

Week6 Assignment

Learning Rate Scheduler

머신러닝과 딥러닝 모델을 학습시키는 과정에서, 학습률(Learning Rate, LR)은 모델의 성능과 학습 속도에 결정적인 영향을 미친다. 학습률이 너무 높으면 모델이 학습 과정에서 발산할 수 있으며, 너무 낮으면 학습이 느려지고 지역 최소값(local minima)에 갇힐 위험이 있다. 이러한 문제를 해결하기 위해 고안된 다양한 학습률 조정 기법 중에서, "Cosine Annealing Scheduler"이 있다.

- Cosine Annealing Scheduler의 기본 원리

Cosine Annealing Scheduler는 이름에서 알 수 있듯이, 학습률을 조정하는 데 코사인 함수의 주기적인 특성을 활용한다. 이 스케줄러의 핵심 아이디어는 학습률을 초기에는 높게 설정하여 모델이 빠르게 학습하고 다양한 영역을 탐색하도록 하다가, 점차 학습률을 줄여가며 모델이 세밀한 조정을 통해 최적의 솔루션에 접근하게 하는 것이다. 이 과정이 코사인 곡선의 형태를 따르며, 주기적으로 반복된다.

- Cosine Annealing Schedule의 수식

$$LR_t = LR_{min} + \frac{1}{2}(LR_{max} - LR_{min})(1 + \cos(\frac{T_{current}}{T_{max}}\pi))$$

이 식에서 LR_t 는 현재 시점 t 에서의 학습률, LR_{min} 과 LR_{max} 는 각각 설정된 최소 및 최대 학습률을 의미한다. $T_{current}$ 는 현재 epoch의 번호, T_{max} 는 사이클 당 최대 epoch 수를 나타낸다. 이 공식에 따라, 학습률은 주기적으로 LR_{max} 에서 LR_{min} 으로 감소한 다음 다시 LR_{max} 로 증가하는 패턴을 반복한다.

Cosine Annealing Scheduler의 장점으로는

- **빠른 탐색과 세밀한 조정의 균형:** 초기에 높은 학습률은 모델이 다양한 솔루션 공간을 빠르게 탐색하도록 하며, 점차 학습률을 감소시키는 과정에서 더 정확한 솔루션을 찾도록 한다.
- **전역 최소값 탐색 촉진:** 주기적으로 학습률을 재설정함으로써, 모델이 지역 최소값에서 벗어나 전역 최소값(global minimum)을 찾을 가능성이 증가한다.
- **유연한 적용성:** 이 스케줄러는 다양한 유형의 네트워크와 문제에 적용 가능하며, 특히 큰 데이터셋을 사용하는 복잡한 문제에서 뛰어난 성능을 보인다.

Training Error 와 Generalization Error 사이의 간극 줄이기

1. 데이터 증강(Data Augmentation)

데이터 증강은 특히 이미지 처리, 음성 인식 등의 분야에서 널리 사용되는 기법이다. 이 방법은 기존의 학습 데이터를 변형하거나 조정하여 데이터의 양을 인위적으로 늘리는 방식이다. 예를 들어, 이미지를 회전, 반전, 크기 조정 등을 통해 다양한 변형을 적용하면, 모델이 더 다양한 시나리오에서 일반화할 수 있는 능력을 개발할 수 있다. 데이터 증강은 모델이 과적합되는 것을 방지하고, 일반화 성능을 향상시키는 효과적인 방법이다.

2. 정규화 기법(Regularization Techniques)

정규화는 모델의 복잡도를 제한하여 과적합을 방지하는 기법이다. 대표적인 정규화 기법에는 L1 정규화, L2 정규화(릿지 회귀), 그리고 드롭아웃(Dropout)이 있다. L1 과 L2 정규화는 모델의 가중치에 패널티를 부여하여 가중치의 크기를 줄이는 방식으로 작동하며, 드롭아웃은 학습 과정에서 무작위로 일부 뉴런을 비활성화시켜 모델이 데이터의 작은 변화에 덜 민감하도록 만든다. 이러한 기법들은 모델이 학습 데이터의 노이즈에 과도하게 적응하는 것을 방지하고, 일반화 성능을 높이는 데 도움을 준다.

3. 교차 검증(Cross-Validation)

교차 검증은 모델의 일반화 능력을 평가하고, 과적합을 감지하는 데 유용한 방법이다. 가장 널리 사용되는 방식은 k-겹 교차 검증(k-fold Cross-Validation)으로, 데이터 세트를 k 개의 동일한 크기를 가진 부분 집합으로 나누고, 이 중 하나를 검증 세트로 사용하고 나머지를 학습에 사용하는 과정을 k 번 반복한다. 이 방법은 모델이 다양한 데이터 세트에 대해 일관된 성능을 보이는지 확인하고, 과적합을 방지하는 데 도움이 된다.

4. 앙상블 학습(Ensemble Learning)

앙상블 학습은 여러 개별 모델의 예측을 결합하여 최종 예측을 만드는 기법이다. 이 방법은 각각의 모델이 서로 다른 종류의 오류를 만들 가능성이 있기 때문에, 모델의 일반화 성능을 향상시킬 수 있다. 대표적인 앙상블 기법에는 랜덤 포레스트(Random Forest), 부스팅(Boosting), 배깅(Bagging) 등이 있습니다. 앙상블 기법은 모델의 안정성과 성능을 향상시키며, 과적합 위험을 줄이는 효과적인 방법이다.