

6.1 로지스틱 회귀분석

로지스틱(Logistic) 회귀분석은 회귀분석이라는 명칭과 달리 회귀분석 문제와 분류문제 모두에 사용할 수 있다. 로지스틱 회귀분석 모형에서는 종속변수가 이항분포를 따르고 그 모수 μ 가 독립변수 x 에 의존한다고 가정한다.

$$p(y | x) = \text{Bin}(y; \mu(x), N)$$

위 식에서 보듯이 로지스틱 함수는 y 의 값이 특정한 구간내의 값($0 \sim N$)만 가질 수 있기 때문에 종속변수가 이러한 특성을 가진 경우에 회귀분석 방법으로 쓸 수 있다.

또는 이항 분포의 특별한 경우($N = 1$)로 y 가 베르누이 확률분포인 경우도 있을 수 있다. 여기에서는 베르누이 확률분포를 따르는 로지스틱 회귀분석만 고려하기로 한다.

$$p(y | x) = \text{Bern}(y; \mu(x))$$

종속변수 y 가 0또는 1인 분류 예측 문제를 풀 때는 x 값을 이용하여 $\mu(x)$ 를 예측한 후 다음 기준에 따라 \hat{y} 값을 출력한다.

$$\hat{y} = \begin{cases} 1 & \text{if } \mu(x) \geq 0.5 \\ 0 & \text{if } \mu(x) < 0.5 \end{cases} \quad \uparrow \text{ 확률}$$

회귀분석을 할 때는 \hat{y} 으로 $y = 1$ 이 될 확률값 $\mu(x)$ 를 직접 사용한다.

$$\hat{y} = \mu(x)$$