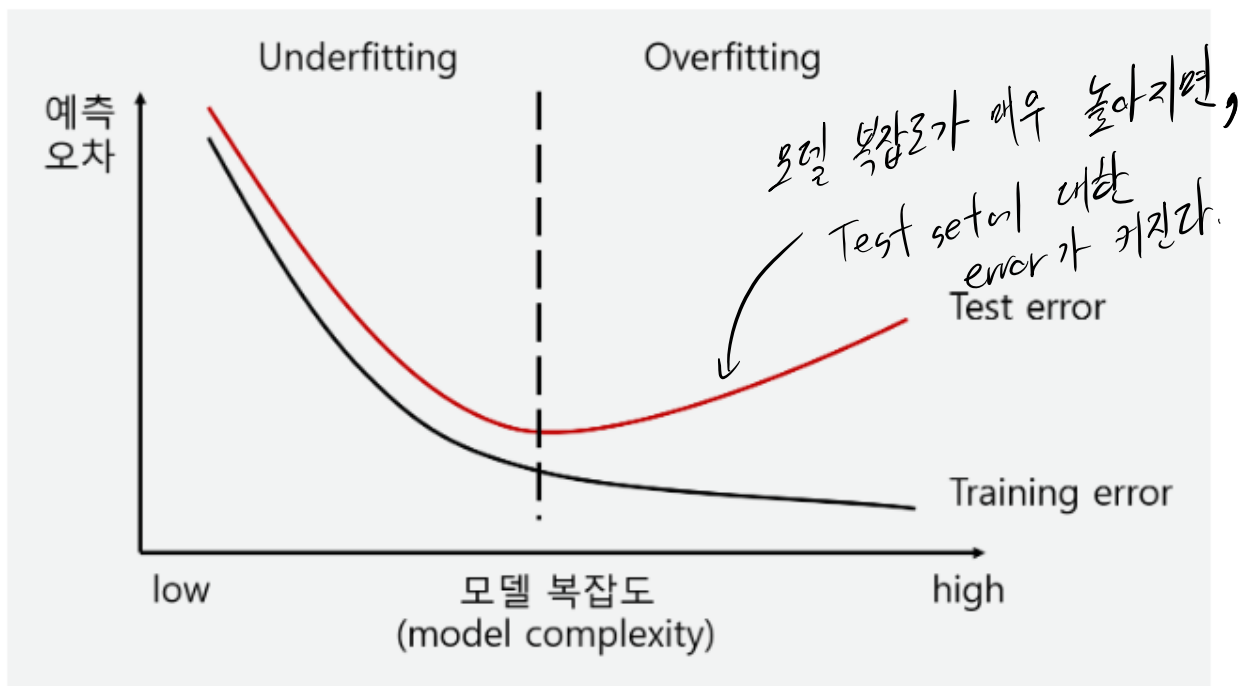


오버 피팅 발생의 근본적인 원인은 학습 데이터의 부족이다.

따라서 근본적인 해결책은 데이터를 충분히, 더 많이 모으는 것이다.

하지만 우리는 학습하려는 실제 모델을 모르기때문에 어느정도의 데이터가 **충분**한 것인지 알 수 없다. 따라서 수집한 데이터 안에서 최대한 오버피팅을 피하면서 일반화 성능이 좋은 모델을 개발해야한다. 이러한 상황으로부터 **모델 복잡도**와 오버피팅의 상관성을 살펴볼 수 있다. **모델 복잡도**란 간단하게 말해서 어떤 모델을 정의하기 위해 필요한 파라미터의 수로 생각할 수 있다. 즉 파라미터의 수가 많을수록 정교하고 복잡한 모델을 표현할 수 있고, 적으면 간단한 모델만 설명할 수 있는 것이다. 모델 복잡도에 따라 발생하는 언더 피팅 오버 피팅의 그래프를 그려보면 다음과 같다.



모델 복잡도와 오버 피팅 그래프

