 중 옳은 것은?

- a) 서버 기반 클라우드 AI로 구현하여 개인정보 보호 문제를 해결했다
- b) 엣지 디바이스의 CPU에서 30초 이내에 실시간으로 한 에폭 클립을 처리할 수 있다
- c) 경량화 기법(pruning, quantization)을 적용하여 모델 크기를 줄였다
- d) 도메인 전문지식을 활용하여 복잡한 영상 AI 파이프라인을 온디바이스 AI로 해결했다

---

## 확인문제 정답 및 해설

3 비접촉식 수면 무호흡증 진단 AI의 온디바이스 구현에 대한 설명입니다. 다음 중 옳은 것은?

- a) 서버 기반 클라우드 AI로 구현하여 개인정보 보호 문제를 해결했다
- b) 엣지 디바이스의 CPU에서 30초 이내에 실시간으로 한 에폭 클립을 처리할 수 있다
- c) 경량화 기법(pruning, quantization)을 적용하여 모델 크기를 줄였다
- d) 도메인 전문지식을 활용하여 복잡한 영상 AI 파이프라인을 온디바이스 AI로 해결했다

## 정답 및 해설

정답: d)

해설: a) 서버 기반이 아니라 "엣지 디바이스(온디바이스)"에서 직접 처리하여 개인정보 보호 b) CPU에서도 가능하지만, 표에 따르면 GPU를 사용할 때 더 빠른 성능 c) 도메인 전문지식을 통한 감지 타겟 전환(무호흡→호흡각성)으로 경량 모델 사용이 가능