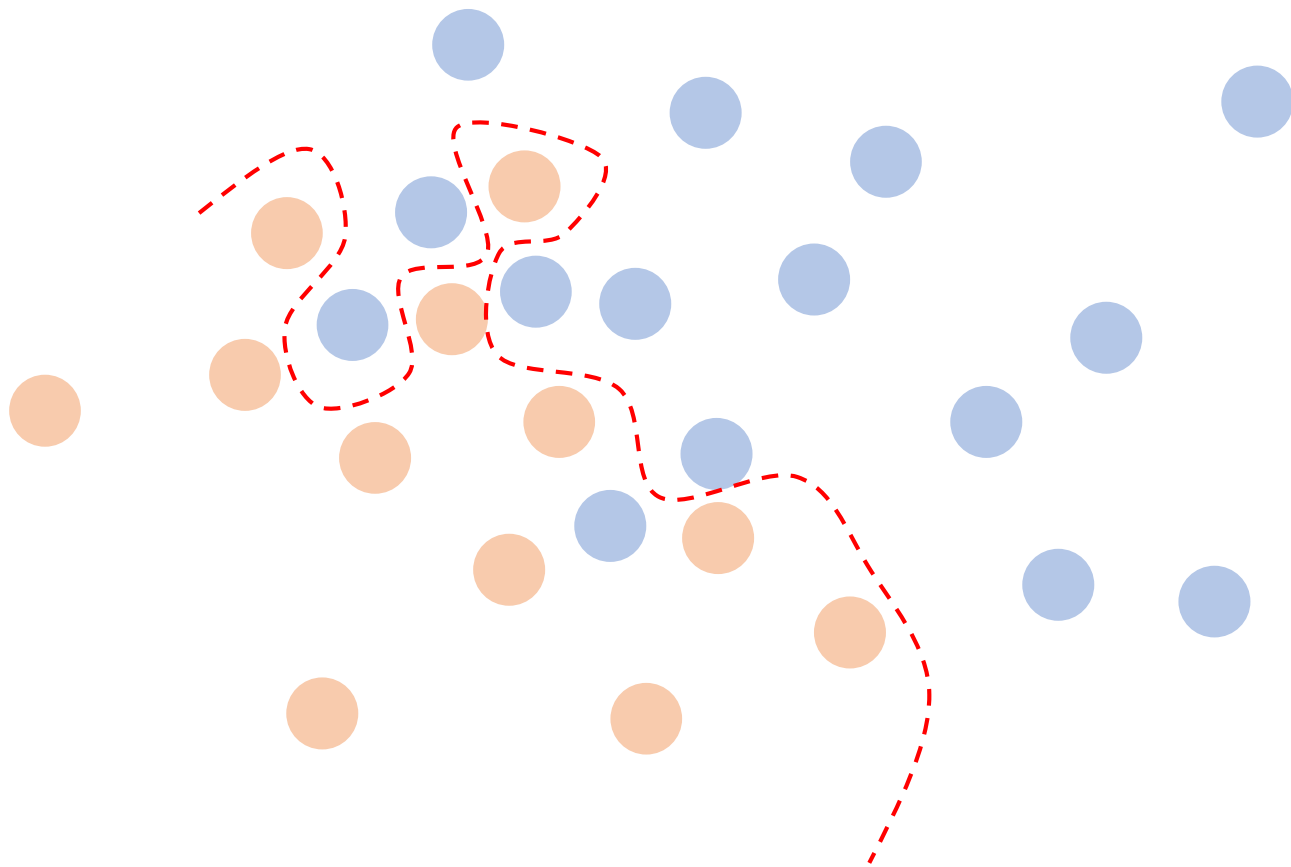


Chapter 08. 효과적이면서도 쉽게 쓸 수 있는 기법들

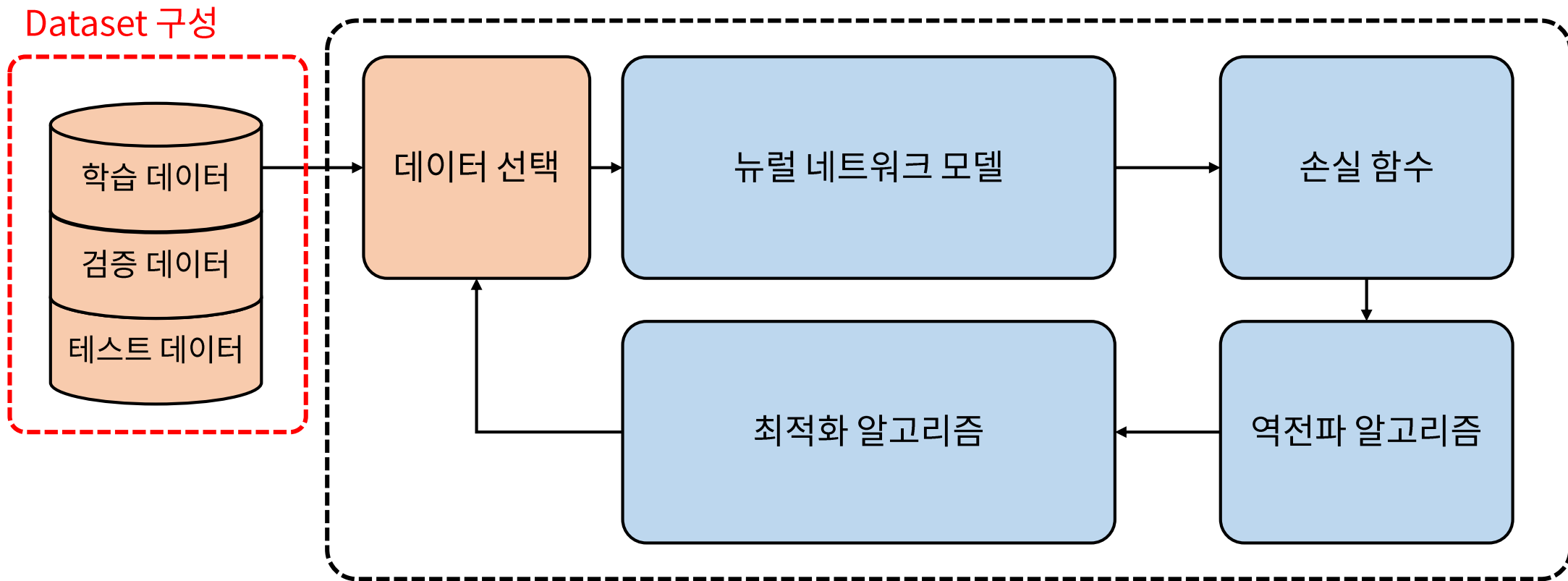
STEP1. 과적합의 해결

과대 적합 Overfitting



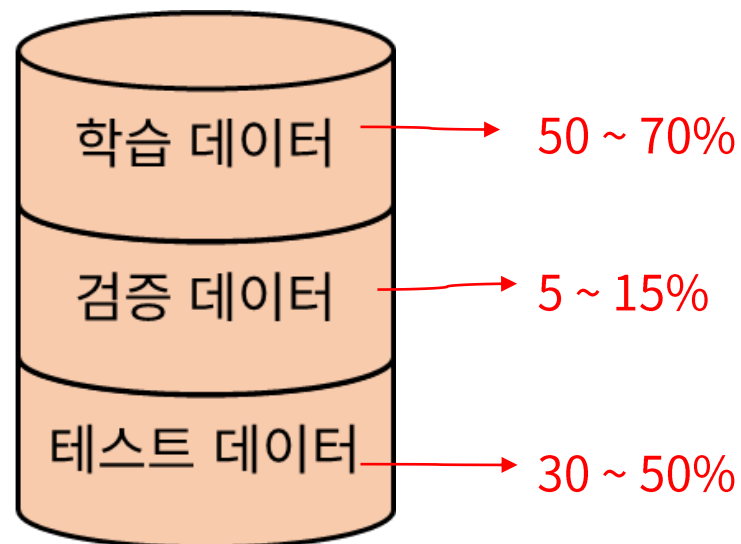
이진 분류 문제에서의 Overfitting. 마치 시험 족보를 외우듯 **주어진 문제의 답을 암기**해 버리는 것과 같다.

데이터셋의 구성



과적합을 더 설명하기에 앞서, 우선 데이터셋의 구성과 활용에 대해 이해할 필요가 있다.

데이터셋의 의미



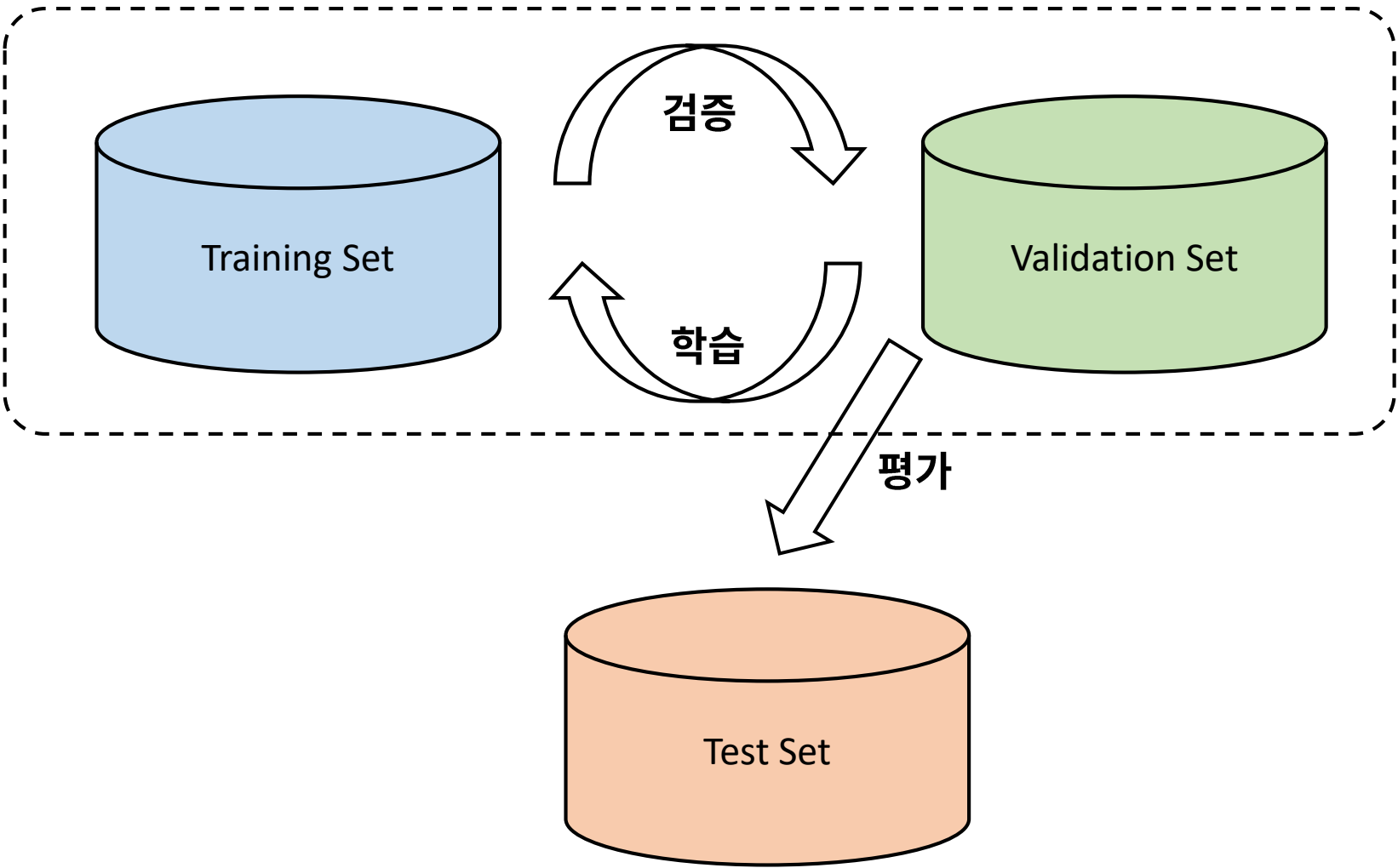
학습 데이터(Training Data) : 학습 과정에 보여지고, **실제 모델을 학습하는 데에 사용**되는 데이터

검증 데이터(Validation Data) : 학습 과정에 보여지는 데이터이지만, 모델 학습에 사용하지 않고
학습이 잘 되는지 검증하는 데에만 사용

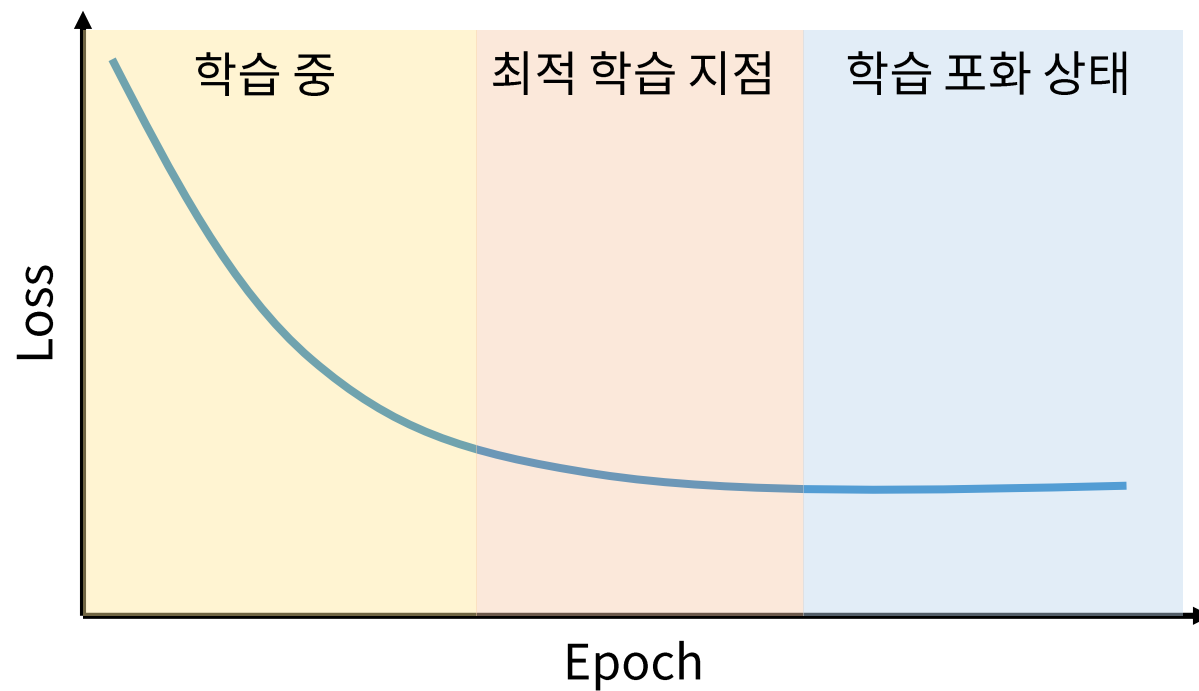
테스트 데이터(Test Data) : 학습 과정에서는 사용하지 않고, **학습을 마친 모델을 평가**하기 위해 단 한번만 사용

데이터셋의 활용

학습 Loop

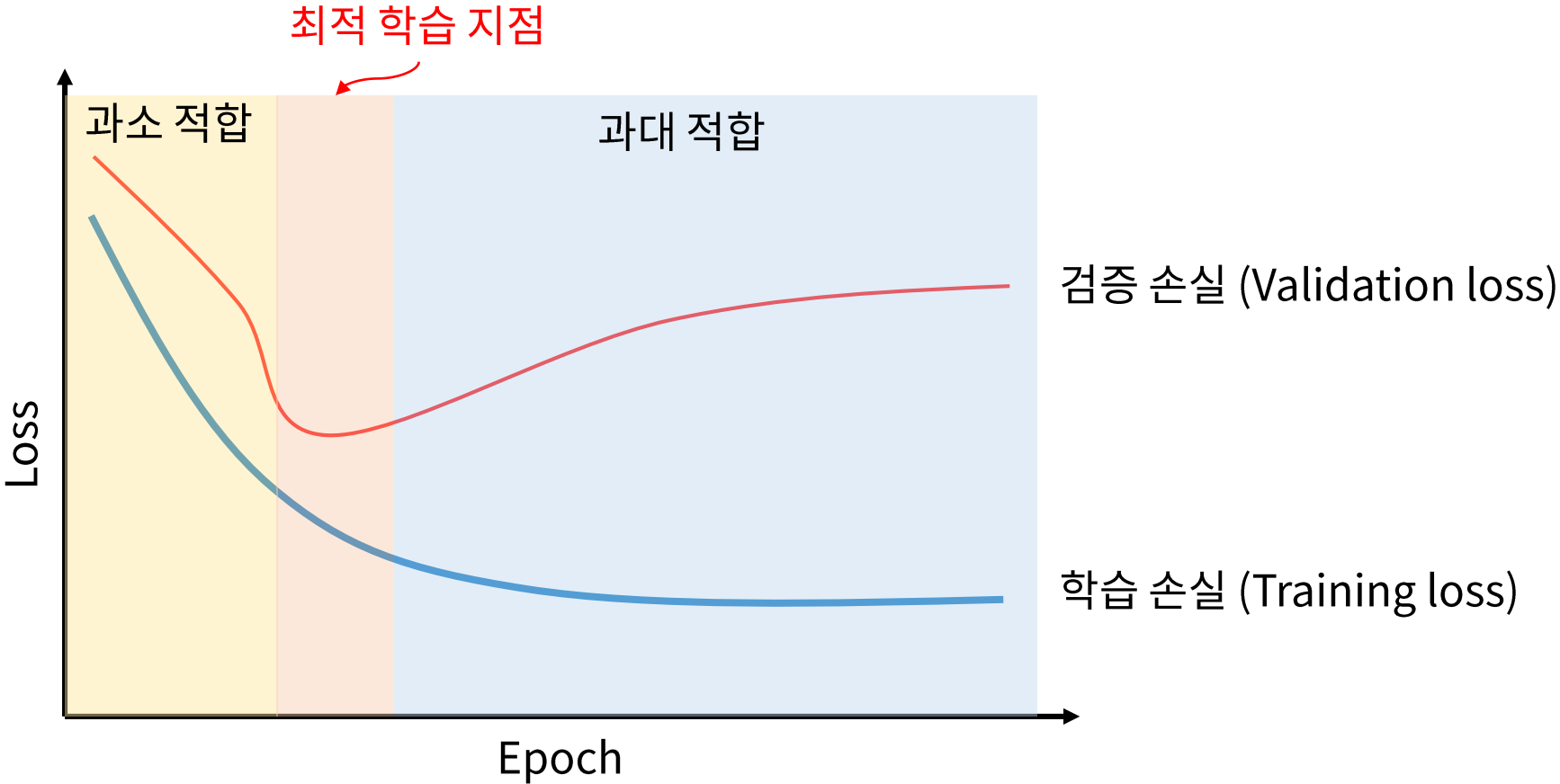


손실 함수 그래프



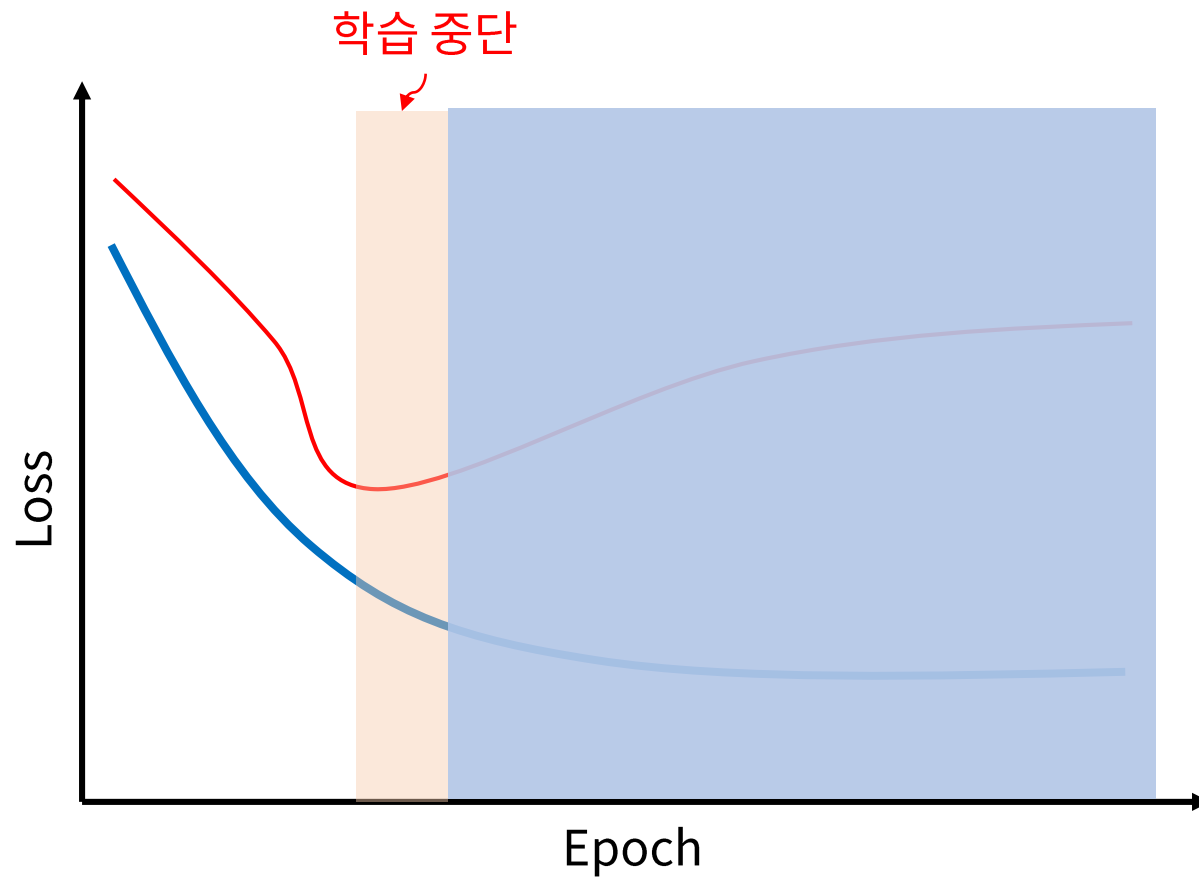
학습 데이터만을 이용해서 학습을 한다면, 위와 같이 판단하는 것이 합리적이다.

손실 함수 그래프 (검증)



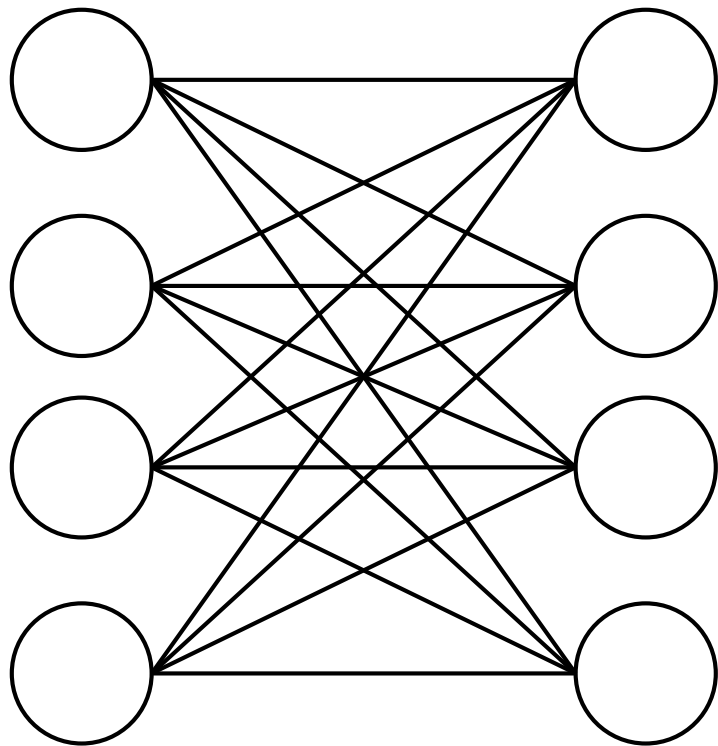
실제로 학습을 진행하다 보면, 학습 데이터에 과적합(Overfitting)되는 현상이 발생한다.

Early Stopping

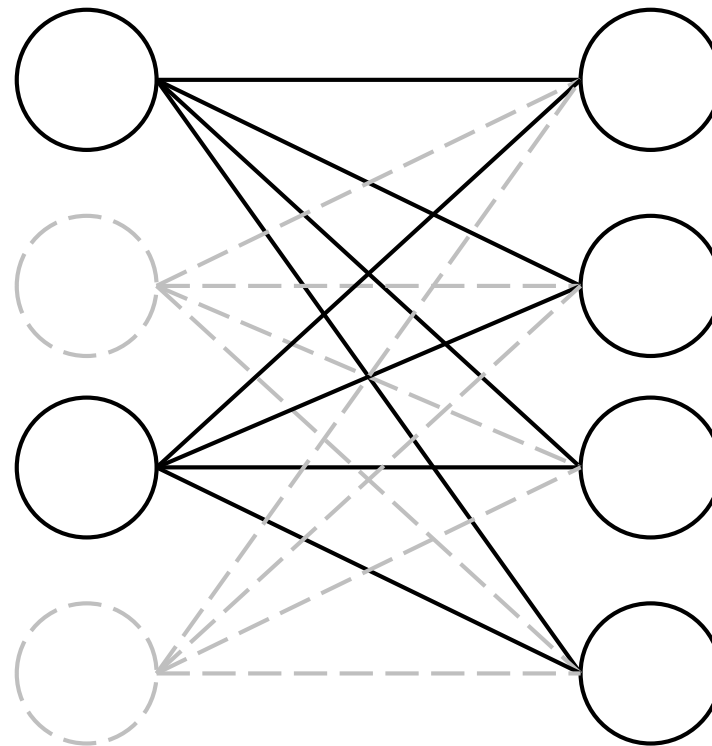


Validation loss가 여러 Epoch 동안 감소하지 않으면 **Overfitting**으로 간주하여 학습을 중단한다.

Drop out



원본 전결합 계층

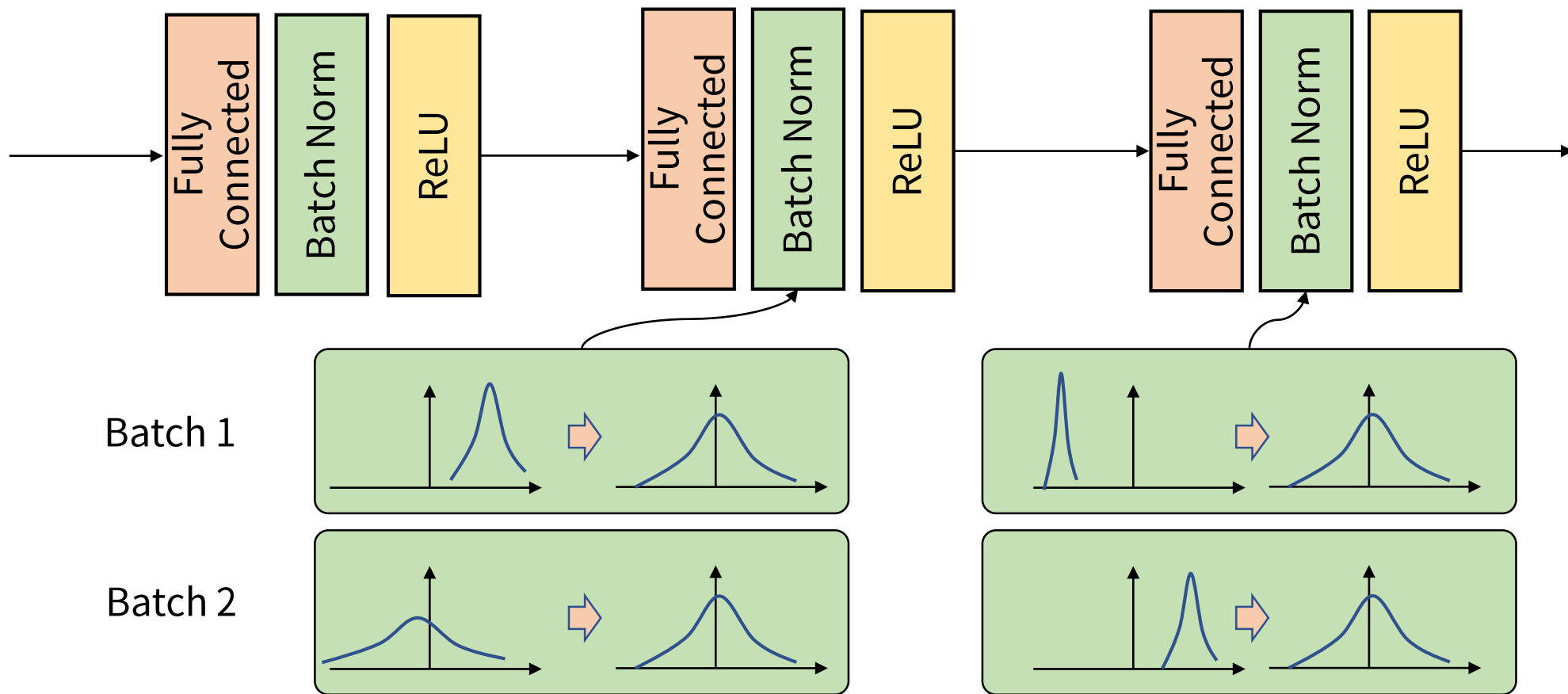


Drop out rate 0.5를 적용한 계층

지정한 비율의 뉴런을 제거(drop out)하고 학습하는 방법.

테스트 시에는 모든 뉴런을 사용하기 때문에, 여러 Network를 Ensemble한 효과를 가진다.

배치 정규화 Batch Normalization



배치 정규화는 중간 **Feature**들을 그대로 사용하지 않고 변형하여 학습하기 때문에 Overfitting을 개선하는데 도움이 된다.