3_Optimization

문제 정의

- **최적화는 손실 함수(loss function) 또는 목적 함수(objective function)의 값을 최소화하는 매개변수를 찾는 과정**을 의미한다. 이 과정은 **제약 조건의 유무에 따라 제약이 있는 최적화와 제약이 없는 최적화**로 나눌 수 있다. **볼록 함수(convex function)**의 경우, 어떤 시작점에서 시작하더라도 **전역 최소값(global minimum)을 찾을 수 있다**. 반면, **오목 함수(concave function)**는 **시작점에 따라 지역 최소값(local minimum)에 도달할 위험**이 있다. 기본적인 최적화 방법 중 하나는 **함수의 기울기가 0이 되는 지점을 찾는 것**이다. 그러나 실제 머신러닝 및 딥러닝에서 다루는 많은 함수들은 **미분하기 어려운 경우**가 많다. 이러한 경우, **경사 하강법(gradient descent)이라는 방법을 사용하여 최솟값을 찾는다**. 이 방법은 **학습률 (learning rate)과 스텝 사이즈(step size)를 활용하여 각 반복(iteration)마다 파라미터를 업데이트**한다. 경사 하강법에는 몇 가지 변형이 있다. **확률적 경사 하강법(SGD)**은 **각 반복에서 단일 데이터 포인트 또는 소규모 배치를 사용하여 파라미터를 업데이트**한다. 반면, **배치 경사 하강법(BGD)**은 **전체 데이터셋을 사용하여 각 반복에서 파라미터를 업데이**트한다.

해당 문제에 대한 일반적인 접근

- 경사 하강법은 기계 학습에서 매개변수를 최적화하는 데 널리 사용되는 기법이다. 이 방법은 크게 세 가지 변형으로 나뉜다: 배치 경사 하강법(BGD), 확률적 경사 하강법(SGD), 그리고 미니 배치 경사 하강법(MSGD). **BGD**는 전체 데이터셋을 사용하여 각 반복에서 매개변수를 업데이트하므로 **계산 비용이 높지만 더 안정적이고 정확한 결과를 제공**할 수 있다. 반면, **SGD**
  는 각 반복에서
  단 하나의 데이터 포인트를 사용하여 빠른 수렴 속도를 제공하며 **메모리 효율성**이 뛰어나다. 그러나 이 방법은 **업데이트가 불안정할 수 있어 다양한 미니 배치를 사용하는 MSGD가 대안으로 제시**된다. MSGD는 **소규모 데이터 그룹을 활용하여 학습 속도와 메모리 사용량을 효율적으로 관리**하면서도, BGD와 비교할 때 **더 빠른 수렴을 보장**하고 SGD보다 **더 안정적인 학습 과정을 제공**한다.

일반적인 접근법의 제한 사항

- 경사 하강법의 각 변형은 특정 상황에서 유용하며 각각 고유의 단점도 가지고 있다. **SGD 방식은 노이즈가 많은 데이터에서 불안정한 수렴을 초래**할 수 있으며, 각 반복마다 업데이트 방향이 크게 변동할 수 있다. 이는 최적화 과정에서 전반적인 진행 방향이 일관되지 않다는 것을 의미하며, 결과적으로 **수렴 속도가 불규칙**해질 수 있다. **BGD**는 데이터셋 크기가 클 경우 많은 메모리를 요구하고, 연산 비용이 많이 든다. 결과적으로 전체 데이터셋을 처리하는 데 필요한 시간이 길어져 **학습 속도가 매우 느려질 수 있다. MSGD**은 BGD와 SGD의 중간 형태로, **무작위로 선택된 소규모 데이터 배치를 사용**한다. 이 방법은 더 빠른 수렴을 위해 메모리와 연산 비용을 절감하려는 장점이 있지만, **선택된 미니 배치가 대표성을 띠지 않을 경우 최적의 수렴을 보장하기 어려울 수 있다.** MSGD는 SGD에 비해 수렴 과정에서의 변동성을 줄이긴 하지만, 여전히 무작위성 때문에 완전히 안정적인 수렴을 보장하지는 않는다.

제한 사항에 대한 해결 방안

- 데이터의 양이 많아지고 모델의 구조가 복잡해짐에 따라, 효과적인 학습 방법을 찾는 연구가 활발히 진행되고 있다. 이러한 연구 중 하나로 경사 하강법의 보완 방법인 **'모멘텀 (Momentum)'**이 있다. 모멘텀은 **이전 단계에서의 기울기를 현재 기울기에 일정 비율로 반영하여 파라미터를 업데이트하는 방법**이다. 이 접근법은 단순히 현재의 기울기만을 고려하는 것이 아니라, 과거의 움직임을 통합하여 업데이트의 방향과 속도를 결정하므로, **안장점(saddle point)이나 지역 최소값(local minima)과 같은 잠재적인 최적화 함정을 해결**할 수 있다. 결과적으로 모멘텀은 **더 빠르고 안정적인 수렴을 가능**하게 한다. 특히, **모멘텀과 적응형 스텝 사이즈를 결합한 최적**

화 알고리즘인 'Adam'은 더욱 효율적인 학습을 지원한다. Adam은 각 파라미터에 대한 학습 속도를 개별적으로 조정하여, 경사 하강 과정을 더욱 최적화한다.