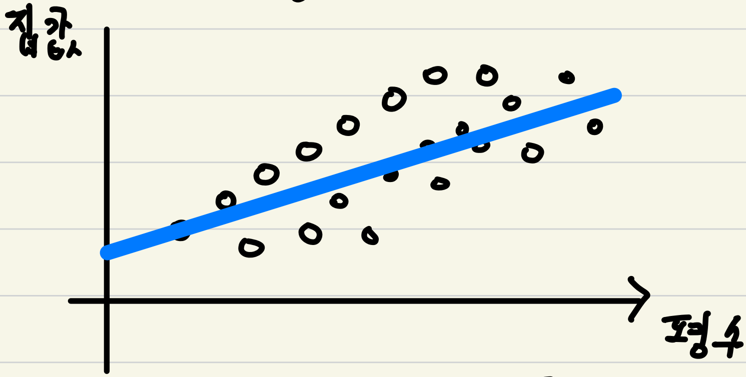
Logistic Regression

Linear Regression.

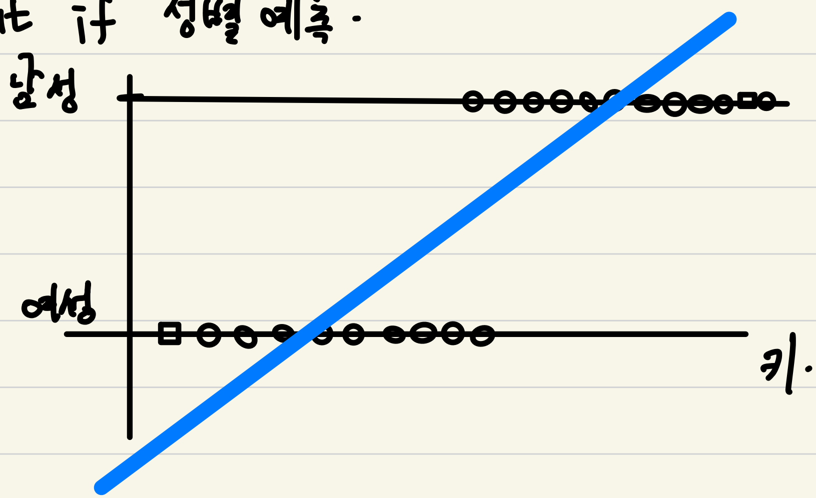
$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k. = WX + b.$$

example) 집값 예측.



if) 변수가 '평수' 하나라면? → 파란 선

but if 성별 예측.



⇒ 너무 이상하다...?

⟶ 확률적 관점으로 보자!

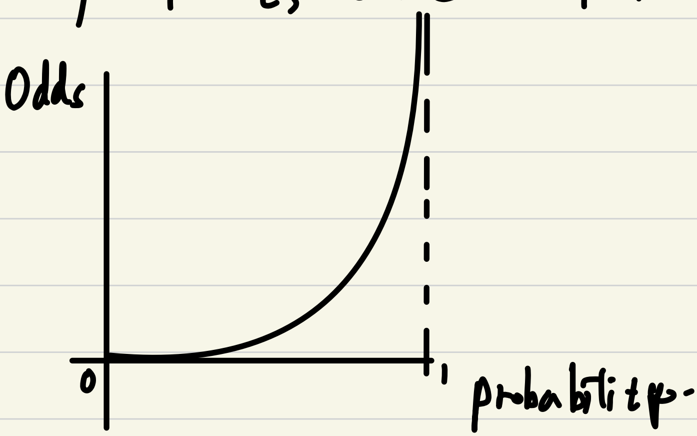
확률 : Probability : 0~1 사이에 분포.

* Linear Regression은 $-\infty$에서 ∞ 사이에 분포.
* 확률도 $-\infty \sim \infty$ 사이에 재분포하자.

승산 : Odd : $\frac{p}{1-p}$: $0 \sim \infty$ 사이에 분포.

$$\Rightarrow \frac{\text{내가 이길 확률}}{\text{내가 질 확률}}$$

$p \sim 1$ 이면, 승산은 ∞에 가까워짐.

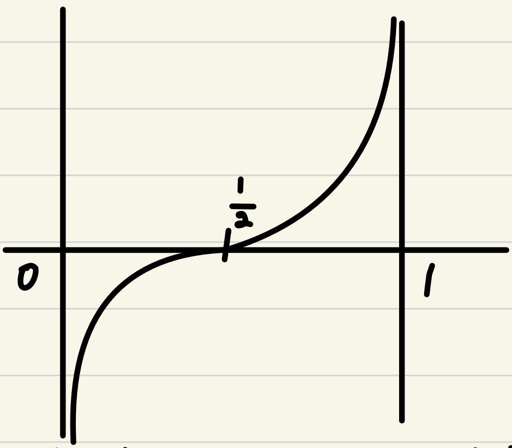


* 여전히 $-\infty \sim \infty$는 아님

log 승산 · log Odds : $\log\left(\frac{p}{1-p}\right)$

* log는 0에 가까이 가면 $-\infty$가 됨.
* p는 1에 〃 ∞가 됨,

  log ∞도 ∞가 됨.



$\log(0) \sim -\infty$

$\log\left(\frac{\frac{1}{2}}{\frac{1}{2}}\right) = \log 1 = 0$

$\log(1) \sim \infty$.

* 이제 $-\infty \sim \infty$ 사이에 분포 시켰다.
* 이를 수식으로 나타내면

$$\text{log-odds} = \log\left(\frac{p}{1-p}\right) = WX + b$$

확률로 풀어내면.

$$\frac{p}{1-p} = e^{(wx+b)} \Rightarrow p = (1-p)e^{(wx+b)} \Rightarrow (1+e^{wx+b})p = e^{wx+b}$$

$$\Rightarrow p = \frac{e^{wx+b}}{1+e^{wx+b}} = \frac{1}{1+e^{-wx+b}}$$

→ logistic function

$$\boxed{\frac{e^x}{1+e^x} = \frac{1}{\frac{1}{e^x}+1} = \frac{1}{1+e^{-x}}}$$

확률 변형값 = 선형 예측치

logistic function 정리

$$\mathrm{logit}(Y|X) = \mathrm{logit}(p) = \ln\frac{p}{1-p} = WX+b.$$

X 가 주어졌을 때, Y=1 이 될 확률

$$= \mathrm{logit}\left(Y=1 \,\middle|\, x_{1,1},\ x_{2,1},\ \cdots\ x_{n,1}\right)$$

$$\ln\left(\frac{p}{1-p}\right) = WX+b.$$

$$\downarrow$$

$$p = \frac{1}{1+e^{-(WX+b)}}$$

$$\downarrow$$

확률 질량함수로 해석

$$\Pr(Y_i = y_i \mid X_i) = p_i^{y_i}(1-p_i)^{1-y_i}$$

y가 1일 때, 1일 확률

y=0일 때, 0일 확률.

Logistic Regression 결정 결계 (Binary)

$$\underbrace{P(Y=1 \mid X)} > \underbrace{P(Y=0 \mid X)}$$ 이면 1이다.

$$P(x) > 1 - P(x).$$

$$\Rightarrow 2P(x) > 1$$

$$\Rightarrow P(x) > \frac{1}{2}.$$

$$\Rightarrow \frac{1}{1+e^{-wx+b}} > \frac{1}{2}.$$

$$\Rightarrow e^{-(wx+b)} = 1$$

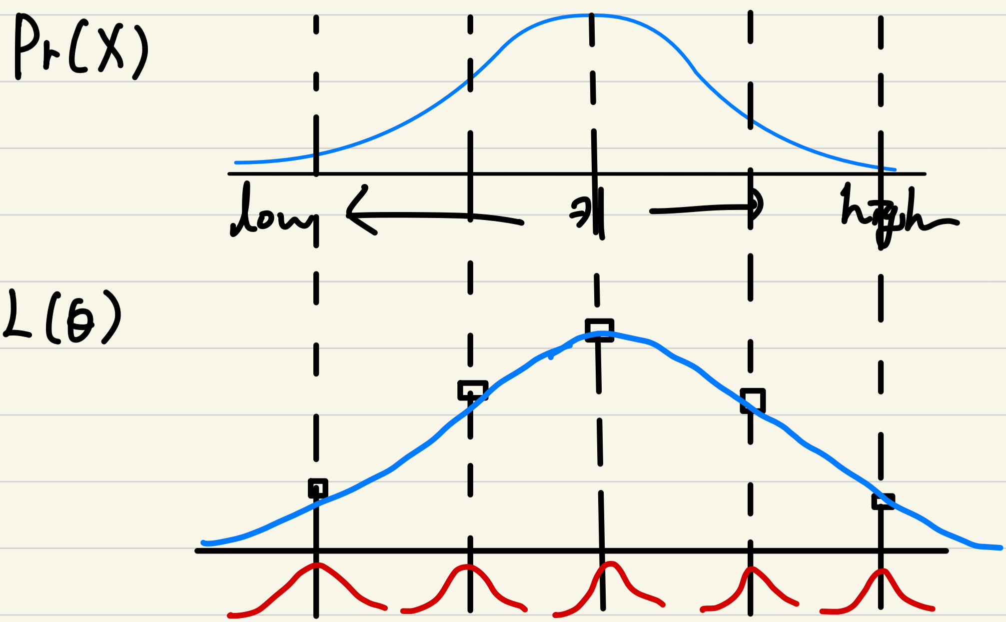
$\Rightarrow wx+b = 0$ 이면 결정 경계.

Logistic Regression 추정

목표: Maximum likelihood Estimation (MLE) 이해.
최대화 우도 추정.
↳ ...?

MLE란 '어떤 모수가 주어졌을 때, 원하는 값이 나올 우도를 최대화 하는 방법'

그림으로 이해



어디가 likelihood를 최대로 하는지 추정하자.
↳ 어떤 값이 주어졌을 때, 이 값이 어떤 확률 분포에서 왔는지 추정

Likelihood = 우도 : 주어진 표본 x를 통해

모집단의 모수 θ를 예측한 값.

$$L(\theta|x) = P_r(X=x|\theta)$$

확률변수 X과 모수 θ에 대해 $P_\theta(X)$를 가짐.

확률함수

$$h_\theta(X) = \frac{1}{1+e^{-\theta X+b}} = P_r(Y=1 \mid X;\theta)$$

우도함수.

p-4 참고

$$L(\theta|x) = P_r(Y|X;\theta) = \prod_i P_r(y_i|x_i;\theta)$$

$$= \prod_i h_\theta(x_i)^{y_i}(1-h_\theta(x_i))^{(1-y_i)}$$

log-우도함수 → log함수는 단조증가함수 → 계산 easy.

$$\log(L(\theta|x)) = \sum_i^N \log P_r(y_i|x_i;\theta)$$

Negative Log Likelihood (NLL).

$$-\frac{1}{N}\sum_{n=1}^N \left[y_n \log \hat{y}_n + (1-y_n)\log(1-\hat{y}_n)\right]$$

⇒ Cross Entropy와 동일.

⇒ Gradient Descent로 minimize 가능.

NLL with Gradient Descent.

$$-\frac{1}{N}\sum_{n=1}^{N}\left[y_n \log \hat{y}_n + (1-y_n)\log \hat{y}_n\right].$$

① $y_n = 0$

$$NLL(Y=0) \rightarrow -\frac{1}{n_0}\sum_{n_0=0}^{0\text{의 수}}\left[0 + (1-0)\log \hat{y}_n\right]$$

② $y_n = 1$

$$NLL(Y=1) \rightarrow -\frac{1}{n_1}\sum_{n_1=0}^{1\text{의 수}}\left[1 \cdot \log \hat{y}_n + 0\right]$$

⋮

즉, y 값을 정확히 예측할 확률

if) $y_n = 1$ $\begin{cases} \hat{y}_1 \sim 0 \rightarrow \log(0) \sim \infty \rightarrow \text{값이 커짐} \\ \hat{y}_1 \sim 1 \rightarrow \log(1) \sim \infty \rightarrow \text{작아짐.} \end{cases}$

실제 값에 대한 예측을 정확히 할 수록, 값이 작아짐

→ 이 값을 최소화 하도록 Gradient Descent 적용.

⇒ 최대한 모수와 유사해지도록 만들면 예측도 좋아짐