

## 1. 다양한 Learning Rate Scheduler 중 적어도 한 가지를 소개

Learning Rate는 gradient의 보폭으로, 신경망 모델이 backpropagation 과정에서 가중치를 업데이트 할 때의 업데이트 규모를 의미한다. 이때 손실 함수의 Error Estimates를 감소시키도록 모델의 가중치가 업데이트된다.

Learning Rate Schedules는 훈련 과정에서 에포크 혹은 이터레이션 사이에 러닝레이트 값을 조정하는 작업을 의미하는데, 주로 Constant Learning rate와 Learning Rate decay 방법론이 있다. 전자는 러닝레이트를 설정하고 훈련과정에서 이를 수정하지 않는 것이고, 후자는 처음 Peak Learning rate를 설정하고 점차적으로 줄여나가는 방식이다. pytorch에서는 다양한 learning rate scheduler를 제공하고 있다.

### - LambdaLR

초기 learning rate에 lambda 함수에서 나온 값을 곱해줘서 learning rate를 계산

$$lr_{epoch} = lr_{initial} * \text{Lambda}(\text{epoch})$$

```
optimizer = torch.optim.SGD(model.parameters(), lr=0.001)
scheduler = optim.lr_scheduler.LambdaLR(optimizer=optimizer,
                                         lr_lambda=lambda epoch: 0.95 ** epoch)
```

### - CosineAnnealingLR

learning rate가 cos함수를 따라 eat\_min까지 떨어졌다가 다시 초기 learning rate까지 올라옴

$$\eta_t = \eta_{\min} + 1/2(\eta_{\max} - \eta_{\min})(1 + \cos((T_{\text{cur}}/T_{\max}) * \pi))$$

```
scheduler = optim.lr_scheduler.CosineAnnealingLR(optimizer, T_max=50, eta_min=0)
```

<참고>

<https://www.kaggle.com/code/isbhargav/guide-to-pytorch-learning-rate-scheduling/notebook>

## 2. 수업에서 소개되지 않은, Training Error와 Generalization Error 사이 간극을 줄이는 방안

Training Error(훈련 오차)는 모델이 학습 데이터를 얼마나 잘 예측하는지를 나타내는 오차이다. Training error가 낮을수록 학습 데이터에 더 잘 적합되어 있다는 것을 의미하나, 과적합 문제로 인해 Training error가 낮다고 항상 좋은 모델이라고는 할 수 없다.

Generalization Error(일반화 오차)는 모델이 이전에 본 적이 없는 새로운 데이터에 대해 얼마나 잘 예측하는지를 나타내는 오차이다.

### - 조기 중단(early stopping)

검증 데이터셋(validation dataset)의 loss가증가하기 시작하는 시점에서 훈련을 조기에 중단시키는 기법을 의미한다. 불필요한 학습을 방지할 수 있어 효율적인 훈련이 가능하고, 과적합을 막을 수 있어서 일반화 성능을 향상시킬 수 있다.

### - 모델 복잡도 축소

모델의 층 수나 은닉 노드 수를 줄여 모델의 복잡성을 감소시켜 일반화 성능을 향상시킬 수 있다.