

## AI-05-02 딥러닝 Quiz 풀이

2024254012 배인호

1. 일반 신경망과 딥러닝 신경망에 대한 설명으로 옳지 않은 것은?

- 가. 일반 신경망인 다층 퍼셉트론에서 은닉층의 개수는 2~3개 정도로 적다.
- 나. 일반 신경망에서 특징 추출을 별도로 할 경우 개발자가 이를 처리해줘야 한다.
- 다. 딥러닝 신경망에서는 데이터에 대한 특징 추출 방법이 학습을 통해서 결정될 수 있다.
- 라. 기울기 소멸 문제는 계단 모양 활성화 함수를 사용하여 완화시킬 수 있다.

→ 계단 모양 함수를 사용하면 오히려 기울기 소멸문제가 더 심화될 수 있음.

2. 다음 신경망의 학습에 대한 설명으로 옳지 않은 것은?

- 가. 시그모이드 함수의 미분값은 0.25이하인 양의 값이기 때문에, 이를 활성화 함수로 사용하면 기울기 소멸 문제가 발생할 수 있다.
- 나. ReLU 함수의 출력값은 0이 될 수 없다. → 입력이 0보다 작을 때 출력값은 0이 된다.
- 다. 제이비어(Xavier) 기법을 사용할 때 가중치의 초기값은 층(layer)에 있는 노드 개수에 영향을 받는다.
- 라. 동일한 학습 알고리즘을 적용하더라도 신경망에 있는 가중치의 초기값은 성능에 영향을 크게 줄 수 있다.

3. 과적합에 대한 설명으로 옳지 않은 것은?

- 가. 과적합 상태이면 테스트 데이터에 대한 성능이 학습 데이터에 대한 성능보다 좋다.
- 나. 오차함수를 오차항과 모델 복잡도항으로 구성함으로써 과적합을 완화시킬 수 있다.
- 다. 학습할 때 일정 확률로 노드들을 무작위로 선택하여 해당 노드에 대한 연결선이 없는 것처럼 학습에 배제하는 드롭아웃은 과적합 해소에 도움이 된다.
- 라. 미니배치 단위로 가중치를 갱신하는 학습을 하면 과적합을 완화시키는데 도움이 될 수 있다.

→ 학습데이터 > 테스트 데이터 보다 성능이 좋다. 테스트 데이터 성능이 가장합시 안좋은 수 있음.

4. 다음 가중치 학습 기법에 대한 설명으로 옳지 않은 것은?

- 가. 딥러닝 신경망의 학습 알고리즘은 기본적으로 경사 하강법에 기반한다.
- 나. 모멘텀 사용 경사 하강법에서는 직전 시점의 가중치 갱신 정보를 일부 활용하여 가중치를 갱신한다.
- 다. Adadelata 알고리즘은 가중치별로 학습율에 다르게 적용될 수 있도록 한다.
- 라. 현재 알려진 가중치 학습 기법 중 ADAM이 가장 성능이 우수한 기법이다.

→ 학습기법의 선택은 문제의 특성, 데이터의 특성, 상황에 따라 달라질 수 있다.

5. 딥러닝 모델의 성능을 향상시키기 위해 사용하는 기법은 아닌 것은?

가. 데이터 확대

나. 정규화

다. 배치 정규화

라. 특성 삭제

→ 과적합을 방지하기 위해 사용하긴 하나 딥러닝은 가능한 많은 특성 사용이 일반적.

6. 일반신경망과 딥러닝 신경망에 대한 설명으로 옳지 않은 것은?

가. 딥러닝 신경망은 층의 수가 많을수록 과적합에 취약하다.

나. 일반 신경망은 과적합에 상대적으로 덜 취약하다

다. 딥러닝 신경망은 자동으로 특징을 추출한다.

라. 일반 신경망은 깊은 구조를 가진다.

→ 딥러닝 신경망

7. 일반 신경망과 딥러닝 신경망의 주요 차이점은?

가. 사용되는 뉴런의 수

나. 네트워크의 깊이 → 다층 (여러개의 은닉층을 가짐)

다. 활성화 함수의 종류

라. 그래디언트 소멸 문제의 발생

8. 기울기 소멸 문제는 주로 어떤 모델에서 발생하는가?

가. 얇은 신경망

나. 깊은 신경망 → 역전파 과정에서 기울기가 점점 작아져 가중치가 제대로 업데이트 되지 않음.

다. 단일 계층 신경망

라. SVM

9. 어떤 모델이 주로 복잡한 계층적 특징을 학습하는 데 적합한가?

가. 얇은 신경망

나. 깊은 신경망

다. 단일 계층 신경망

라. 선형 회귀 모델

10. 딥러닝 신경망에서는 주로 어떤 유형의 활성화 함수를 사용하여 그래디언트 소멸 문제를 완화하는가?

가. Sigmoid 함수

나. Step 함수

다. ReLU

라. 선형 함수

이 그래프는 ReLU 함수의 모양을 나타내며, 기울기 소멸 문제를 완화한다.

11. 기울기 소멸 문제가 발생할 때, 가중치 업데이트는 어떤 특성을 보이는가?

가. 지나치게 큰 업데이트

나. 매우 작은 업데이트 - 가중치가 점점 작아짐

다. 무작위한 업데이트

라. 일정한 크기의 업데이트

12. 기울기 소멸 문제는 어떤 활성화 함수와 함께 발생할 가능성이 높은가?

가. ReLU

나. Leaky ReLU

다. Sigmoid - 미분값 최댓값 0.25로 가중치 업데이트가 제한될 수 있음.

라. Softmax

13. 어떤 활성화 함수가 입력 값의 모든 부분에서 미분 가능한가?

가. ReLU

> 0에서 미분 불가.

나. Leaky ReLU

다. Sigmoid - 모든 부분에서 미분 가능함

라. Step function

14. 기울기 소멸 문제가 주로 발생하는 경우는?

가. 층이 많은 신경망에서

나. 학습률이 너무 높을 때

다. 가중치가 무작위로 초기화될 때

라. 활성화 함수로 ReLU를 사용할 때

15. 기울기 소멸 문제를 일으키는 활성화 함수는?

가. ReLU

나. 시그모이드 - 위의 문제 발생 ↑

다. Leaky ReLU

라. ELU

16. 기울기 소멸 문제를 완화하는 방법 중 가장 잘못된 것은?

가. 적절한 가중치 초기화 방법 사용

나. 활성화 함수로 ReLU 계열 함수 사용

다. 학습률을 점차 증가시키는 방법

라. 배치 정규화(Batch Normalization)

★ 학습률이 높으면 불안정한 수 있음



17. 신경망에서 초기 가중치 설정의 중요성에 대한 잘못된 설명은?

가. 초기 가중치는 학습의 시작점을 결정한다.

나. 너무 큰 초기 가중치는 학습 과정을 불안정하게 만들 수 있다.

다. 모든 가중치를 같은 값으로 초기화하는 것이 바람직하다 (X)

라. 초기 가중치는 무작위로 설정되거나 특정 기법을 사용하여 초기화된다.

→ 신경망이 깊은 경우 초기화하는 것이 어려움.

18. 신경망에서 배치 크기(Batch Size)의 역할에 대한 잘못된 설명은?

가. 한 번에 처리하는 데이터의 수를 결정한다.

나. 너무 크면 학습 속도가 느려질 수 있다.

다. 아주 작게 하면 모델의 일반화 성능이 향상된다. (X)

라. 메모리 용량과 계산 속도에 영향을 준다.

19. 신경망에서 손실 함수(Loss Function)의 목적에 대한 잘못된 설명은?

가. 예측값과 실제값 간의 차이를 측정한다. (O)

나. 학습 과정에서 최소화되어야 한다.

다. 손실 함수의 선택은 문제 유형에 따라 달라진다.

라. 손실 함수의 값은 학습 과정에 영향을 주지 않는다.

→ 손실값 vs 예측값 차이 측정. 0에 가까울수록 좋음.

차이를 최소화하는 방법이다. 가중치 업데이트. 성능에 영향을 많이 줌.

20. 가중치 초기화에서 0으로 초기화하는 방법의 문제점은?

가. 신경망이 대칭적인 가중치를 갖게 되어 학습이 제대로 이루어지지 않는다.

나. 0으로 초기화하면 뉴런의 출력이 항상 0이 되어 학습이 이루어질 수 없다.

다. 가중치 감소(weight decay) 기법을 적용할 때 문제가 발생할 수 있다.

라. 모든 가중치가 같은 값을 갖기 때문에 기울기 소멸 문제가 발생한다.

21. Xavier 초기화 방법의 주된 특징은?

가. 가중치를 일정 범위 내에서 균등하게 분포시키는 방법이다.

나. 초기 가중치를 큰 값으로 설정하여 활성화 함수의 비선형 영역을 활용한다.

다. 각 층의 입력 노드의 개수에 따라 분산의 크기를 조정한다.

라. 무작위성을 최대화하여 모든 가중치가 다른 값을 갖도록 한다.

→ 신경망의 각 층에서 출력값이 분산이 입력 값의 분산과 비슷하게 유지되도록 함. 신경망 학습 안정화.

22. He 초기화 방법에서 사용하는 분산의 계산에 사용되는 요소는?

가. 은닉층의 노드 수

나. 학습률

다. 각 층의 입력 노드 수

라. 가중치의 개수

→ 입력 노드 수를 고려하여 가중치 크기 분산을 설정.

23. L1 규제화는 어떤 특징을 가지고 있는가?

가. 가중치의 제곱을 규제

나. 가중치의 절댓값을 규제

다. 가중치의 합을 규제

라. 가중치의 제곱근을 규제

→ 절댓값 바뀌는 손실함수를 추가.

24. 규제화는 어떤 상황에서 효과적인가요?

가. 훈련 데이터가 충분할 때

나. 모델이 너무 단순할 때

다. 모델이 너무 복잡할 때

라. 훈련 데이터가 아주 적을 때

25. 규제화를 사용할 때 주의해야 할 점은?

가. 너무 강한 규제는 과소적합을 일으킬 수 있음

나. 규제는 항상 모델의 성능을 향상시킴

다. 규제는 항상 모델의 학습 속도를 늦춤

라. 규제는 모든 문제에 적합하다

26. 배치 정규화(Batch Normalization)의 주된 목적은?

가. 과적합 방지

나. 학습 속도 개선

다. 모델의 계산 복잡도 감소

라. 모델의 크기 축소

27. 배치 정규화는 어느 위치에 주로 적용되는가?

가. 활성화 함수 전

나. 활성화 함수 후

다. 손실 함수 전

라. 최적화 함수 후

→ 평균, 분산, 정규화된 값, 비정규화된 값, 이동평균, 이동 분산.

28. 배치 정규화에서 계산할 필요가 있는 것은?

- 가. 평균과 분산
- 나. 가중치와 편향
- 다. 학습률과 모멘텀
- 라. 손실 값과 그래디언트 값

29. 배치 정규화는 어떤 값을 이용해 정규화하는가?

- 가. 한 배치 내의 데이터들의 중앙값
- 나. 전체 데이터의 평균값
- 다. 한 배치 내의 데이터들의 평균값
- 라. 전체 데이터의 중앙값

→ 스케일 조정, 이동, 정규화

30. 배치 정규화에서 발생하는 연산이 아닌 것은?

- 가. 스케일(scale) 조정
- 나. 이동(shift)
- 다. 회전(rotation)
- 라. 정규화(normalization)

정답 3개.

31. 배치 정규화의 장점은?

- 가. 더 큰 학습률을 사용할 수 있게 해줌
- 나. 가중치 초기화의 영향을 줄여줌
- 다. 내부 공변량 변화를 완화
- 라. 모델의 크기를 물리적으로 줄여줌

☆

32. 경사하강법(Gradient Descent)의 핵심 아이디어는?

- 가. 손실 함수의 최대값을 찾는 것
- 나. 모든 가중치를 무작위로 조정하는 것
- 다. 손실 함수의 그래디언트를 사용하여 가중치를 업데이트하는 것
- 라. 모든 샘플에 대해 개별적으로 가중치를 업데이트하는 것



33. Momentum 방법을 사용할 때, 어떤 변수가 이전 그래디언트 값을 사용하는가?

가. 학습률

나. 가중치

다. 손실 함수

라. 속도(velocity)

34. AdaGrad의 특징은?

가. 과거의 모든 그래디언트의 제곱을 누적한다.

나. 속도와 방향을 모두 고려한다.

다. 학습률을 일정하게 유지한다.

라. 모멘텀만을 사용하여 가중치를 업데이트한다.

35. RMSProp의 핵심 아이디어는?

가. RMSProp + AdaGrad

나. RMSProp + Momentum

다. Gradient Descent + Momentum

라. AdaGrad + Gradient Descent

학습률이 너무 빨리 감소하는 문제를 완화.

36. 미니배치 경사하강법은 어떤 방식으로 데이터를 처리하는가?

가. 한 번에 하나의 샘플만 처리한다.

나. 전체 데이터 세트를 한 번에 처리한다.

다. 데이터 세트의 작은 부분집합을 한 번에 처리한다.

라. 데이터의 순서를 고려하지 않고 무작위로 처리한다.

37. 어떤 최적화기가 학습률을 동적으로 조정하는 기능을 가지고 있는가?

가. Gradient Descent

나. Momentum

다. AdaGrad

라. Batch Gradient Descent

38. 최적화기(optimizer)의 학습률을 너무 높게 설정하면 어떤 문제가 발생할 수 있는가?

가. 모델이 과소적합될 수 있다.

나. 손실 함수의 최소값에 수렴하기 위해 더 많은 시간이 걸릴 수 있다.

다. 손실 함수의 최소값 주변에서 진동할 수 있다.

라. 학습이 전혀 진행되지 않는다.

→ 학습률이 지나칠 수 있음.

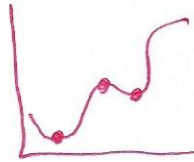
39. saddle point(안장점)에 빠질 경우, 어떤 최적화기가 이를 해결하는 데 가장 도움이 안될까?

가. Gradient Descent

나. Momentum

다. Batch Gradient Descent

라. AdaGrad



40. 학습 데이터의 특징 Z가 평균 0, 표준 편차 1로 정규화되어야 할 때, 원본 데이터 값 [2, 8, -1, 4]를 정규화하고 변환된 값을 계산하세요. (Z-Score 정규화)

각 데이터 포인트에서 평균을 빼고, 데이터의 표준 편차로 나눔.

$$\frac{2+8-1+4}{4} = 3.25 \quad \hookrightarrow \sqrt{\frac{(2-3.25)^2 + (8-3.25)^2 + (-1-3.25)^2 + (4-3.25)^2}{4}}$$

$$\sqrt{\frac{42.95}{4}} = 3.26917 \approx 3.3$$

계산식

$$\hookrightarrow Z = \frac{X - \mu}{\sigma}$$

$$z = \frac{(2-3.25)}{3.3} = -0.37$$

$$z = \frac{(8-3.25)}{3.3} = 1.44$$

$$z = \frac{(-1-3.25)}{3.3} = -1.29$$

$$z = \frac{(4-3.25)}{3.3} = 0.23$$



41. 학습률이 0.01이고, 현재 가중치가 0.5일 때, 손실 함수의 그래디언트가 0.3으로 계산되었을 때, 다음 단계의 가중치를 계산하시오.

$$W_{\text{new}} = W_{\text{old}} - \eta * \nabla J$$

(현재가치)  $W_{\text{old}} = 0.5$

(학습률)  $\eta = 0.01$

(손실함수 그래디언트)  $\nabla J = 0.3$

$$W_{\text{old}} \quad \eta \quad \nabla J$$

$$0.5 - 0.01 * 0.3 =$$

$$0.5 - 0.003 = 0.497$$

42. 데이터 배치가 [1, 3, 5, 7]로 주어졌을 때, 이 데이터의 평균과 표준편차를 계산하시오.

$$\mu(\text{평균}) = \frac{(1+3+5+7)}{4} = \frac{16}{4} = 4$$

$$\sigma(\text{표준편차}) = \sqrt{\frac{(1-4)^2 + (3-4)^2 + (5-4)^2 + (7-4)^2}{4}}$$

$$\sqrt{\frac{9+1+1+9}{4}} = \sqrt{\frac{20}{4}} = \sqrt{5}$$

$$\approx 2.236$$

43. 위에서 계산한 평균과 표준편차를 사용하여 각 데이터 포인트를 정규화하시오.

$$Z = \frac{x - \mu}{\sigma}$$

$$Z = \frac{(1-4)}{2.236} = -1.34$$

$$Z = \frac{(3-4)}{2.236} = -0.45$$

$$Z = \frac{(5-4)}{2.236} = 0.45$$

$$Z = \frac{(7-4)}{2.236} = 1.34$$

44. 위에서 정규화된 데이터에 학습 가능한 매개변수  $\gamma = 2$ 와  $\beta = 0.5$ 를 적용하여 출력을 계산하시오.

$$\text{출력} = \gamma * Z + \beta$$

$$2 \times -1.34 + 0.5 = -2.18$$

$$2 \times -0.45 + 0.5 = -0.4$$

$$2 \times 0.45 + 0.5 = 1.4$$

$$2 \times 1.34 + 0.5 = 3.18$$