

14. Activation function 함수를 사용하는 이유? Softmax, Sigmoid 함수의 차이?

활성화 함수는 모델에 비선형성을 부여하여 신경망이 복잡한 문제 (비선형 문제)를 학습할 수 있게 한다. 선형함수만 사용하면 층을 쌓아도 하나의 선형 모델과 같아 의미가 없다

Sigmoid 함수는 0.0 ~ 1.0 사이의 연속적인 실수를 출력하며, 이진 분류에 사용된다. 반면 Softmax 함수는 다중 클래스 분류에 사용된다. 각 클래스의 확률로 변화되며, 총합이 1이 되도록 조정된다

15. Forward Propagation, Backward Propagation이란?

Forward propagation은 입력 데이터를 받아 가중치와 편향을 통해 예측값을 계산하는 과정이며, 출력값을 생성한다. Backward propagation은 예측값과 실제값의 오차를 기반으로 오차를 각 층으로 전파하며 가중치의 변화량을 계산하는 과정이다. Chain Rule (연쇄법칙)을 사용해 미분한다.

16. 손실함수란 무엇인가? 가장 많이 사용하는 손실함수 4가지 종류는?

손실함수 (Loss Function)은 모델의 예측값과 실제값 사이의 차이를 수치로 나타낸 것이다. 모델 학습의 목표는 이 손실을 최소화하는 것이다

대표 손실함수 4가지로는 MSE, MAE, Binary Cross-Entropy, Categorical Cross-Entropy가 있다. MSE는 오차를 제곱한 값들의 평균으로 회귀 문제에 사용되며, MAE는 오차의 절대값의 평균으로 역시 회귀 문제에 사용된다. Binary Cross-Entropy는 이진 분류 문제에, Categorical Cross-Entropy는 다중 클래스 분류에 사용하는 손실함수다

17. 옵티마이저 (Optimizer)란 무엇인가? 옵티마이저와 손실함수의 차이점은?

옵티마이저는 손실함수의 값을 줄이기 위해 가중치들을 업데이트하는 역할을 하며, 알고리즘과 경사하강법을 구현하는 것이다. 대표적으로 SGD, Adam, RMSprop 등이 있다. 손실함수는 실제값과 예측값이 얼마나 틀렸는지 평가한다면, 옵티마이저는 그 차이를 어떻게 줄일지 방향을 결정하고 가중치를 조정한다

18. 경사하강법 의미는? (확률적 경사하강법, 배치 경사하강법, 미니 배치 경사하강법)

손실함수 값을 최소화 하는 가중치를 찾는 최적화 알고리즘이다. 배치 경사하강법은 전체 데이터를 한 번에 학습하고 가중치를 변경하는 방법으로, 안정적이지만 느리다. 확률적 경사하강법은 무작위로 샘플 하나를 선택하여 가중치를 변경하는 방법으로, 빠르지만 불안정한 진동이 크다. 미니 배치 경사하강법은 훈련 샘플들을 작은 배치들로 분할한 후 배치 단위로 가중치를 계산하는 방법으로, 속도와 안정성의 균형을 갖춘 방법이다.

21. CNN의 합성곱의 역할은?

합성곱(convolution)을 통해 이미지의 공간적 특성(모양, 윤곽, 패턴) 등을 추출할 수 있다. CNN은 이미지 전체를 한 번에 보지 않고, 지역적 특징(local feature)을 감지한다.

22. CNN의 풀링층의 역할은?

풀링층은 특징 맵의 크기를 줄이고 연산량을 감소시켜 계산 효율성을 높인다. 중요한 특징만 남기며 레이어의 크기를 줄이기 때문에 입력 데이터를 요약하는 효과가 있다. 레이어의 크기가 작아진다는 것은 매개변수가 작아진다는 것을 의미하므로 과잉적합을 방지할 수 있다. 또한 물체의 공간이동에 대해 둔감하게 만든다. 풀링의 종류로는 최대 풀링, 평균 풀링, 가중치 풀링 등이 있다.

23. CNN의 Dense Layer의 역할은?

Dense Layer는 CNN의 마지막 단계에 위치하며, 추출된 특징을 기반으로 최종 예측을 수행하는 역할을 한다. 즉, 모델의 최종 분류 결과를 산출한다. Dense Layer는 Fully Connected Layer로, 모든 뉴런이 이전 층과 연결된다.

24. CNN의 stride, filter의 역할? 필터의 크기는 어떻게 결정되는가?

stride는 필터가 이미지 위를 이동하는 간격이다. 보폭의 크기가 크면 출력 크기가 줄어드는데, 일반적으로 stride를 k 로 설정하면 출력은 $1/k$ 로 줄어든다. 좀 더 구체적으로는 수평 보폭 sh , 수직 보폭 sv 로 나눌 수 있다.

filter는 입력 이미지에서 특징을 추출하는 역할로, 가중치가 저장된 2차원 배열이다. 필터는 커널, 마스크라고 불리며, 여러 개의 필터는 각각 다른 특징을 감지한다.

영상 처리에서는 필터의 크기가 미리 결정되지만, 권역추선 신경망에서는 학습을 통해 필터의 크기가 자동으로 업데이트 된다. 역전파를 통해 손실을 줄이는 방향으로 조정되며, CNN에서 어떤 필터가 어떤 특징을 추출하는지는 알 수 없다.

25. RNN을 사용하는 이유와 한계점은?

RNN은 시계열 데이터나 문장처럼 순서가 중요한 데이터를 처리하기 위해 사용된다. RNN은 수많은 저층의 행렬이 존재하기 때문에 역전파(BPTT) 과정에서 행렬과 행렬의 곱셈의 대부분이 여러 번 공해지며 그라디언트 소실 문제가 발생할 수 있다. 그라디언트가 소실되면 학습이 진행되어도 먼 거리의 의존 관계는 파악하지 못하고 근거리의 의존 관계만을 중시하게 된다. 반대로 그라디언트 폭증 문제도 발생할 수 있으며, 병렬 처리가 불가능해 학습 속도가 느리다는 한계가 있다.



26. LSTM을 사용하는 이유와 한계점은?

LSTM은 그래디언트 소실로 인해 장기 문맥을 기억하지 못하는 RNN의 장기 의존성 문제를 해결한다. 게이트 구조와 장기기억을 선택에 저장하는 방식으로 중요 정보는 기억하고 불필요한 정보는 잊는다. LSTM에서는 셀 상태 C_t 를 이용하여 그래디언트가 끊기지 않고 역방향으로 흐를 수 있다. LSTM의 한계점으로는 복잡한 구조와 높은 학습 비용, RNN보다 느린 속도가 있다.

27. GRU를 사용하는 이유와 차별성은?

GRU는 LSTM보다 간단한 구조이면서도 비슷한 성능을 낸다. GRU는 LSTM보다 게이트 수가 적고 계산이 간단하기 때문에 학습 속도가 빠르며, LSTM보다 더 적은 파라미터로 동작한다는 차별성이 있다.

33. Dense Layer란 무엇인가?

Dense Layer는 신경망의 기본적인 완전 연결 계층(Fully Connected Layer)이다. 이전 층의 모든 뉴런과 연결되어 있으며, 입력의 모든 특징을 고려한다. CNN, RNN 등에서 마지막 분류기 역할을 수행한다.