



일반화,과대적합,과소적합

일반화 (Generalization)

- 훈련 세트로 학습한 모델이 테스트 세트에 대해 정확히 예측하도록 하는 것 .

과대적합 (Overfitting)

- 훈련 세트에 너무 맞추어져 있어 테스트 세트의 성능 저하.

과소적합 (Underfitting)

- 훈련 세트를 충분히 반영하지 못해 훈련 세트, 테스트 세트에서 모두 성능이 저하.

과대적합 (Overfitting)

요트 회사의 고객

45세 이상, 자녀 셋 미만,
이혼하지 않은 고객

나이	보유차량수	주택보유	자녀수	혼인상태	애완견	보트구매
66	1	yes	2	사별	no	yes
52	2	yes	3	기혼	no	yes
22	0	no	0	기혼	yes	no
25	1	no	1	미혼	no	no
44	0	no	2	이혼	yes	no
39	1	yes	2	기혼	yes	no
26	1	no	2	미혼	no	no
40	3	yes	1	기혼	yes	no
53	2	yes	2	이혼	no	yes
64	2	yes	3	이혼	no	no
58	2	yes	2	기혼	yes	yes
33	1	no	1	미혼	no	no

과소적합 (Underfitting)

요트 회사의 고객

집이 있는 고객

나이	보유차량수	주택보유	자녀수	혼인상태	애완견	보트구매
66	1	yes	2	사별	no	yes
52	2	yes	3	기혼	no	yes
22	0	no	0	기혼	yes	no
25	1	no	1	미혼	no	no
44	0	no	2	이혼	yes	no
39	1	yes	2	기혼	yes	no
26	1	no	2	미혼	no	no
40	3	yes	1	기혼	yes	no
53	2	yes	2	이혼	no	yes
64	2	yes	3	이혼	no	no
58	2	yes	2	기혼	yes	yes
33	1	no	1	미혼	no	no



아이에게 공이 무엇인지 알려주자

공이라는 것은..




과대적합

두근두근 새겨다

개 붙어있다.
으로 구성된다.
만씩만씩 광이 난다.

공이라는 것은..

두근두근 새겨다



과소적합

일반화 성능이 최대화 되는 모델을 찾는 것이 목표



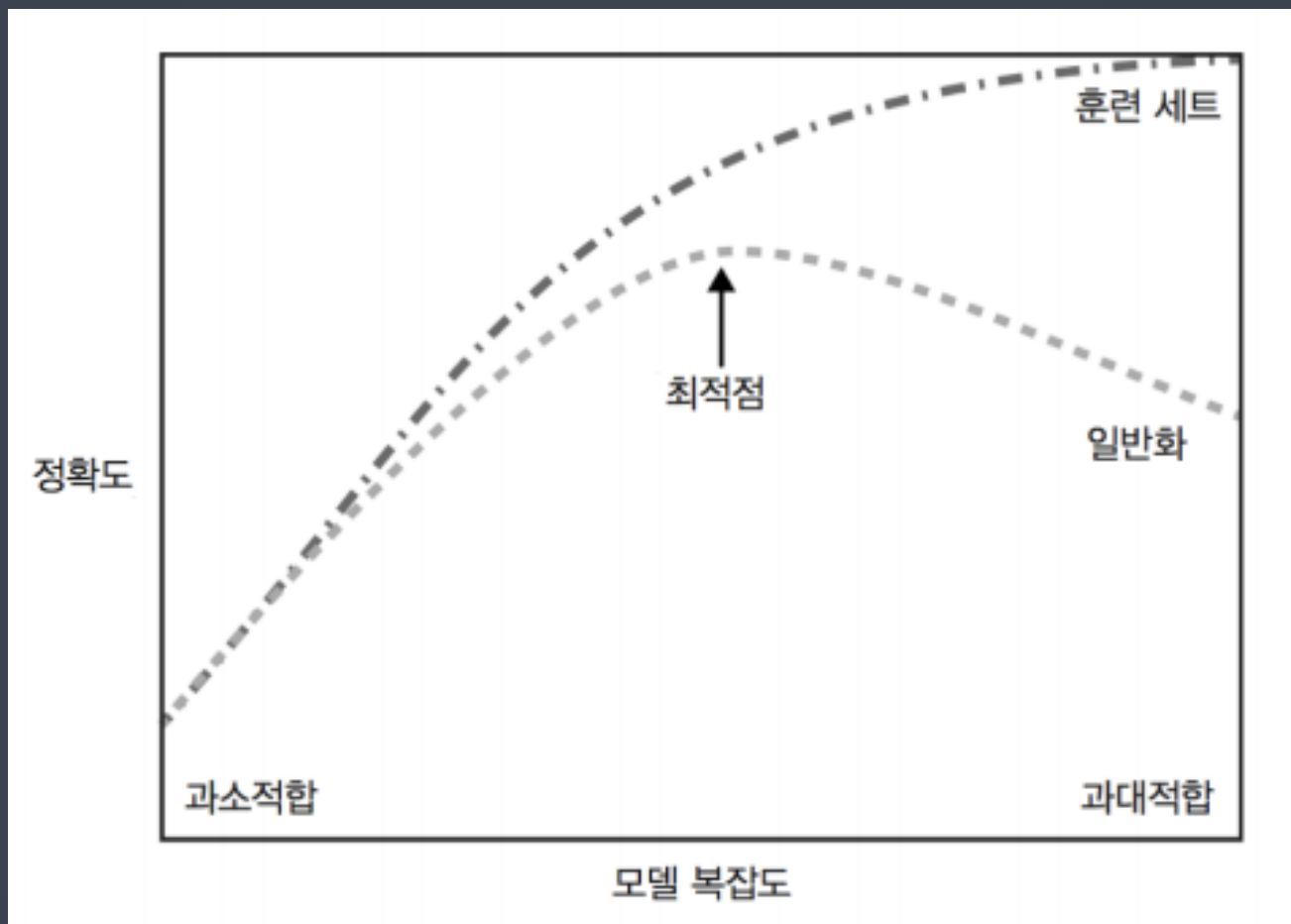
과대적합 (Overfitting)

- 너무 상세하고 복잡한 모델링을 하여 훈련데이터에만 과도하게 정확히 동작하는 모델.

과소적합 (Underfitting)

- 모델링을 너무 간단하게 하여 성능이 제대로 나오지 않는 모델.

모델 복잡도 곡선



해결방법

- 주어진 훈련데이터의 다양성이 보장되어야 한다. 다양한 데이터 포인트를 골고루 나타내야 한다.
- 일반적으로 데이터 양이 많으면 일반화에 도움이 된다.
- 하지만 편중된 데이터를 많이 모으는 것은 도움이 되지 않는다.
- 규제(Regularization)을 통해 모델의 복잡도를 적정선으로 설정한다.

Machine Learning

교차검증

학습



Train
Data

검증



Validation
Data

평가



Test
Data

K-fold cross-validation 동작 방법

Train Data



1번 세
트

2번 세
트

3번 세
트

4번 세
트

5번 세
트

K-fold cross-validation 동작 방법

test



train

Cross-validation 장/단점

- 데이터의 여러 부분을 학습하고 평가해서 일반화 성능을 측정하기 때문에 안정적이고 정확하다.
- 데이터 세트 크기가 충분하지 않은 경우에도 유용하게 사용 가능하다.
- 여러 번 학습하고 평가하는 과정을 거치기 때문에 계산량이 많아진다