[23차시]

표준화(Standardization),정규화(Regularization),일반화(Normalization) 의 구별

참조사이트: https://skyil.tistory.com/50

1. Normalization(일반화)

너무 큰 값을 가질 때 weight가 편향되는 문제가 있다. 이에
데이터 값들을 0~1 사이의 값으로 축소하는 것, 식은 아래와 같다.

$$x = \frac{x - x_{min}}{x_{max} - x_{min}}$$

2. Standardization(표준화)

outlier(이상치)의 제거

Standardization을 위해서 먼저 z-score라는 지표를 계산한다.

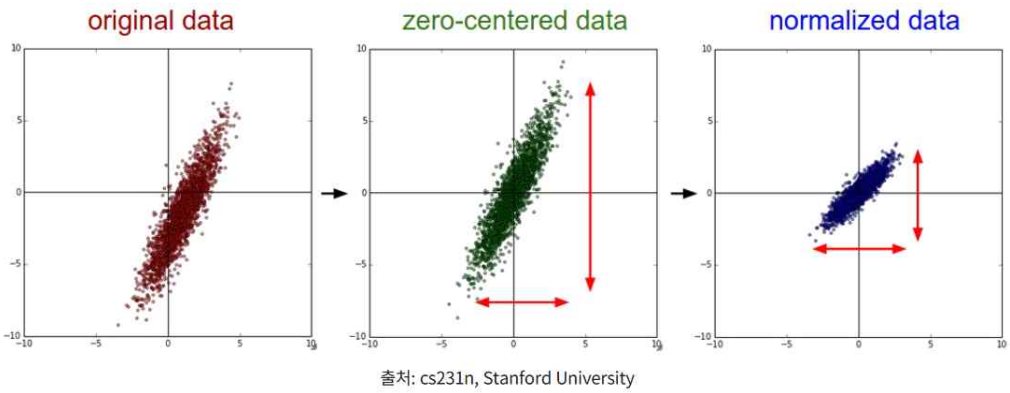
$$z(x) = \frac{x - m}{\sigma}$$

m; 평균

σ; 표준 편차

이렇게 계산된 z-score는 평균에서 크게 벗어난 outlier일수록 큰 절댓값을 갖는다. 이제 이 score가 평균에서 벗어난 값들을 모두 지워주면 된다.

※ Normalization과 Standardization의 활용



출처: cs231n, Stanford University

Normalization은 값을 0과 1 사이로 모아서 값의 규모(scale)를 줄여준다. min과 max의 편차가 크거나, 다른 열에 비해 데이터가 지나치게 큰 열에 사용한다.

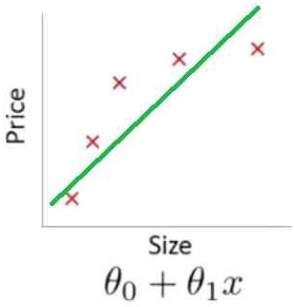
더 작은 scale을 갖는 데이터는 더욱 빠르게 수렴하는 경향이 있어, 머신러닝 학습 성능 향상에 도움을 준다. 특히 MSE와 같이 Cost가 Loss에 비례해 끝없이 커지는 비용 함수에 대해서

는, 비용 값이 너무 커서 발생할 수 있는 문제들을 예방할 수 있다. 또한, 여러 입력 데이터의 scale을 비슷하게 맞춰줌으로써 각 데이터가 비슷한 중요도를 갖도록 학습시킬 수 있다.

Standardization은 z-score의 절댓값 최개치를 넘는 값들을 지워 평균에서 크게 벗어난 outlier를 제거한다. 이는 모델이 대부분의 상황에서 더 높은 precision을 갖게 한다.

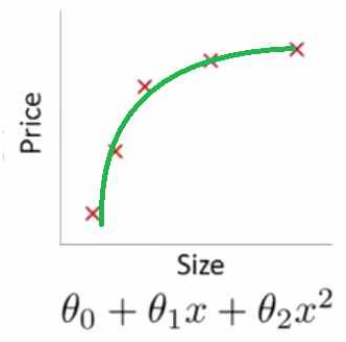
때에 따라선 Standardization 이후에 Normalization을 추가로 하기도 한다.

3. 정규화(Regularization)
- 정규화(regularization)를 통한 과적합(Overfitting) 방지
1) 과적합(overfitting)
가) underfit 또는 high bias의 경우



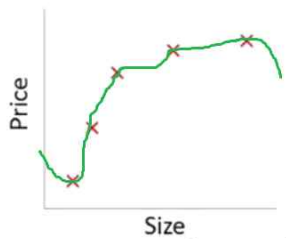
$$\theta_0 + \theta_1 x$$

나) 좋은 모델의 경우



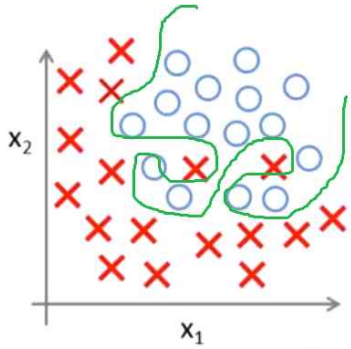
$$\theta_0 + \theta_1 x + \theta_2 x^2$$

다) overfitting, high variance의 경우



$$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$

- 과적합(overfitting)이란 많은 Feature가 있을 때 가설이 학습데이터에만 너무 잘 맞아서 학습데이터가 아닌 일반적인 다른 데이터에 대해서는 제대로 예측하지 못하는 상태를 말한다.
※ 로지스틱 회귀에서의 과적합



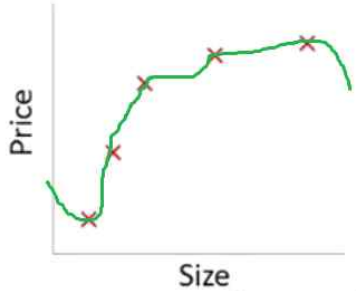
$$g(\theta_0 + \theta_1 x_1 + \theta_2 x_1^2 + \theta_3 x_1^2 x_2 + \theta_4 x_1^2 x_2^2 + \theta_5 x_1^2 x_2^3 + \theta_6 x_1^3 x_2 + \dots)$$

- 과적합(Overfitting) 문제가 발생하는 이유는 다양하다. 이를 해결하기 위한 방법은 크게 2가지가 있다.
가) 특성(Feature)의 갯수를 줄여기
- 쓸만한 특성을 수작업으로 골라낸다.
- 특성을 고르는 알고리즘을 사용한다.
나) 정규화(Regularization)를 수행한다.
- 모든 특성을 사용하되, 파라미터(세타)의 값을 줄인다.
- 특성이 많아도 잘 동작하게 해준다.

■ 정규화(Regularization)
Overfitting이 생긴다고 Feature를 수작업으로 없애는 것은 좋지 않을 수 있다.
이럴때는 정규화(Regularization) 작업을 통해 Overfitting을 조절할 수 있다.
선형회귀에서 다음과 같은 Hypothesis 를 사용한다고 가정

$$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$

과적합인 경우)



$$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$

- 과적합을 최소화하기 위해서 Cost Funcion을 이용해 Parameter인 세타 값을 찾을 때 특정 세타값을 작게 만들어서 영향을 최소화시키는 방법을 사용
- 과적합을 유발하는 파라미터 세타 3과 세타4를 줄여보면

$$\min_{\theta} \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)} \right)^2 + 1000\theta_3^2 + 1000\theta_4^2$$

위 수식을 사용하면 세타3의 값과 세타 4의 값은 점점 0과 가까워진다.
계산 결과가 점점 작아져야하는데 계산할때마다 세타3의 제곱값과 세타4의 제곱값을 계속 더하니까 값이 줄어들려면 세타 3의 값과 세타 4의 값이 작아져야 하는 원리
이런식으로 세타 값을 줄일 수 있고 이 방법을 정규화(Regularization)라고 한다.

선형회귀 정규화)
- 선형회귀의 Cost Function 공식

$$\min_{\theta} \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)} \right)^2 + \lambda \sum_{j=1}^{n} \theta_j^2$$

λ: *regularization parameter*

j: *parameter index*

람다; regularization parameter로 정규화를 어느 정도 시킬건지 정하는 상수값
- 람다값이 너무 크면 underfitting이 나올 수 있음.
- 람다값이 너무 작으면 overfitting이 나올 수 있음.

j; 정규화를 진행하는 parameter의 index값인데 1부터 시작하는 이유는 선형회귀에서 0은 Feature값이 무조건 1이 들어가는 상수 값이기 때문

최종적으로 구한 Cost Function을 이용해 구한 선형회귀의 Gradient Descent 공식은 다음과 같다.

$$\begin{aligned}
&\textbf{Repeat } \{ \\
&\quad \theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_0^{(i)} \\
&\quad \theta_j := \theta_j - \alpha \left[\left(\frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)} \right) + \frac{\lambda}{m} \theta_j \right] \qquad j \in \{1, 2 \ldots n\} \\
&\}
\end{aligned}$$

세타0의 경우는 어차피 상수라서 정규화를 할 필요가 없으므로, 각자 쓴것
이를 미분하면

$$\theta_j := \theta_j \left(1 - \alpha \frac{\lambda}{m}\right) - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)} \right) x_j^{(i)}$$

로지스틱 회귀 정규화)

- 로지스틱 회귀도 선형회귀와 마찬가지로 람다를 이용해 파라미터의 크기를 줄인다.
- 변경된 Cost Function으로, 끝에 람다를 이용한 파라미터를 조정하는 수식이 하나 추가됨

$$J(\Theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\left(h_\theta\left(x^{(i)}\right)\right) + \left(1-y^{(i)}\right)\log\left(1-h_\theta\left(x^{(i)}\right)\right)\right] + \frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2$$