- 로지스틱 회귀분석은 추론과 분류문제에 모두 활용될 수 있다. 로지스틱 회귀모형에서는 종속변수가 이항분포를 따르고 그 모수 μ가 독립변수 x에 의존한다고 가정한다.

$$p(y|x) = \underbrace{Bin}_{\text{이항분포}}(y \mid \mu(x), \underbrace{N}_{y\text{는 N개의 구간에서의 횟수}})$$

- 또는 이항 분포의 특별한 경우 $(N=1)$로 y가 베르누이 확률분포인 경우로 있을 수 있다. 여기서는 베르누이 확률 분포를 따르는 로지스틱 회귀분석만 고려한다.

$$p(y|x) = Bern(y \mid \mu(x))$$

- 종속변수 y가 0 또는 1인 분류 예측 문제를 풀때 x값에 따라 $\mu(x)$를 예측한 후 다음 기준에 따라 $\hat{y}$를 예측한다.

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

- 로지스틱 회귀 모형에서는 베르누이 확률분포의 모수 μ가 x의 함수라고 가정한다.

$$\mu = f(w^T x)$$

- 시그모이드 함수 (Sigmoid function)
  - 로지스틱 회귀모형에서는 베르누이 확률분포의 모수 μ가 x의 함수라고 가정한다. $\mu(x)$는 x에 대한 선형함수를 0부터 1 사이의 값만 나올 수 있도록 하는 시그모이드 함수를 이용한다.
  - 시그모이드 함수는 두 가지 공통특성을 갖는다.
    - 유한한 구간의 한정된 (bounded) 값과,
    - 항상 0 또는 양의 기울기를 가지는 함수이다.

  - 로지스틱 함수

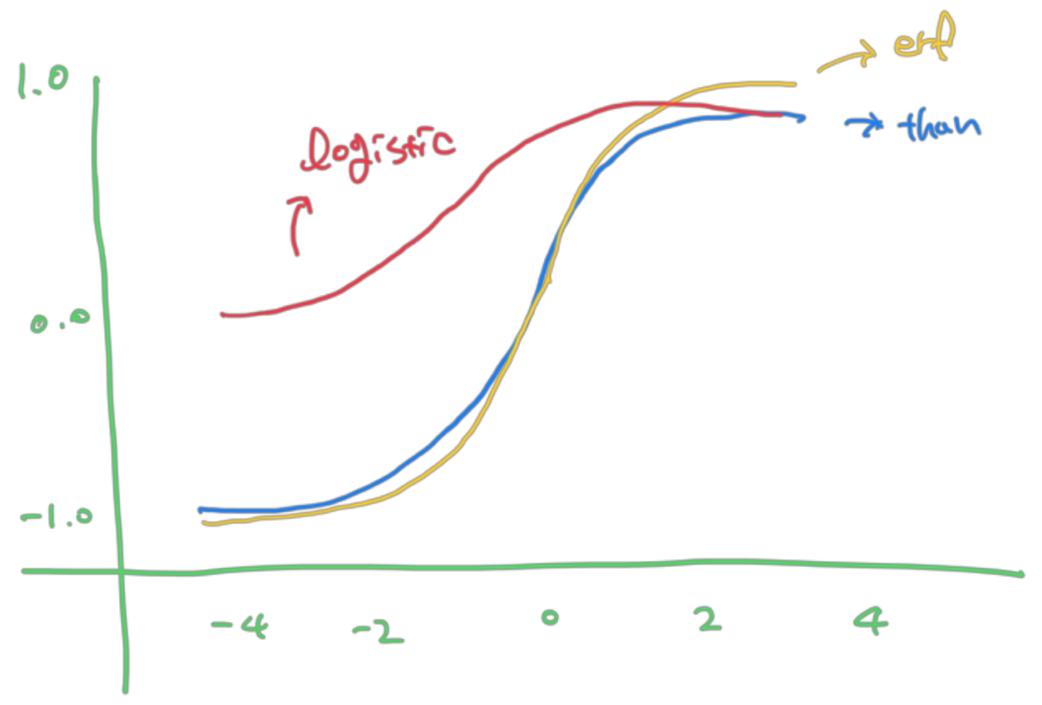
$$logistic(z) = \sigma(z) = \frac{1}{1+\exp(-z)}$$

  - 하이퍼볼릭 탄젠트 함수 (Hyperbolic tangent)

$$\tanh(z) = \frac{\sinh z}{\cosh z} = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}} = 2\sigma(2x) - 1$$

  - 오차함수 (Error Function)

$$erf(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$$



- 로지스틱 함수는 베르누이 시도에서 1이 나올 확률 μ과 0이 나올 확률 $1-\mu$의 비(ratio)를 odds ratio라 한다.

$$odds\ ratio = \frac{\mu}{1-\mu}$$

- 0부터 1사이의 값만 가지는 μ를 odds ratio로 변환하면 0부터 ∞의 값을 가질 수 있다. 이 odds ratio를 로그 변환한 것이 로지트 함수 (logit function)이다.

$$z = logit(odds\ ratio) = \log\left(\frac{\mu}{1-\mu}\right)$$

- 로지트 함수는 로그변환에 의해 $-\infty$에서 ∞의 값을 가질 수 있다.
- 로지스틱 함수는 이 로지트함수의 역함수 이다. 즉 $-\infty$에서 ∞의 값을 가지는 변수를 0에서 1까지의 값으로 변환한 결과이다.

$$logistic(z) = \mu(z) = \frac{1}{1+\exp(-z)}$$