

Part. 06

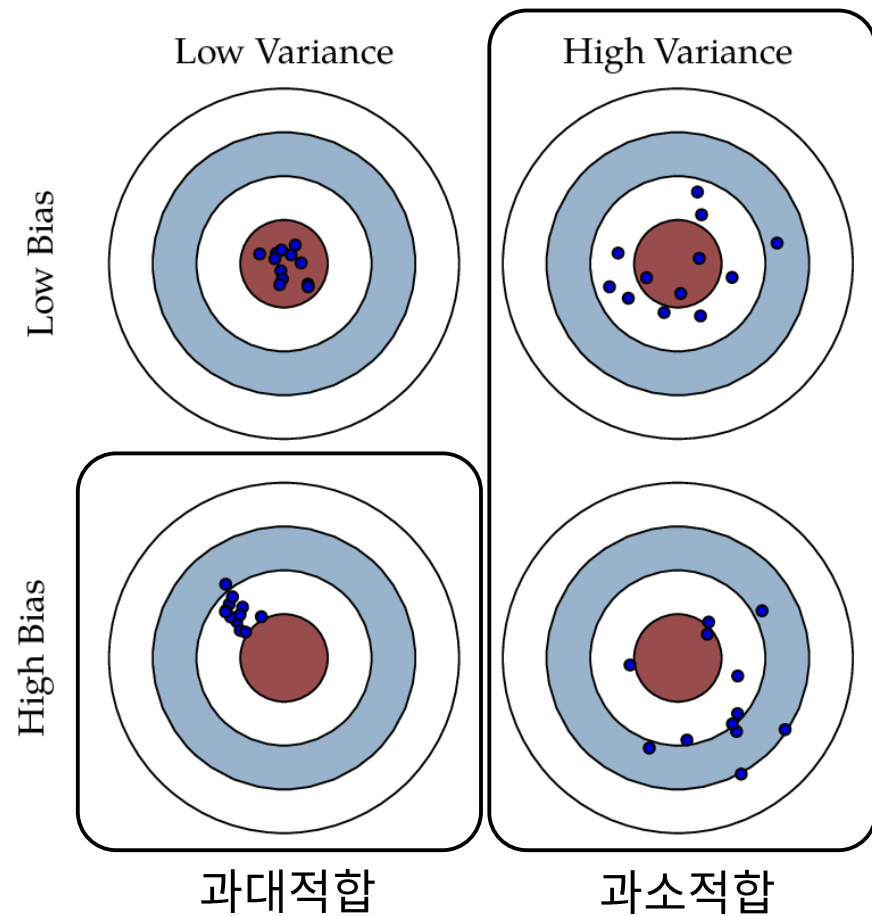
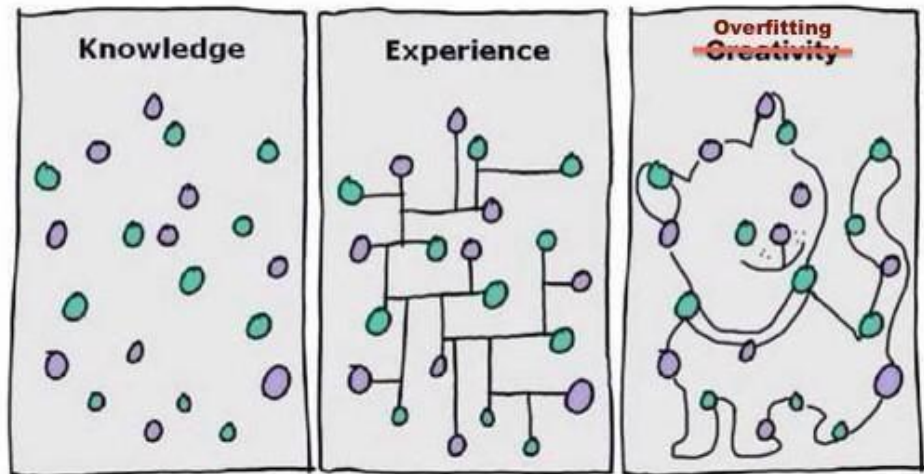
Generalization

| 과소적합과 과대적합

FASTCAMPUS
ONLINE

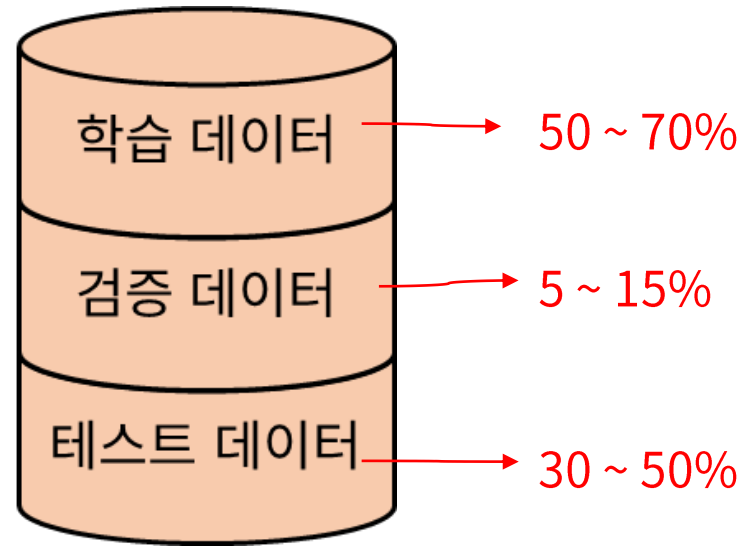
강사. 신제용

I 과대적합? 과소적합?



과대적합과 과소적합에 관한 가장 직관적인 설명

I 데이터셋의 구성

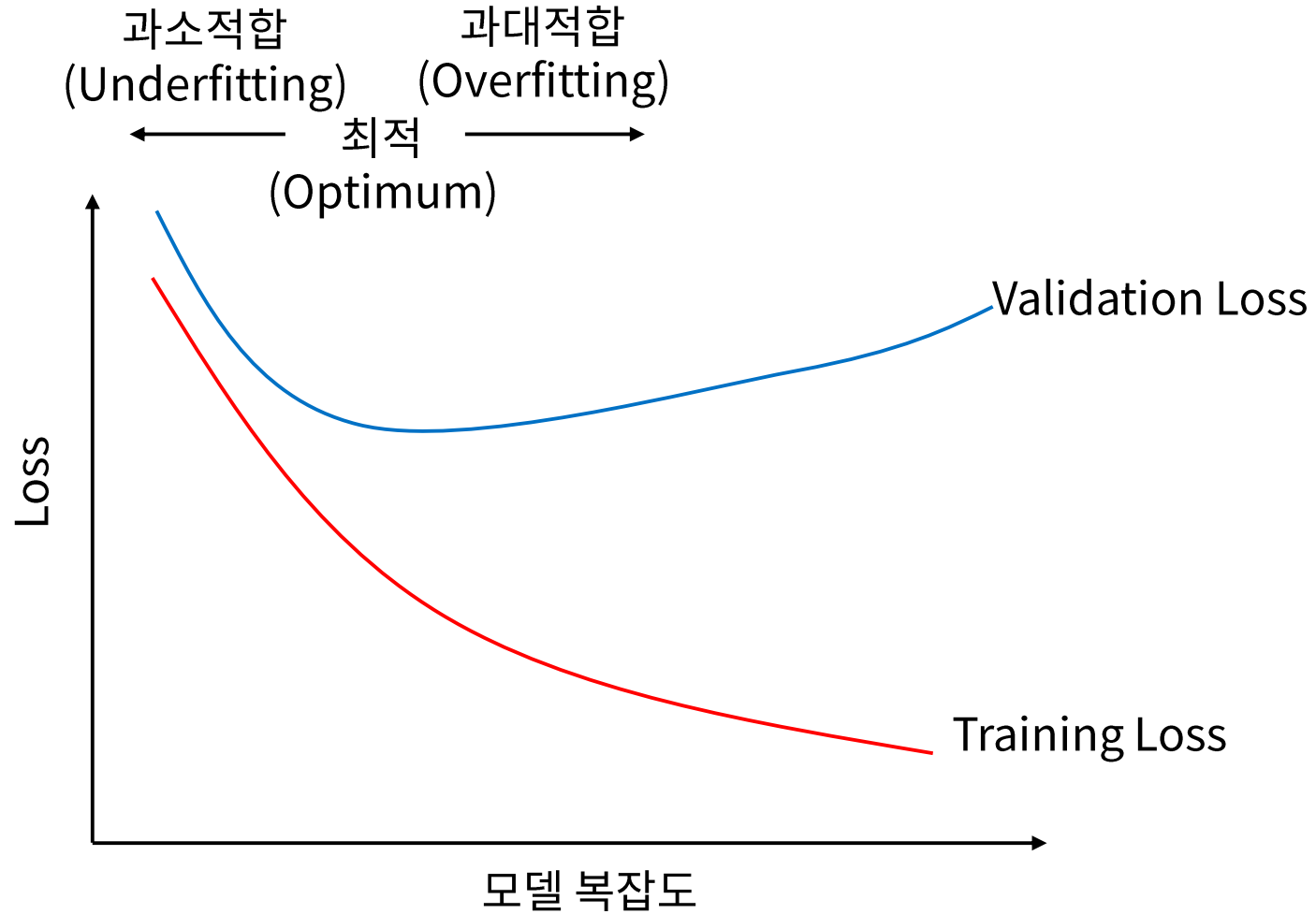


학습 데이터(Training Data) : 학습 과정에 보여지고, **실제 모델을 학습하는 데에 사용**되는 데이터

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

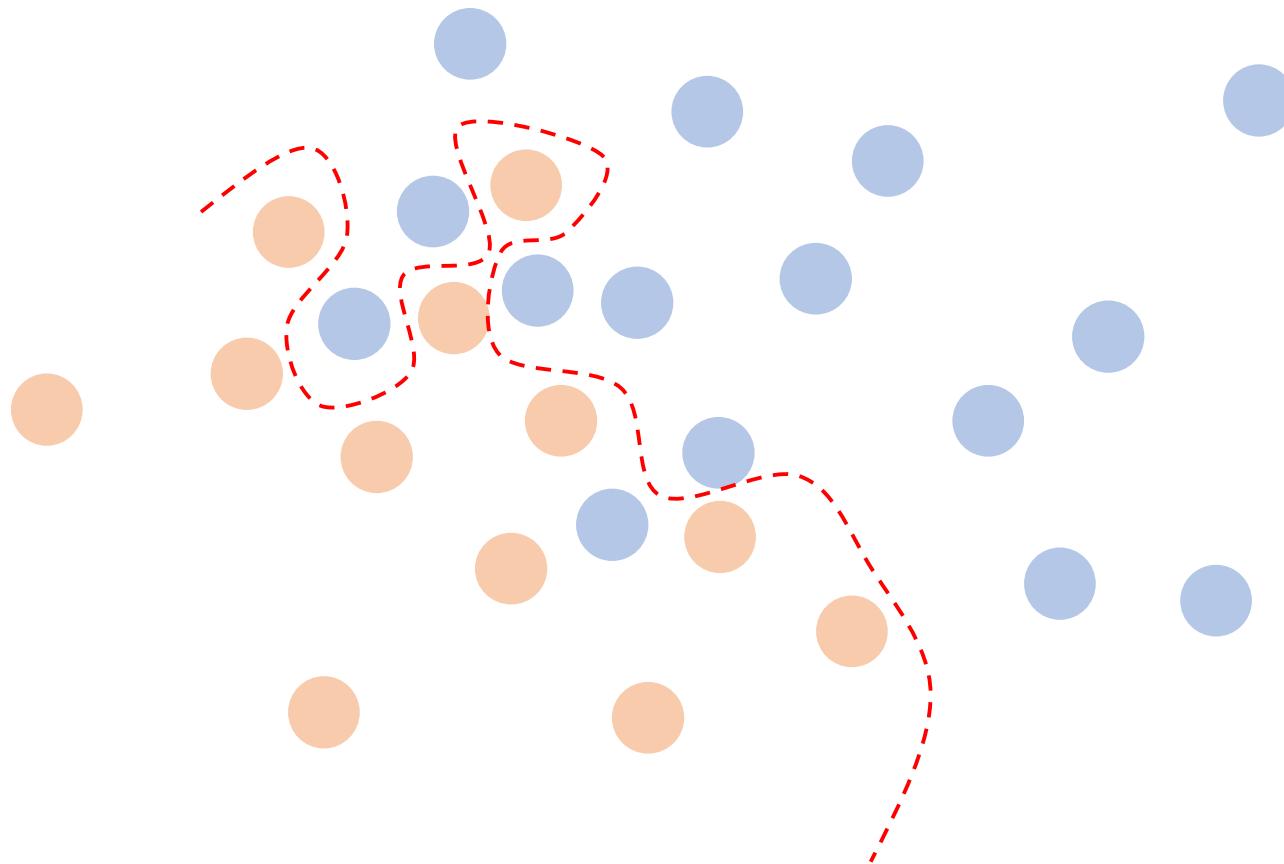
I 과소적합, 과대적합



문제의 난이도에 비해 **모델 복잡도가 너무 낮으면 과소 적합**이 발생한다.

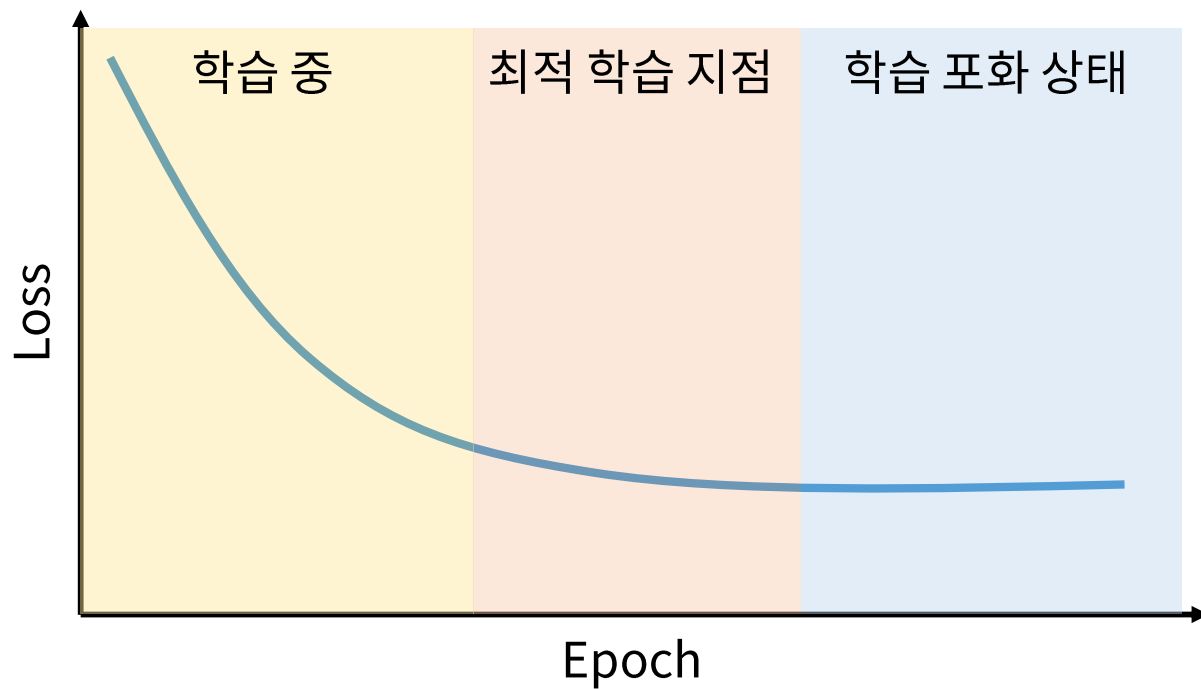
반대로 문제를 **너무 어렵게 풀다 보면 과대 적합**이 발생한다.

I 과대 적합 (Overfitting)



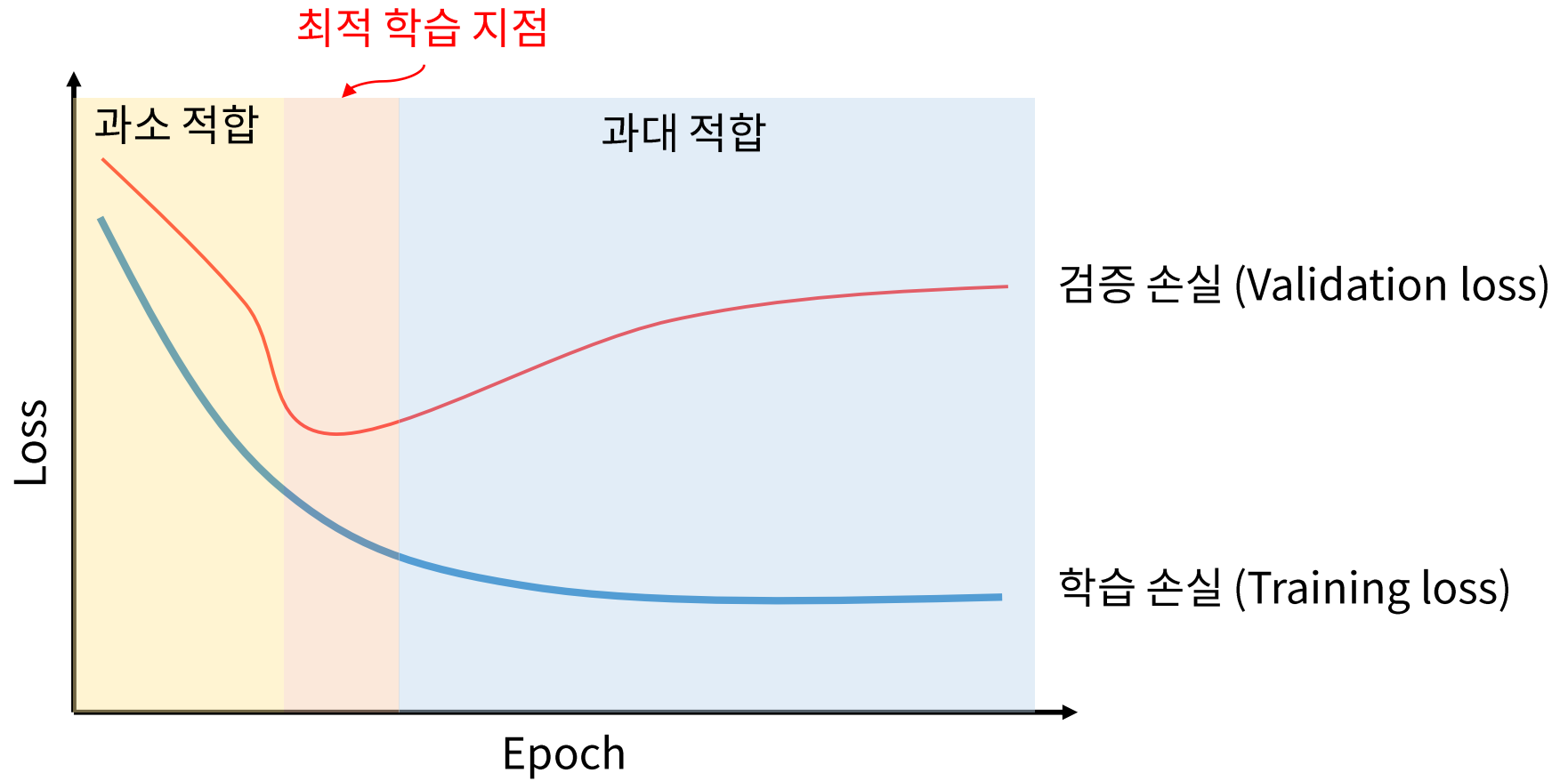
이진 분류 문제에서의 Overfitting. 마치 시험 족보를 외우듯 **주어진 문제의 답을 암기**해 버리는 것과 같다.

I 손실 함수 그래프



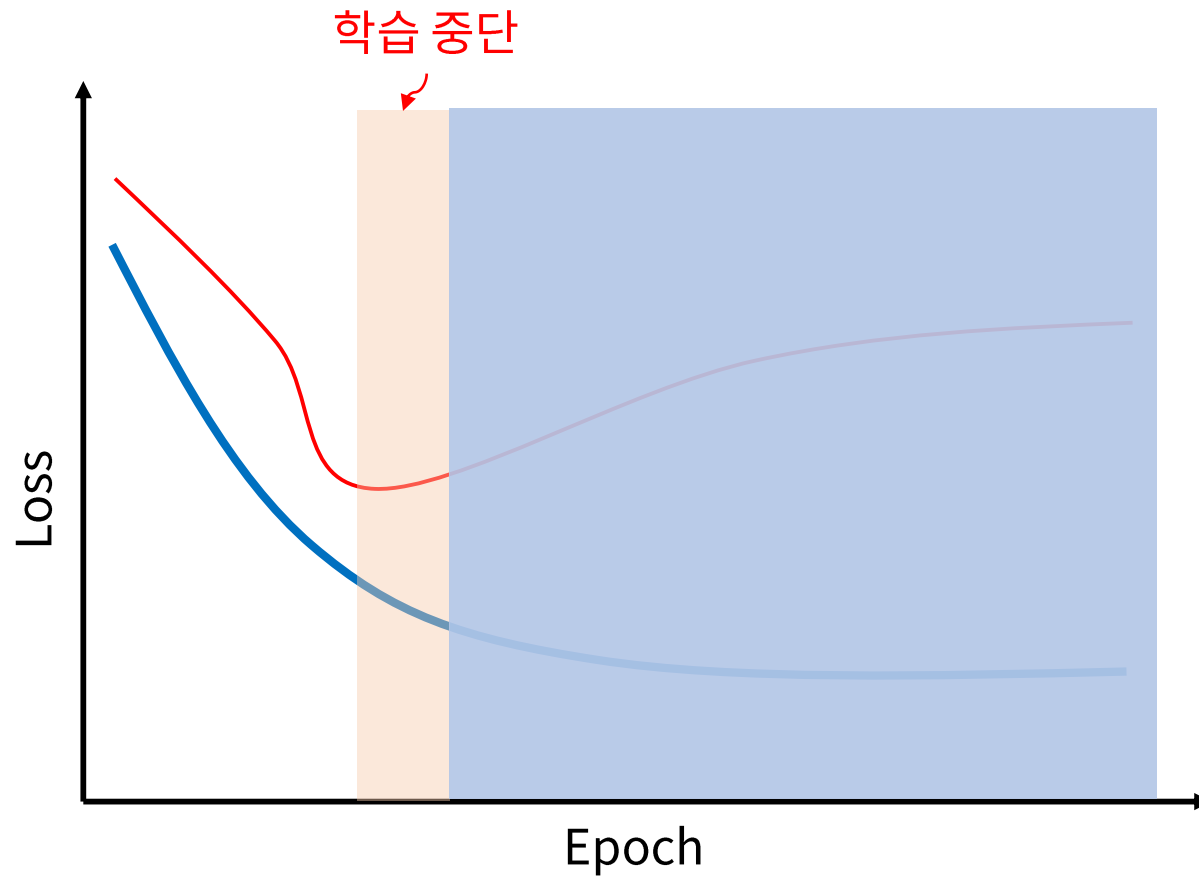
학습 데이터만을 이용해서 학습을 한다면, 위와 같이 판단하는 것이 합리적이다.

I 손실 함수 그래프 (학습 + 검증)



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

I 학습 조기 종료 (Early Stopping)



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