양성 샘플일 확률 $= P(\underbrace{y=1}_{\text{양성}} | \vec{x}) = P$

Odds ratio(오즈비) $= \dfrac{P}{1-P} = \dfrac{\text{양성일 확률}}{\text{음성일 확률}}$

$\dfrac{0}{1-0} = \dfrac{0}{1} = \boxed{0 \le \text{오즈비} < \infty} \approx \dfrac{1}{0} = \dfrac{1}{1-1}$

(단, $P \neq 0$ (by $\frac{P}{1-P} \neq 0$))

$\mathrm{logit}(P(y=1|\vec{x})) = \mathrm{logit}(P) = \log \dfrac{P}{1-P}$

$= W_0 x_0 + W_1 x_1 + \dots + W_m x_m = \sum_{i=0}^{m} W_i x_i = \vec{w} \cdot \vec{x} = \vec{w}^T \vec{x}$

(단, $x_0 = 1$)

$\log \dfrac{0}{1-0} = \log 0 \approx \boxed{-\infty < \mathrm{logit}(P) < \infty} \approx \log \dfrac{1}{0} = \log \dfrac{1}{1-1}$

$\vec{w}^T \vec{x} = \mathrm{logit}(P) = z$ 라 하자

$\rightarrow \mathrm{logit}(P) = \log \dfrac{P}{1-P} = z \rightarrow e^z = \dfrac{P}{1-P} \xrightarrow{P \neq 1} e^z(1-P) = P$

$\rightarrow e^z - e^z P = P \rightarrow e^z = (1+e^z)P \rightarrow P = \dfrac{e^z}{1+e^z}$

$\rightarrow P = \dfrac{e^z}{1+e^z} = \dfrac{e^z(e^{-z})}{(1+e^z)(e^{-z})} = \dfrac{1}{e^{-z}+1}$

$\rightarrow P = P(y=1|\vec{x}) = \dfrac{1}{1+e^{-z}} = \dfrac{1}{1+e^{-\mathrm{logit}(P)}} = \dfrac{1}{1+e^{-\vec{w}^T\vec{x}}}$

$$\boxed{P(y=1|\vec{x}) = \frac{1}{1+e^{-\vec{w}^T\vec{x}}} = \frac{1}{1+e^{-z}} = \phi(z) = \phi(\vec{w}^T\vec{x})}$$

such that (s.t)

$P(y=1|\vec{x};\vec{w}) = P(y=1|\vec{x}) = \phi(z) = \phi(\vec{w}^T\vec{x})$

→ 모수가 $\vec{w}$ 일때
입력값이 $\vec{x}$ 일때
$y=1$ 일 확률

모수가 $\vec{w}$ 인 확률밀도 함수 ϕ 는
$\vec{x}$ 입력되었을 때 $y=1$ 일 확률을
함숫값으로 가짐

(가능도)
최대 우도 추정법(MLE)으로
입력데이터 $\vec{x}_i$에 대응하는 y_i 데이터가 n개 입력될 때
$y=1$일 확률밀도함수 ϕ의 모수 $\vec{w}$을 추정해보자

$\underbrace{\text{가능도 함수}}_{\text{Likelihood}} = L(\vec{w}) = P(\vec{y} \mid X; \vec{w})$

→ 모수가 $\vec{w}$일 때 X가 입력되었을 때 $\vec{y}$가 나올 확률(가능도)

$$L(\underbrace{\vec{w}}_{\text{모수 } \vec{w}\text{가 변수}}) = P(\vec{y} \mid X; \vec{w}) = \prod_{i=1}^{n} P(\overbrace{y_i}^{y=y_i} \mid \vec{x}_i; \vec{w})$$

$$= \prod_{i=1}^{n} \left(P(y=1 \mid \vec{x}_i; \vec{w})\right)^{y_i} \left(1 - P(y=1 \mid \vec{x}_i; \vec{w})\right)^{1-y_i}$$

$$= \prod_{i=1}^{n} \left(\phi(z_i)\right)^{y_i} \left(1-\phi(z_i)\right)^{1-y_i}$$

$$= \prod_{i=1}^{n} \left(\phi(\vec{w}^T \vec{x}_i)\right)^{y_i} \left(1-\phi(\vec{w}^T \vec{x}_i)\right)^{1-y_i}$$

가능도(Likelihood)에 $\overbrace{\log}^{\text{로그 가능도 함수}}$를 취해주자

→ $\log$ Likelihood $= l(\vec{w}) = \log L(\vec{w})$

$$= \log \prod_{i=1}^{n} \left(\phi(z_i)\right)^{y_i} \left(1-\phi(z_i)\right)^{1-y_i}$$

$$= \sum_{i=1}^{n} \left[\log \phi(z_i)^{y_i} + \log\left(1-\phi(z_i)\right)^{1-y_i}\right]$$

$$= \sum_{i=1}^{n} \left[y_i \log(\phi(z_i)) + (1-y_i)\log(1-\phi(z_i))\right]$$

이때 로그 가능도 함수 $l(\vec{w})$를 경사 상승법을 활용하여
극댓값을 갖게 하는 $\vec{w}$를 찾는다. → 최대 로그 가능도 추정법

⇒ $-l(\vec{w})$를 경사하강법을 활용하여
극솟값을 갖게 하는 $\vec{w}$를 찾는다.

→ Cost function(비용함수) $= J(\vec{w}) = -l(\vec{w})$

$$J(\vec{w}) = -\ell(\vec{w}) = -\sum_{i=1}^{n} \overbrace{\left[y_i \log(\phi(z_i)) + (1-y_i)\log(1-\phi(z_i)) \right]}^{=E_i}$$

$$= \sum_{i=1}^{n} \left[-y_i \log(\phi(\vec{w}^T \vec{x}_i)) - (1-y_i) \log(1-\phi(\vec{w}^T \vec{x}_i)) \right]$$

$\rightarrow \frac{1}{n} J(\vec{w})$ = Binary Cross Entropy (BCE)

$$\frac{\partial J}{\partial \vec{w}} = -\frac{\partial \ell}{\partial \vec{w}} = -\frac{\partial \ell}{\partial \phi} \frac{\partial \phi}{\partial z_i} \frac{\partial z_i}{\partial \vec{w}}$$

$$= -\sum_{i=1}^{n} \left[\frac{\partial E_i}{\partial \phi} \frac{\partial \phi}{\partial z_i} \frac{\partial z_i}{\partial \vec{w}} \right]$$

$$= -\sum_{i=1}^{n} \left[\left(\frac{y_i}{\phi(z_i)} + \frac{-(1-y_i)}{1-\phi(z_i)} \right) \phi(z_i)(1-\phi(z_i)) \vec{x}_i \right] = 0$$ 을

만족하는 $\vec{w}$ 구하기