

## 6. Additional Problem: DNN, CNN, RNN (8)

6-1. 경사 하강(Gradient Descent)법의 학습 속도를 개선하기 위한 대표적 방법에는 모멘텀(Momentum)과 적응형 학습률(Adaptive Learning Rate)이 있다. 이 중, 모멘텀 방식이 경사 하강법에 어떤 개념을 도입하여 최적화 과정을 개선하는지 설명하라. 또한, 이 방법이 왜 필요한지, 어떤 문제를 해결하려고 하는지에 대해 서술하라.

**답:**

모멘텀 방식은 물리학의 운동량 개념을 경사 하강법에 도입하여 최적화 과정을 개선하는 방법입니다. 이 방법은 이전 단계의 그래디언트 방향을 일정 비율로 현재 업데이트에 반영함으로써 학습 과정에 관성을 부여합니다. 이를 통해 학습 과정이 지속적으로 같은 방향으로 진행될 때 더 빠르게 이동하고, 방향이 자주 바뀌는 경우에는 진동을 감소시키는 효과를 얻을 수 있습니다.

모멘텀 방식은 특히 경사가 가파른 협곡이나 지역 최소값이 많은 복잡한 손실 함수 지형에서 필요합니다. 이 방법은 지역 최소값에 빠지는 문제를 완화하고, 안장점에서의 느린 수렴 문제를 해결하는 데 효과적입니다. 또한 그래디언트의 방향이 자주 변하는 경우에 발생하는 지그재그 현상을 줄여 더 빠르고 안정적인 수렴을 가능하게 합니다.

6-2. Gradient Vanishing 문제의 해결방안을 제시하라.

**답:**

Gradient Vanishing 문제 해결을 위한 첫 번째 방안은 활성화 함수의 변경으로, 시그모이드나 tanh 대신 ReLU와 같은 함수를 사용하여 그래디언트가 사라지는 현상을 방지할 수 있습니다. 배치 정규화를 적용하면 각 층의 입력 분포를 정규화하여 그래디언트 흐름을 안정화시키고 학습 속도를 향상시킬 수 있습니다. 또한 잔차 연결을 통해 그래디언트가 네트워크의 이전 층으로 직접 흐를 수 있는 지름길을 제공함으로써 그래디언트 소실 문제를 완화할 수 있습니다.

LSTM이나 GRU와 같은 게이트 메커니즘을 가진 순환 신경망 구조를 사용하면 장기 의존성을 효과적으로 학습하고 그래디언트 소실 문제를 해결할 수 있습니다. 가중치 초기화 방법으로 Xavier나 He 초기화를 사용하면 각 층의 활성화 값이 적절한 분포를 유지하도록 하여 그래디언트 소실을 방지할 수 있습니다. 마지막으로 그래디언트 클리핑을 적용하여 그래디언트 폭발 문제를 방지하고 안정적인 학습을 가능하게 할 수 있습니다.

6-3. CNN에서 풀링(Pooling)이 이미지를 처리하는데 어떠한 이점을 제공하는지 설명하라. 또한, 가능하다면 Pooling 과정에 있어 최대(Max)가 아닌 평균(Average)를 사용했을 경우 어떠한 차이가 생기는지에 대해서도 설명하라.

**답:**

풀링은 CNN에서 특성 맵의 공간적 크기를 줄이는 연산으로, 주요 목적은 연산량을 줄이고 과적합을 방지하며 위치 변화에 대한 불변성을 확보하는 것입니다. 또한 풀링은 네트워크의 수용 영역을 확장시켜, 보다 넓은 영역의 특징을 추출할 수 있게 합니다.

최대 풀링(Max Pooling): 주어진 영역에서 가장 큰 값을 선택함으로써, 가장 강한 특징을 강조하고 잡음에 강한 특성을 보입니다.

평균 풀링(Average Pooling): 영역 내 모든 값의 평균을 취하여, 배경 정보나 전체적인 분포를 더 잘 보존합니다.

일반적으로 Max Pooling은 객체의 존재 여부나 경계 등 명확한 특징이 중요한 작업(예: 객체 검출)에 적합하며, Average Pooling은 전체적인 질감이나 패턴 분포가 중요한 작업(예: 장면 분류)에서 사용될 수 있습니다.

최근에는 풀링 대신 Strided Convolution을 사용하는 경우도 늘고 있으며, 이는 네트워크의 파라미터와 학습 가능성을 늘리는 방향으로 발전하고 있습니다.

6-4. RNN 모델의 각 타임 스텝에서 입력 데이터가 은닉층을 통해 출력층으로 어떻게 흐르는지 설명하라. 또한, RNN의 "재귀적(recurrent)" 특성은 무엇인지, 그리고 이러한 특성이 시계열 데이터를 처리할 때 어떤 이점을 제공하는지 설명하라. 이 때, RNN의 신경망 모델을 반드시 이용하라.

**답:**

RNN 모델에서는 각 타임 스텝  $t$ 에서 입력 데이터  $x_t$ 가 은닉층에 전달되어 이전 타임 스텝의 은닉 상태  $h_{t-1}$ 와 결합된 후 활성화 함수를 통과하여 새로운 은닉 상태  $h_t$ 를 생성합니다. 이 은닉 상태  $h_t$ 는 출력층으로 전달되어 현재 타임 스텝의 예측값  $y_t$ 를 생성하는 동시에, 다음 타임 스텝의 계산을 위해 저장됩니다. 이 과정은

$$h_t = \tanh(W_{hx}x_t + W_{hh}h_{t-1} + b_h) \text{와 } y_t = W_{yh}h_t + b_y$$

와 같은 수식으로 표현되며, 여기서  $W$ 와  $b$ 는 각각 가중치와 편향을 나타냅니다.

RNN의 재귀적 특성은 은닉 상태가 이전 타임 스텝의 정보를 현재 계산에 반영하는 순환 연결 구조를 의미합니다. 이러한 특성은 시계열 데이터에서 시간적 의존성을 모델링할 수 있게 하여, 텍스트나 음성과 같은 순차적 데이터에서 문맥 정보를 유지하는 데 중요한 역할을 합니다. 또한 이 구조는 가변 길이의 입력 시퀀스를 처리할 수 있게 하여 자연어 처리, 음성 인식, 시계열 예측 등 다양한 응용 분야에서 유연성을 제공합니다.