모델 훈련

1. 선형 회귀: 훈련과정이 단순하여 기초 개념 설명에 유용.
   대다수의 머신러닝 모델이 훈련 과정에서 선형 회귀 활용

   선형 회귀 모델 일반화

   $$\hat{y} = \theta_0 + x_1 \cdot \theta_1 + \cdots + x_n \cdot \theta_n$$

   파라미터: 세타1, 세타2 ......
   편향: 세타0
   가중치: 편향을 제외한 나머지 파라미터

   일반적으로 여러 개의 입력값에 대하여 동시에 예측값을 계산
   모델의 예측 성능을 최대화 하고, 비용 함수를 이용해 성능 계산

   비용 함수

   모델의 성능이 얼마나 나쁘지 평가

   모델의 종류와 목표에 따라 다른 비용 함수 선택

   회귀 모델 : 일반적으로 평균 제곱 오차(MSE)를 비용 함수로 사용

   $$\mathrm{MSE}(\theta) = \frac{1}{m} \sum_{i=1}^{m} \left(\mathbf{x}^{(i)}\,\theta - y^{(i)}\right)^2$$

   최종 목표는 비용함수가 최소가 되는 세타를 찾는 것

2. 다항 회귀: 비선형 데이터를 선형 회귀를 이용해 학습하는 기법

   단점: 몇 차 다항 회귀를 사용해야 할지 알 수 없음, 심층 신경망처럼 비선형 데이터를 분석하는 좋은 모델이 개발되어 사용할 필요가 없음

3. 모델 규제 & 조기 종료: 둘 다 성능을 높이기 위한 방법
   모델 규제:
   릿지 회귀: 가중치의 절댓값을 최대한 작게 유지, 모델의 분산을 줄이지만

편향은 커짐.

비용함수

$$J(\theta) = \text{MSE}(\theta) + \frac{\alpha}{m_b} \sum_{i=1}^{n} \theta_i^2$$

편향은 규제 안함

알파가 규제 강도이고, 이 값을 키우면 가중치의 절댓값은 0에 가까워짐

특성 스케일링 전처리를 해야 규제 모델의 성능이 높아짐

라쏘 회귀: 중요하지 않은 특성의 가중치를 0으로 설정. 자유도가 줄어들어 역시 분산은 줄이고 편향은 커짐.

비용함수

$$J(\theta) = \text{MSE}(\theta) + 2\alpha \sum_{i=1}^{n} |\, \theta_i|$$

덜 중요한 특성을 무시하기 위해 해당 특성의 가중치를 0에 수렴하도록 유도

엘라스틱 넷: 위의 두 회귀의 혼합

엘라스틱 넷 회귀

비용함수

$$J(\theta) = \text{MSE}(\theta) + r \cdot \left(2\alpha \sum_{i=1}^{n} |\, \theta_i \,|\right) + (1-r) \cdot \left(\frac{\alpha}{m_b} \sum_{i=1}^{n} \theta_i^2\right)$$

조기 종료: 훈련 중에 훈련셋에 과대적합하는 것을 방지하는 규제