

2강. 선형회귀(1)

■ 주요용어

용어	해설
선형회귀	관측값 $\{y_i\}$ 과 이에 해당하는 표적값 $\{t_i\}$ 이 훈련집합으로 주어졌을 때 회귀모델의 목표는 새 변수 x 의 표적값 t 를 예측하는 것인데, 이때 회귀모델로 선형함수를 사용하면 선형회귀라고 한다.
기저함수	선형회귀에서 입력변수에 다음과 같은 고정 비선형 함수들의 선형결합을 사용할 수 있다. $y(\mathbf{x}, \mathbf{w}) = w_0 + \sum_{j=1}^{m-1} w_j \phi_j(\mathbf{x})$ 이 때 $\phi_j(\mathbf{x})$ 를 기저함수(basis function)라고 한다.
정규화항	regularization term을 말함. normalization(정규화)와 혼동하기 쉽기 때문에 원서 그대로 표현하는 것이 좋다. 많은 수의 매개변수를 가진 모델들의 과적합문제를 어느정도 조절하기 위해 매개변수의 크기 자체를 penalty로 주는 항을 오차함수에 추가한다. 이 때 추가된 항을 regularization term이라고 한다.
편향-분산 트레이드오프	기대오류는 편향, 분산, 노이즈로 분해되는데, 유연한 모델은 낮은 편향값과 높은 분산값을 가지며, 엄격한 모델은 높은 편향값과 낮은 분산값을 가진다. 이 둘은 트레이드 오프 관계에 있다.

■ 정리하기

1. 선형회귀는 기저함수들의 선형결합함수를 통해 타겟변수를 예측하는 것이 목표이다.
2. 기저함수는 비선형 함수들을 사용할 수 있어서, 선형결합을 통해서도 복잡한 형태를 표현하는 것이 가능하다. (아래에서 ϕ 가 기저함수)

$$y(\mathbf{x}, \mathbf{w}) = \sum_{j=0}^{M-1} w_j \phi_j(\mathbf{x}) = \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x})$$

3. 시그모이드함수, 가우시안함수등을 기저함수로 사용할 수 있다.
4. 제곱손실함수를 최소화 하는 예측값은 조건부 평균이다.
5. 기저함수를 어떤 형태를 선택하든지, 제곱합 오차함수를 통해 w 를 알아낼 수 있다.

6. 제곱합 오차함수를 전개함으로써 w_0 가 무엇을 보정하는지 알 수 있다.
7. 큰 데이터는 w 를 바로 계산하기 힘들기 때문에 순차학습을 통해 알아낸다.
8. 오차함수에 정규화항을 넣어서 과적합을 막을 수 있다.
9. 정규화항의 차수에 따라 추정되는 w 의 성질이 다르다.(lasso -> sparse)
10. 손실함수를 적절히 전개하여 편향과 분산으로 나눠볼 수 있다.