5장 Regression

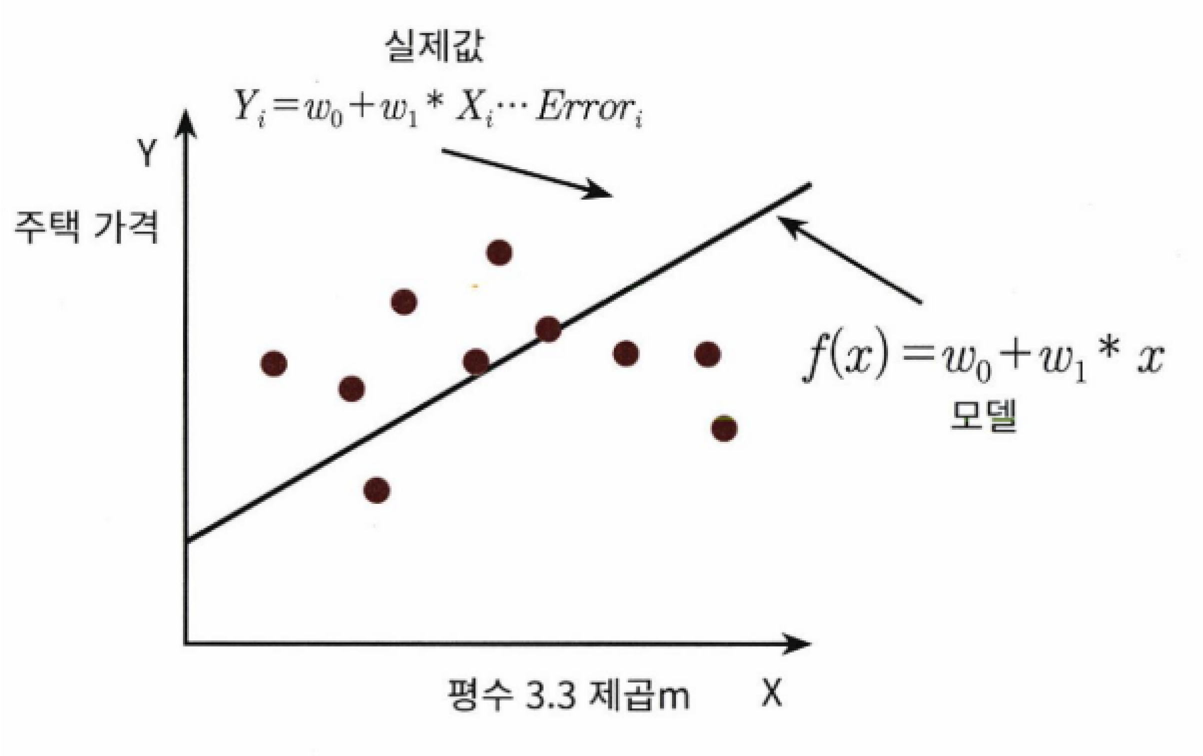
5.1 회귀 소개

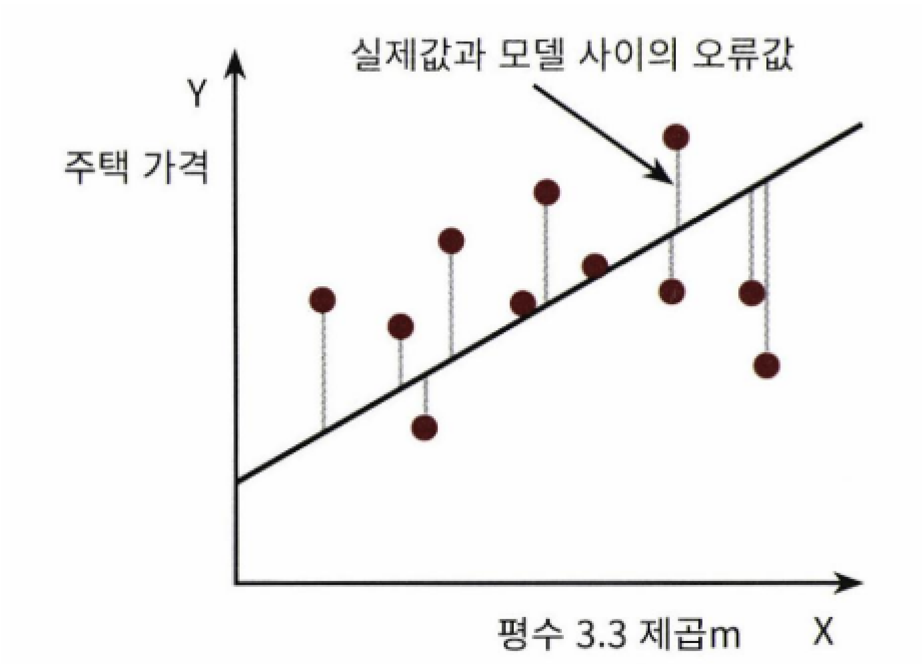
- 회귀 : 여러 개의 독립 변수와 한 개의 종속변수 간의 상관관계를 모델링하는 기법을 통칭한다.
- <회귀 유형 구분>

독립변수 개수	회귀 계수의 결합
1개 : 단일 회귀	선형 : 선형 회귀
여러 개 : 다중 회귀	비선형 : 비선형 회귀

- **선형 회귀** : 실제 값과 예측값의 차이(오류의 제곱 값)를 최소화하는 직선형 회귀선을 최적화하는 방식
- **규제(Regularization)**에 따른 선형 회귀 방법 모델
    - 일반 선형 회귀 : 예측값과 실제 값의 RSS(Residual Sum of Squares)를 최소화 할수 있도록 회귀 계수를 최적화하며, 규제(Regularization)를 적용하지 않은 모델
    - 릿지(Ridge) : 릿지 회귀는 선형 회귀에 L2 규제를 추가한 회귀 모델
    - 라쏘(Lasso) : 라쏘 회귀는 선형 회귀에 L1 규제를 적용한 회귀 모델
    - 엘라스틱넷(ElasticNet) : L2, L1 규제를 결합한 모델
    - 로지스틱 회귀(Logisitic Rregression) : 사실상 분류에 사용되는 회귀 모델(매우 강력한 분류 알고리즘)

5.2 단순 선형 회귀를 통한 회귀 이해





- 목표(최적의 회귀 모델) : 전체 데이터의 잔차(오류 값) 합이 최소가 되는 모델을 만든다는것!

그렇다면 전체 데이터의 잔차(오류 값) 합은 어떻게 계산될 수 있을까?

$$RSS(w_0, w_1) = 1/N * (\sum(y_i - (w_o + w_i * x_i))^2)$$

- RSS: 비용 함수(Cost Function)이라고 부르고, 이 비용 함수가 반환하는 값을 지속해서 감소시키고 최종적으로는 **더 이상 감소하지 않는 최소의 오류 값**을 구하는게 목적이다.

5.3 비용 최소화 하기 - 경사 하강법(Gradient Descent) 소개

지금까지 우리의 목적이 잔차의 최소화 즉 w를 줄이는 것이라는 것을 알았다! 그렇다면 어떻게 줄일 수 있을까?

- 경사하강법 : ‘’점진적으로’ 반복적인 계산을 통해 W 파라미터 값을 업데이트하면서 오류 값이 최소가 되는 W 파라미터를 구하는 방식입니다.
- 핵심 아이디어 : “**어떻게 하면 오류가 작아지는 방향으로 W 값을 보정할 수 있을까**?”

$$R(w) = 1/N * (\sum(y_i - (w_o + w_i * x_i))^2)$$

윗 식을 w_1에 관해 편미분하면 아래와 같다.

$$\frac{\sigma R(w)}{\sigma w_1} = 2/N * (\sum x_i * (y_i - (w_o + w_i * x_i))$$

$$\frac{\sigma R(w)}{\sigma w_1} = -2/N * (\sum x_i * (\text{실제값}_i - \text{예측값}_i)$$

마찬가지로 w_0에 관해 편미분하면 아래와 같다.

$$\frac{\sigma R(w)}{\sigma w_0} = 2/N * (\sum -(y_i - (w_o + w_i * x_i))$$

$$\frac{\sigma R(w)}{\sigma w_0} = -2/N * (\sum(\text{실제값}_i - \text{예측값}_i)$$

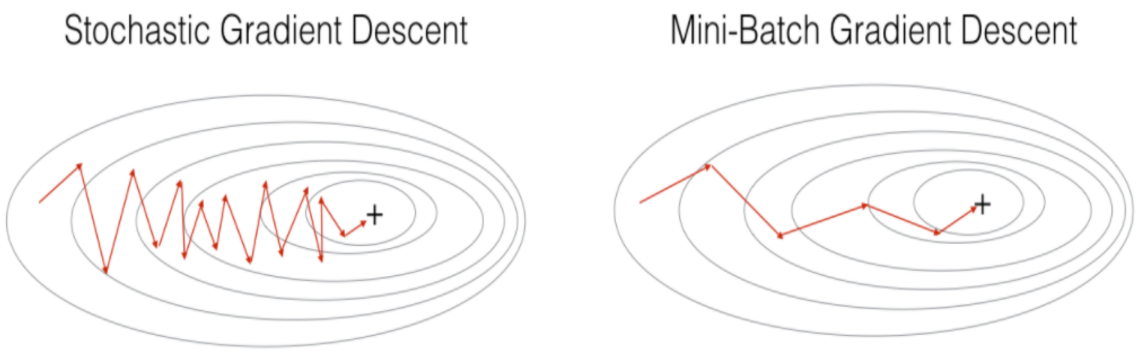
이 후, 이렇게 편미분된 결과값을 마이너스하면서 적용한다.

$$w_1(new) = w_1(old) + \eta\frac{2}{N} * (\sum x_i * (\text{실제값}_i - \text{예측값}_i)$$

$$w_0(new) = w_0(old) + \eta\frac{2}{N} * (\sum(\text{실제값}_i - \text{예측값}_i)$$

- 이를 **업데이트 한 후에 다시 비용함수의 값을 계산**한다. 이 를 반복적으로 수행하며 더 이상 비용 함수의 값이 감소하지 않으면 그때의 **w_1, w_0**을 구하고 반복을 중지한다.

- 일반적인 경사 하강법은 모든 데이터에 대해서 반복적으로 비용함수 최소화를 업데이트 하기 때문에 수행시간이 오래걸린다.
- 개선한 방법 : (미니 배치) 확률적 경사 하강법



- (**미니 배치**) 확률적 경사 하강법 : **일부 데이터**만 이용해 w가 업데이트되는 값을 계산하므로 경사 하강법에 비해서 빠른 속도를 보장합니다.

지금까지는 피쳐가 1개, 독립변수가 1개인 단순 선형 회귀에만 경사 하강법을 적용하였다. 그렇다면 **다중 선형 회귀**에서는 어떤식으로 경사 하강법을 적용할 수 있을까?

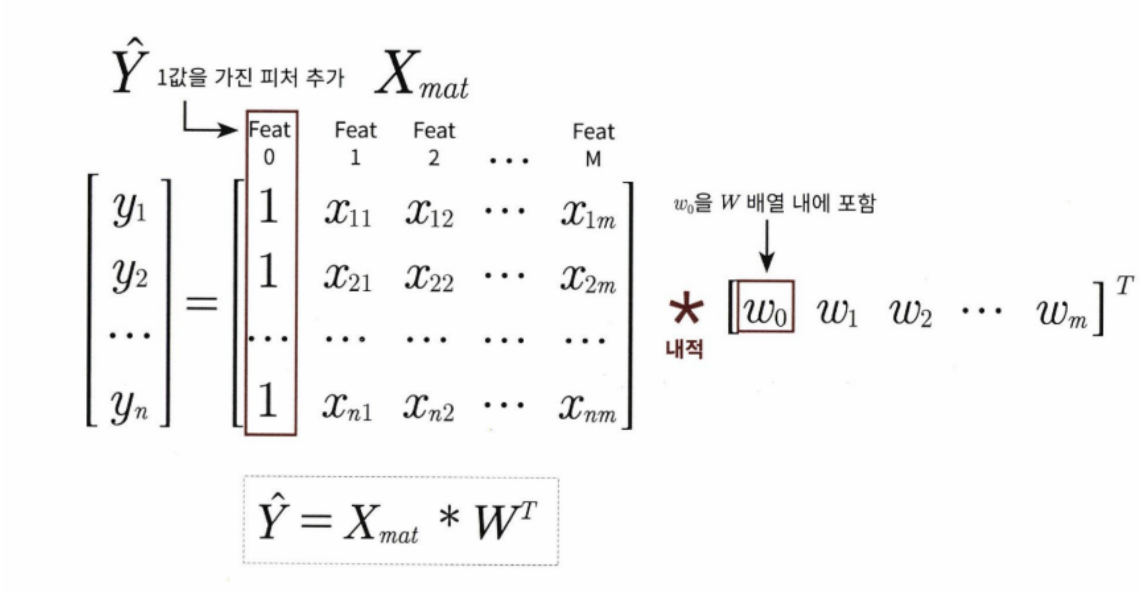
- 만약 피처가 M개(X_1, X_2,,, X_100)있다면 그에 따른 회귀 계수는 M_+1개로 도출된다.

$$\hat{Y} = w_0 + w_1 * X_1 + w_2 * X_2 + \cdots + w_100 * X_100$$

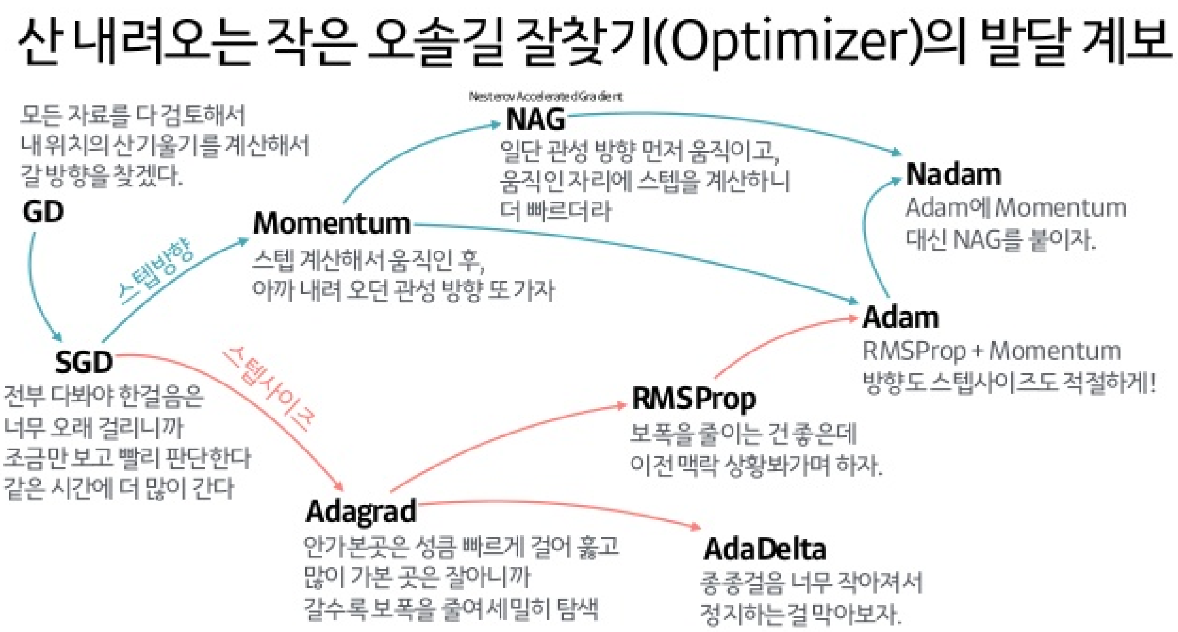
이를 아래와 같은 그림으로 도식화 시킬수 있다.

$$\underset{\hat{Y}}{\begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix}} = \underset{X_{mat}}{\begin{array}{cccc} \text{Feature 1} & \text{Feature 2} & \cdots & \text{Feature M} \\ \end{array}} \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \underset{\text{내적}}{*} \begin{bmatrix} w_1 & w_2 & \cdots & w_m \end{bmatrix}^T + w_0$$

하지만, 위의 그림은 w_0를 포함하지 못하므로 이를 포함시키는 새로운 X_mat을 만들어 준다.



< 대표적인 딥러닝 optimizer 예시)



코드 구현 부분 자세히 살펴보기

5.4 사이킷런 LinearRegression을 이용한 보스턴 주택 가격 예측

5.5 다항 회귀와 과(대)적합/과소적합 이해

5.6 규제 선형 모델 - 릿지(Ridge), 라쏘(Rasso), 엘라스틱넷(ElasticNet)

5.7 로지스틱 회귀

5.8 회귀 트리

5.9 회귀 실습 - 자전거 대여 수요 예측

5.10 회귀 실습 - 캐글 주택 가격 : 고급 회귀 기법