선형 회귀와 경사 하강법

정보컴퓨터공학부 Washington

1. 선형 회귀 란?

1.1 선형 회귀 소개

통계학에서 선형 회귀(Linear Regression)는 종속 변수 y와 한 개 이상의 독립 변수 X와의 선형 상관 관계를 모델링하는 회귀분석 기법이다. 한 개의 독립 변수에 기반한 경우에는 단순 선형 회귀, 둘 이상의 독립 변수에 기반한 경우에는 다중 선형 회귀라고 한다.

선형 회귀는 선형 예측 함수를 이용하여 회귀식을 모델링하며, 알려지지 않은 인자는 데이터로부터 추정하는데 이렇게 만들어진 회귀식을 선형 모델이라고 한다. 선형 회귀는 현재에도 널리 사용되는데, 알려지지 않은 파라미터에 대해 선형 관계를 갖는 모델을 세우는 것이 비선형 관계의 모델을 세우는 것보다 용이하기 때문이다.

1.2 선형 회귀의 예

선형 회귀는 대개 두 가지 분류로 요약할 수 있는데, 첫째, 값을 예측하는 것이 목적인 경우, 선형 회귀를 사용해 데이터에 적합한 **예측 모형**을 개발하고, 개발한 선형 회귀식을 이용해 대응되는 y값이 없는 x값에 대해 y를 예측하기 위해서 사용한다. 둘째, **종속 변수 y와 이와 연관된 독립변수 x_1, x_2, ..., x_n이 존재하는 경우** 선형 회귀 분석을 사용하여 x_i와 y의 관계를 정량화 할 수 있다. 이를 통해 x_i와 y가 전혀 관계없는 변수인지, 추가적인 정보를 제공할 수 있는 변수인지를 판단하여 사용할 수 있다.

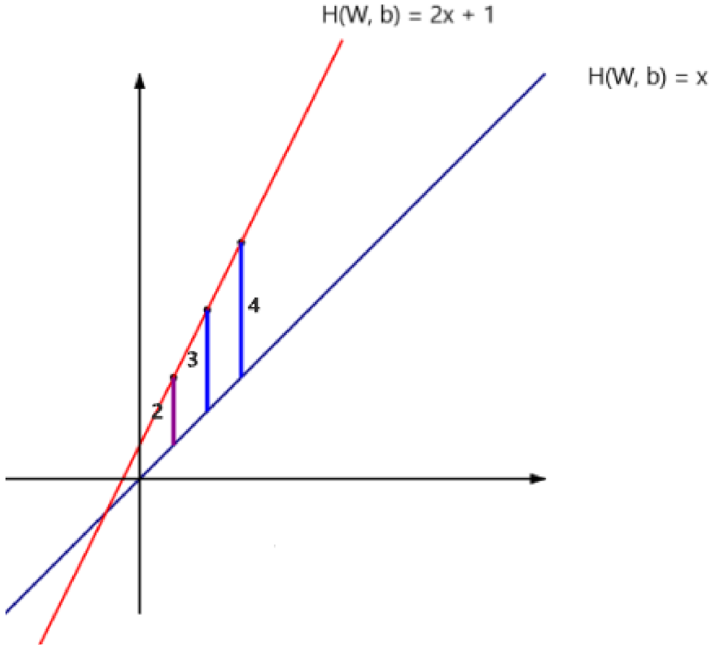
1.3 선형 회귀 모델 구성 기법

선형 회귀 모델을 만드는 기법에는 일반적으로 최소 제곱법(Least square method)이 활용된다. 최소 제곱법 외에도 손실 함수(Loss function)을 최소화하는 방식으로 모델을 만들 수 있다. 최소 제곱법은 선형 회귀 모델뿐만 아니라 비선형 회귀 모델에서도 활용할 수 있는 기법이다.

2. 선형 회귀 모델링

2.1 최소 제곱법

X = [1 3 5]에 대응되는 Y = [3 5 7]이라고 하면 인간은 바로 f(x) = 2x + 1이라는 것을 유추해 낼 수 있으며, H(W, b) = Wx + b라고 함수를 만들면 목표 W는 2이고, b는 1이라는 것을 알 수 있지만, 컴퓨터는 이렇게 바로 함수를 유추해낼 수 없기 때문에 이런 방법을 이용한다. 먼저 가설을 초기화 해준다. 가설 초기화 값이 W = 1, b = 0이라고 하면 Cost(W, b) 함수를 통해서 얼마나 잘못되었는지를 확인하는데 이 때, 최소 제곱법을 활용한다.



[그림 1] 가설 함수와 목표 함수의 차이를 나타내는 그래프

위의 예시 값들과 그림을 통해서 확인하면 초기 가설 세운 값을 통해 얻은 비용이 $(2^2 + 3^2 + 4^2)/3 = 29/3$ 라는 것을 계산할 수 있으며, 목표 값에 가까울 수록 0에 가까워진다는 것을 알 수 있다. 이 최소 제곱법을 통한 비용을 계산하는 방법은

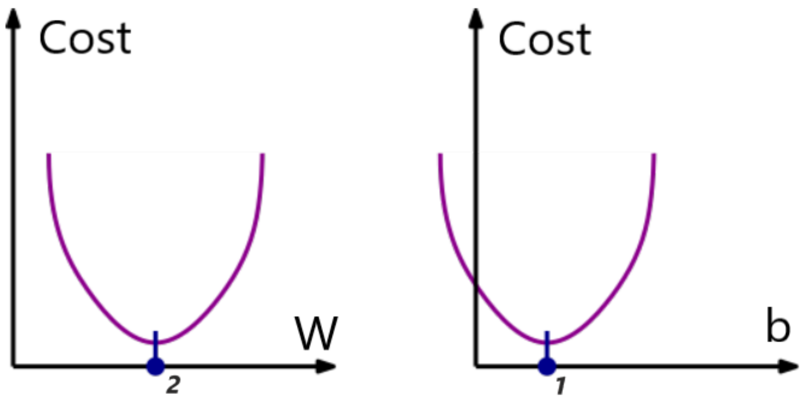
$$Cost(W, b) = 1/m \times \Sigma(Wx_i + b - y_i)^2$$

이다.

최소 제곱법에서 제곱을 하는 이유는 오차를 더 부각시켜서 빠르게 학습이 가능하며, 비용을 계산하기 위해서 절댓값을 이용하면 컴퓨터 내부에서 조건문을 사용하기 때문에 학습 속도가 느려 지는데, 이를 방지하기 위해서 제곱을 사용하여 빠르게 학습을 할 수 있기 때문이다.

2.2 경사 하강법(Gradient Descent)

최소 제곱법으로 계산한 비용을 이용해서 점진적으로 W, b를 목표에 맞게 해야 하는데 이때 사용되는 방법이 경사 하강법이다. 최소 제곱법은 말 그대로 제곱을 사용하므로 기울기의 변화에 따라 목표 값과의 거리가 기하급수적으로 증가 및 감소하게 된다. W, b 각각에 대해 편미분을 적용하여 기울기가 0이 되는 값을 구하면 비용이 최소가 된다. (Global Minima)



[그림 2] W, b 편미분을 통해 각각 기울기가 최소가 되는 값

[그림 2]를 통해서 기울기가 0이 되는 값이 목표 값(Global Optima) W = 2와 동일한 것을 확인할 수 있다. W에 대한 기울기를 구한 것과 같이 b에 대한 기울기도 구할 수 있는데, 각각 따로 갱신해줌으로써 최적의 W와 b를 구할 수 있다. 이처럼 반복적으로 기울기를 구하고 증가 혹은 감소시켜 최종적으로 구하고자 하는 목표 값(Global Optima)에 도달하기 위한 것이 경사 하강법이다.

2.3 수학적 방법

단순 선형회귀는 수학적인 방법으로도 간단하게 해결할 수 있는데, W의 기울기와 b의 기울기를 구하고, 연립방정식으로 두 계산식을 엮어 행렬식으로 변경한 후 역행렬을 크래머 공식을 통하여 적용하면 공분산/분산의 식으로 최적 값을 구할 수 있다. 위의 예와 같이 X = [1 2 3]을 통해 평균 E(X) = 2를 계산하고, Y = [3 5 7]을 통해 E(Y) = 5를 계산하여 공분산을 계산하면,

Cov(X, Y) = ((1-2)(3-5) + (2-2)(5-5) + (3-2)(7-5))/3 = **4/3**

을 계산할 수 있고, 분산 역시

Var(X) = $(1^2 + 2^2 + 3^2)/3 - 2^2$ = **2/3**

으로 계산할 수 있다.

구한 값을 통하여 Cov(X, Y)/Var(X) = 2라는 W 값을 계산할 수 있다. b 값 역시 마찬가지 방식으로 계산할 수 있다.

3. 결론

단순 선형회귀는 수학적인 공식을 이용해서도 Global Optima가 계산이 가능하지만, 실제로는 곡선의 형태가 이처럼 이차함수 형식처럼 간단하게 나오는 것이 아니라 매우 복잡하므로 적용이 어렵고, 식도 매우 복잡하므로 이런 방식으로 해를 구하기 어렵다. 위의 두 가지 방식 이외에도 Global Optima를 구하는 방법은 더 있으나, 현재 복잡한 문제에 있어 경사 하강법(Gradient Descent)이 가장 적절한 해결방법으로 제시되고 있다.