

Week6-Part2. Neural Network Advanced

1. 다양한 Learning Rate Scheduler 중 적어도 한 가지 소개하기

Cyclic Learning Rate (CLR)

: Leslie N. Smith가 제안한 방법으로, 전통적인 학습률 스케줄링 방법과 달리 학습률을 주기적으로 조정함으로써 학습 과정을 개선하는 기법이다. 핵심 아이디어는 학습률을 사이클마다 특정 범위 내에서 변화시켜 학습 과정을 동적으로 조절하는 것이다. 이를 통해 local minimum에서 쉽게 벗어나 global minimum을 찾을 수 있도록 돕는다.

CLR의 작동 원리

CLR은 두 가지 주요 파라미터에 의존한다. 학습률의 하한값(base_lr)과 상한값(max_lr). 이 두 값 사이에서 학습률은 사이클에 따라 조정된다. 사이클의 각 단계에서 학습률은 하한값에서 시작하여 상한값까지 증가한 다음 다시 하한값으로 감소합니다. 이러한 변화는 주로 선형적(linear)이거나 sin 곡선(sinuous) 패턴을 따른다.

사이클의 유형

- **Triangular**: 가장 간단한 형태의 CLR로, 학습률이 선형적으로 상한값과 하한값 사이를 오간다.
- **Triangular2**: Triangular의 변형으로, 각 사이클이 진행됨에 따라 학습률의 최대값이 절반으로 줄어든다.
- **Exp Range**: 각 사이클의 최대 학습률이 지수적으로 감소합니다. 이 방법은 학습률을 더 미세하게 조정할 수 있게 해준다.

CLR의 장점

1. **빠른 수렴**: 동적인 학습률 조정을 통해 모델이 더 빠르게 수렴할 수 있도록 합니다.
2. **지역 최소값 탈출**: 주기적인 학습률 변화가 모델을 지역 최소값에서 벗어나게 만들어 글로벌 최소값을 찾는 데 도움을 줍니다.
3. **하이퍼파라미터 조정의 단순화**: 전통적인 학습률 감소 방식에 비해 덜 민감하며, 실험을 통해 최적의 하한값과 상한값을 비교적 쉽게 찾을 수 있습니다.

2. Training Error와 Generalization Error 사이 간극을 줄이는 방안 소개하기

- 훈련 기법 개선

Early Stopping

조기 종료는 과적합을 방지하는 간단하면서도 효과적인 방법입니다. 이 기법은 훈련 과정에서 모

델의 성능이 검증 데이터셋에서 더 이상 개선되지 않는 지점을 감지하고, 그 시점에서 훈련을 중단합니다. 이를 통해 모델이 훈련 데이터에 과적합되는 것을 방지할 수 있습니다. 일반적으로 성능이 일정 기간 동안 개선되지 않으면 훈련을 중단한다.

Cross-Validation

교차 검증은 모델의 일반화 능력을 더 정확하게 평가하기 위해 사용되는 기법입니다. 데이터를 여러 개의 부분집합으로 나누고, 이 중 하나를 검증 데이터로, 나머지를 훈련 데이터로 사용합니다. 이 과정을 모든 부분집합에 대해 반복하고, 각 반복에서 얻은 성능을 평균내어 모델의 성능을 평가합니다. 가장 흔한 형태는 k-겹 교차 검증(k-fold Cross-Validation)으로, 데이터를 k개의 부분집합으로 나누고 이 과정을 k번 반복합니다.

- 앙상블 기법

배깅(Bagging)

배깅(Bagging, Bootstrap Aggregating의 약자)은 여러 개의 독립적인 결정 트리를 훈련시키고, 그 예측을 평균내거나 다수결을 통해 최종 결정을 내리는 방법이다. 배깅은 병렬로 모델을 학습시키며, 각 모델은 원본 훈련 데이터셋에서 부트스트랩(랜덤 샘플링)을 통해 생성된 데이터셋을 사용한다. 대표적인 예로 랜덤 포레스트(Random Forest)가 있다.

부스팅(Boosting)

부스팅은 약한 학습기(weak learner) 여러 개를 순차적으로 훈련시키면서, 잘못 예측된 데이터에 대한 가중치를 증가시키는 방식으로 모델의 성능을 향상시킨다. 각 단계에서의 모델은 이전 모델들이 어려워했던 데이터 포인트들에 더 잘 맞추도록 학습된다. 이 방법으로 AdaBoost, Gradient Boosting, XGBoost 등이 있다.

스태킹(Stacking)

스태킹은 다양한 모델로부터의 예측을 입력으로 사용하여, 새로운 모델(메타 학습기)을 훈련시키는 방법이다. 기본 아이디어는 여러 다른 모델들의 예측을 결합하여 최종 예측을 생성하는 것으로, 이 과정에서 각 기본 모델의 예측 성능을 최대화하는 방향으로 메타 학습기가 학습된다. 스태킹은 모델의 다양성을 활용하여 성능을 향상시키는데, 이때 기본 모델들이 서로 다른 종류여야 한다.