

# 인공지능개론 과제 (기말)

2020335기 유수빈

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

활성화 함수를 쓰는 가장 큰 이유는, ~~레이어를 쌓아도~~ 활성화 함수가 없으면 모든 레이어가 선형 변환만 하게 때문에 레이어를 많이 쌓아도 1차 함수(선형 함수)로 바뀐다. 이러한 선형 함수는 복잡한 패턴을 배우지 못한다. 따라서 비선형 함수(활성화 함수)를 두어서 고차원 특징, 더 다양한 구조를 하기 위해 쓴다.

대표적인 활성화 함수로는 softmax와 sigmoid가 있는데, sigmoid 함수는 각 출력 0과 1 사이의 연속값으로 매핑하며, 이진 분류 문제나 출력 노드가 1개일 때 사용된다. 출력 값들은 서로 독립적이며, 출력의 총합이 1이 되지 않아도 된다. 반면 softmax 함수는 다중 클래스 분류에 사용되며, 각 클래스에 대한 출력값을 전체 클래스에 대한 확률 분포로 변환한다. 각 출력의 합은 1 사이이며, 전체 확률의 합은 1이 된다. 클래스 중 하나를 선택하는 상황에 적합하다.

## #2. Forward propagation, Backward propagation이란?

Forward propagation(순전파)는 인공신경망에서 입력 데이터를 입력층에서 출력층까지 전달하며 예측값을 계산하는 과정이다. 이때, 각 층의 가중치와 편향을 이용하여 선형 변환을 수행하고, 활성화 함수를 통해 비선형성을 추가하여 최종 출력을 얻는다. Back propagation은 순전파로 계산된 예측값과 실제값의 차이를 최소화하기 위해 가중치와 편향을 조정하는 과정이다. 손실 함수를 통해 오차를 계산하고, 체인 룰을 이용하여 기울기를 구한 뒤 경사하강법으로 가중치를 업데이트 한다.

## #3. 손실 함수란 무엇인가? 가장 많이 사용하는 손실 함수 4가지는?

손실 함수란 모델이 예측한 값과 실제 값의 차이를 계산하여 모델의 성능을 평가하는 함수이다.

종류는 다음과 같다.

1. 평균 제곱 오차(MSE) : 회귀 문제에서 주로 사용하며, 예측값과 실제값의 차이를 제곱하여 평균한 값.
2. 평균 절대 오차(MAE) : 예측값과 실제값 차이의 절대 값을 평균한 값. 이상치의 영향을 적게 받는다.
3. 교차 엔트로피 : 분류 문제에서 사용, 예측 확률과 실제 정답 확률 간의 차이를 계산하여 모델의 불확실성을 평가한다.

4. 이진 교차 엔트로피 : 이진 분류에서 사용하며, 두 클래스(0 또는 1) 사이의 확률 차이를 계산하여 손실을 구한다.



#### #4. 손실함수와 오티마이저 차이? 오티마이저란?

오티마이저는 경사하강법과 같은 최솟값을 알고리즘을 구현한 함수이다. 예를 들어 Adam은 경사하강법을 개선하여 더 효율적으로 학습할 수 있도록 구현한 대표적인 오티마이저이다.

경사하강법이 알고리즘이라면, 오티마이저는 이를 구현하는 구체적인 구현체이다.

손실함수는 모델이 예측한 값과 실제 값의 <sup>차이를</sup> 비교하여 성능을 평가하는 역할을 한다. 반면 오티마이저는 손실값을 최소화하기 위해 가중치나 편향을 어떻게 업데이트할지 결정하는 역할을 한다.

#### #5 경사하강법의 의미는? (확률적, 배치, 미니 경사하강법)

경사하강법은 손실을 최소화하기 위해 가중치와 편향을 조정하는 알고리즘이다.

모델이 예측한 값과 실제 값 차이를 계산하여 손실 함수를 통해 손실 값을 구하고, 이 손실을 최소화하기 위해 가중치와 편향을 어떻게 업데이트할지 결정한다.

경사하강법에는 다음과 같이 3가지 주요 방식이 있다.

1. 배치 경사하강법: 전체 데이터를 사용하여 가중치를 계산한다. 안정적이지만 느리다.
2. 확률적 경사하강법: 샘플 하나씩 가중치를 계산한다. 빠르지만 진동이 크다.
3. 미니배치 경사하강법: 소규모 ~~데이터~~ 묶음 단위로 계산한다. 속도와 안정성의 균형이 있다.

#### #6 CNN의 합성곱의 ~~역할~~ 역할은?

합성곱은 입력 이미지와 필터를 곱하면서 특징을 뽑는 연산이다. 예를 들어 가로자리, 선, 곡선, 패턴 같은 사각형 특징을 찾아낸다. 필터 1개는 1가지 특징을 탐지하는 역할을 한다. 필터 1개당 1개의 특징만 볼 수 있기 때문에 여러 개의 필터를 사용한다. 필터가 많을수록 더 다양한 시각 정보를 학습할 수 있다. 또한 작은 필터를 겹쳐서 같이 있는 정보를 추출한다. 필터를 많이 쓰고, 여러층을 쌓으면 계산량도 많아진다. 그런 때는 pooling이나 stride 같은 기법으로 연산량을 줄이기도 한다.

#### #7 CNN의 풀링(Pooling)층의 역할은?

풀링은 특징 맵의 크기를 줄여서 연산량을 줄인다. 즉, 복잡한 정보를 간소화하여 다음층에 전달한다. 이미지나 텍스트의 위치가 약간 변해도 동일한 결과를 내도록 돕는다 (위치 변화에 강함). 종류로는 다음과 같다.

1. Max Pooling: 일정 영역에서 최대값 추출. 가장 강한 특징만 남김.
2. Average Pooling: 일정 영역의 평균값 추출. 전체 특징을 평균적으로 반영함.



#8. CNN의 Dense Layer의 역할은?

합성곱과 풀링층에서 추출한 특징을 기반으로 최종 출력을 계산하는 역할을 한다. 모든 노드가 서로 연결되어 특징들을 종합적으로 학습하며, 주로 분류나 레이블 같은 최종 예측 단계에서 사용된다. 비선형성을 추가하여 복잡한 패턴을 학습하고, 이전 층에서 얻은 특징을 바탕으로 최종 결론에 도달한다.

#9. CNN의 stride, filter 역할? 필터의 가중치는 어떻게 결정되는가?

필터는 입력 데이터에서 특징을 추출하는 작은 행렬로, 이미지의 로컬 특징을 추출한다. 필터의 값은 학습을 통해 최적화된다.

Stride는 필터가 이동하는 간격이다. Stride가 크면 출력 특징 맵의 크기가 줄어들고 계산량이 감소한다.

필터의 가중치는 처음엔 무작위로 초기화되지만, 학습 과정에서 역전파를 통해 손실 함수를 최소화 하도록 자동으로 조정된다.

#10. RNN을 사용하는 이유와 한계점은?

RNN은 순서가 있는 시계열 데이터나 문장 처리 연속된 데이터를 처리하기 위해 사용된다. 이전 단계의 정보를 기억하여 시간에 따른 문맥과 관계를 학습할 수 있다는 것이 장점이다. 그러나 긴 시퀀스 문제를 학습할 때 정보가 점차 사라지거나 변형되는 장기 의존 문제나 계산 속도가 느리다는 한계가 있다.

#11. LSTM을 사용하는 이유와 한계점은?

LSTM은 RNN 장기 의존 문제를 해결하기 위해 고안된 구조로, 긴 시퀀스에서도 중요한 정보를 오래 기억하고 불필요한 정보는 잊어버리게 한다. 덕분에 긴 문맥이나 복잡한 시계열 데이터를 잘 처리할 수 있어 자연어 처리, 대화형 데이터나 매우 긴 시퀀스에서는 여전히 한계가 있을 수 있다.

#12. GRU를 사용하는 이유와 차별성은?

GRU는 LSTM의 복잡한 구조를 간소화한 RNN 변형 모델로, 긴 시퀀스에서도 장기의존성을 잘 학습하면서 계산 효율이 높다. 그러나 게이트가 적고 구조가 단순해 학습 속도가 빠르고 메모리 사용량이 적지만 복잡한 패턴 학습에서는 LSTM보다 성능이 떨어질 수 있다.

#13. Dense Layer란 무엇인가?

신경망에서 모든 입력 노드가 모든 출력 노드와 연결된 층을 말한다. 주로 신경망의 마지막 부분에서 특징을 종합해 최종 예측을 만드는 역할을 한다.