Week 3

Sunday, August 10, 2025 3:40 PM

1. 로지스틱 함수.

로지스틱 회귀모델 → 관측치를 같은 범주로 분류.

$f(x) = 1/1+e^{-(\beta_0+\beta_1 x)}$: 로지스틱 / Sigmoid / Squashing func.

→ $0 \le f(x) \le 1$

→ input에 연속증가

$E(y) = \pi(X=x) = 1/1+e^{-(\beta_0+\beta_1 x)}$

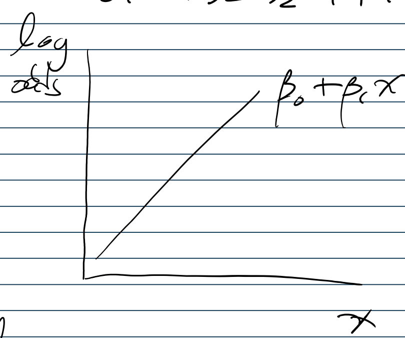
2. 승산 odds

: 실패 대비 성공 확률 $\frac{p}{1-p}$

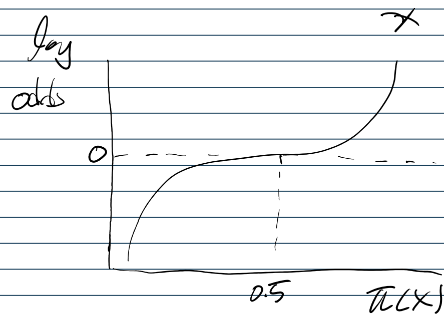
$\pi(X=x) = 1/1+e^{-\beta_0-\beta_1 x}$ → $odds = \frac{\pi(X=x)}{1-\pi(X=x)}$

: 범주 0에 속할 확률 대비 1에 속할 확률.

$\underbrace{\log(odds) = \beta_0+\beta_1 x}_{\text{logit transform}}$



⇒ β_1 : x가 한단위 증가시 log(odds)의 증가량



3. 로지스틱 회귀 모델 파라미터 추정

$$\boxed{\pi(x) = 1/1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2 \cdots \beta_p x_p)}}$$

최대우도 추정법 Maximum Likelihood Est.

$f_i(y_i) = \pi(x_i)^{y_i}(1-\pi(x_i))^{1-y_i}$

$L = \prod f_i(y_i)$ $\quad \ln(L) = \sum_i y_i \ln\left[\frac{\pi(x_i)}{1-\pi(x_i)}\right] + \sum \ln(1-\pi(x_i))$

$= \sum_i y_i(\beta_0+\beta_1 x_1 + \cdots \beta_p x_p) + \sum_i \ln(1+e^{\beta_0+\beta_1 x_1+\cdots \beta_p x_p}$

$\ldots \, 1-\pi(x_i) \,] + \sum \ldots$

$= \sum_i y_i (\beta_0 + \beta_1 x_1 + \cdots \beta_p x_p) + \sum_i \ln(1 + e^{\beta_0 + \beta_1 x_1 + \cdots \beta_p x_p}$

→ $\ln L$; log-likelihood func가 최대가 되는 β 결정.

비선형이므로 명시적인 해가 존재하지 않음.

Cross Entropy $= -\sum p(x) \log q(x)$: 두 확률분포의 차이

= 음의 log likelihood의 기댓값

log likelihood 최대 = cross entropy 최소.

4. 결과 및 해석

$\log(\text{odds}) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_p x_p$: 입력변수 1단위 증가 시 log(odds) 변화량.

Odds Ratio

$\dfrac{\text{odds}(x_1 + 1, x_2 \cdots x_n)}{\text{odds}(x_1, x_2 \cdots x_n)} = e^{\beta_1}$: 나머지 입력 x 고정, 한 변수 1단위 증가 시 변화하는 odds의 비율.

- $\beta > 0$ → 성공확률 증가
- $\beta < 0$ → 〃 감소.

ex) feature① − coef $\begin{cases} > 0 & : \text{성공확률과 양의 상관} \\ < 0 & : \text{성공확률과 음의 상관} \end{cases}$

− odds ($\text{Exp} = 1.058$) : 점수 1단위 더 많으면 더블 확률이 1.058배 증가