

머신러닝(Machine Learning)

지도학습 알아보기(logistic)

학습 목표

▶ 로지스틱 회귀에 대해 알아봅니다.

목차

01 분류형 선형 모델

02 로지스틱 회귀(logistic regression)

01 분류의 선형 모델

- ▶ 로지스틱 회귀(logistic regression)
- ▶ 서포트 벡터 머신(support vector machine)

범주형 변수에서 회귀 모델을 사용하려면 어떻게 해야 할까?

02 로지스틱 회귀(logistic regression)

▶ 로지스틱 회귀(logistic regression)

Y가 범주형(categorical)변수 일때는 다중 선형 회귀 모델을 그대로 적용할 수 없다.

=> 로지스틱 회귀 모델 제안

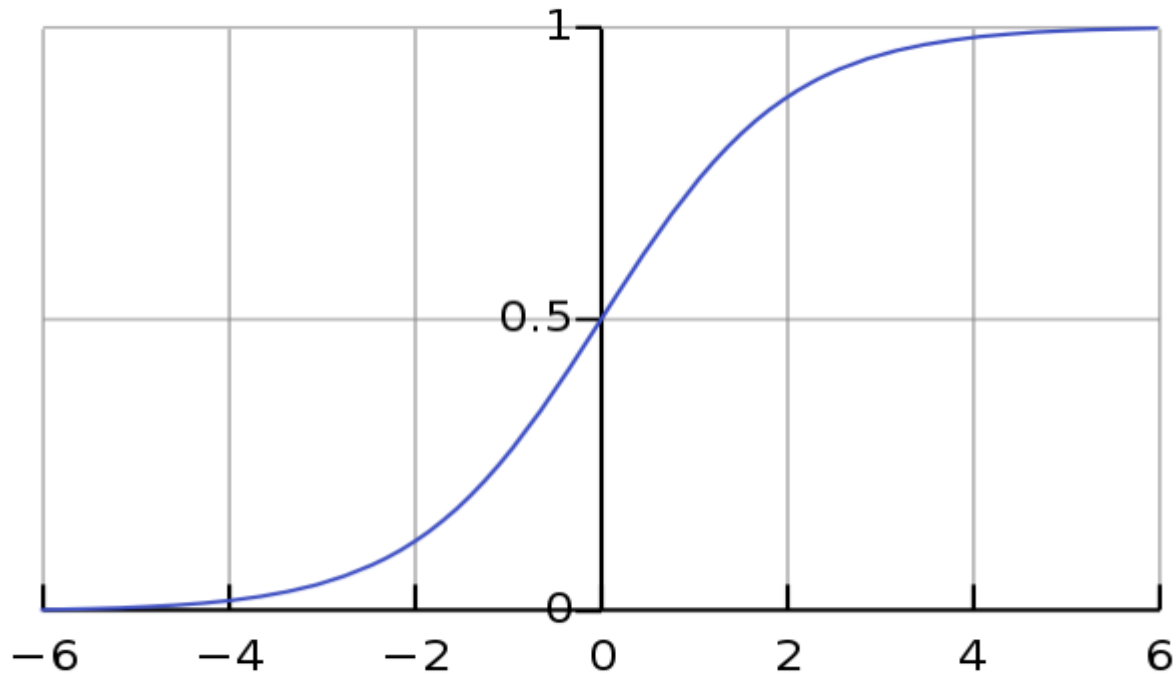
=> 영국의 통계학자인 D.R.Cox가 1958년에 제안한 확률 모델

02 로지스틱 회귀(logistic regression)

▶ 로지스틱 회귀(logistic regression)

회귀를 사용하여 데이터가 어떤 범주에 속할 확률을 0과 1사이의 값으로 예측한다.

로짓 변환을 통해 직선형 Regression을 곡선형으로 적합시킨다.(Fitting)



02 로지스틱 회귀(logistic regression)

▶ 선형 회귀

$$\hat{y}_i = w[0] * x[0] + w[1] * x[1] + \dots + w[p] * x[p] + b$$

▶ 로지스틱 회귀(logistic regression)

$$P(Y = 1) = w[0] * x[0] + w[1] * x[1] + \dots + w[p] * x[p] + b$$

A. $P(Y=1)$ 의 범위는 $0 \sim 1$ 인데, $w[0] * x[0] \dots$ 의 범위는 $-\infty \sim \infty$ 로 맞지 않음.

=> (해결) $P(Y=1)$ 의 범위를 $-\infty \sim \infty$ 로 바꾸기

=> (해결) $P(Y=1)$ 에 Logit를 적용시키기

02 로지스틱 회귀(logistic regression)

▶ 로짓이란 무엇인가?

Logit = Log Odds

▶ Odds란 무엇인가?

$$\text{Odds} = \frac{\text{성공할 확률}}{\text{실패할 확률}} = \frac{P}{1-P}$$

Odds $\in (0, \infty)$

$P \in (0, 1)$



$$\text{Logit}(\text{Odds}) = \log\left(\frac{P}{1-P}\right)$$

Log Odds $\in (-\infty, \infty)$

02 로지스틱 회귀(logistic regression)

▶ $P(Y=1)$ 에 로짓 적용하기

$$\text{Logit}(P(Y=1)) = ax + b = \log\left(\frac{P}{1-P}\right)$$

$$z = \text{Logit}(P) = \log\left(\frac{P}{1-P}\right) = ax + b$$

02 로지스틱 회귀(logistic regression)

▶ 로지스틱 함수는 Logit 함수의 역함수이다.

$$\log\left(\frac{P}{1-P}\right) = ax + b$$

$$\Rightarrow \frac{P}{1-P} = e^{ax+b}$$

$$\Rightarrow P = (e^{ax+b}) * (1 - P)$$

$$\Rightarrow P(1 + e^{ax+b}) = e^{ax+b}$$

$$\Rightarrow P = \frac{e^{ax+b}}{1+e^{ax+b}} = \frac{e^z}{1+e^z} = \frac{1}{1+e^{-z}}$$