로지스틱회귀

- 권수태 교수

JEONJU UNIVERSITY
The Place for Superstars

1. 로지스틱회귀

❖ 다른 변수값을 예측 또는 추정한다면

- 수학 60점이니까 물리는 70점이겠다(계량=>계량) : 회귀분석
- 영어가 550점이니까 불합격 하겠네(계량=>명목) : 로지스틱 회귀분석
- 남자니까 검은색 좋아하겠네(명목=>명목) : 로그선형모형

1. 로지스틱회귀

❖ 로지스틱 회귀

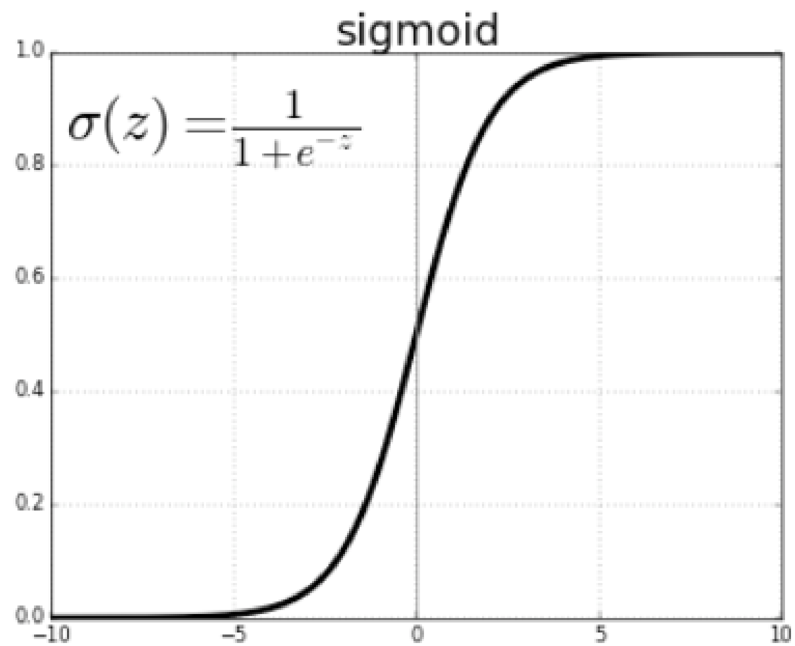
- 사건이 발생할 확률을 결정하는 데 사용되는 통계 모델
- 로지스틱 회귀는 머신러닝에서 정확한 예측을 생성하는 데 사용
- 종류
  - ✓ 이진 로지스틱 회귀: 범주형 (합격, 불합격)
  - ✓ 다항 로지스틱 회귀: 범주형 (3개 이상)
  - ✓ 순서 로지스틱 회귀: 다항 회귀와 마찬가지로 3개 이상의 변수가 있으며, 측정에는 순서가 있음(1-5 척도로 맛집 평가)

1. 로지스틱회귀

❖ 로지스틱 회귀

- 로지스틱 함수 또는 로짓 함수를 x와 y 사이의 방정식으로 사용하는 통계 모델
- 로짓 함수는 y를 x의 시그모이드 함수로 매핑

$$f(x) = \frac{1}{1 + e^{-x}}$$



1. 로지스틱회귀

❖ 로지스틱 회귀

- 여러 독립 변수를 사용한 로지스틱 회귀 분석
  - 여러 설명 변수가 종속 변수의 값에 영향을 미치는데, 이러한 입력 데이터 세트를 모델링하기 위해 로지스틱 회귀 공식은 여러 독립 변수 간의 선형 관계를 가정

$$y = \mathrm{f}(w_0 + w_1x_1 + w_2x_2 + \cdots + w_nx_n)$$

$$y = \frac{1}{1 + e^{-(w_0+w_1x_1+w_2x_2+\cdots+w_nx_n)}}$$

$$y = \mathrm{f}(w_0 + w_1x_1) = \frac{1}{1 + e^{-(w_0+w_1x_1)}}$$

1. 로지스틱회귀

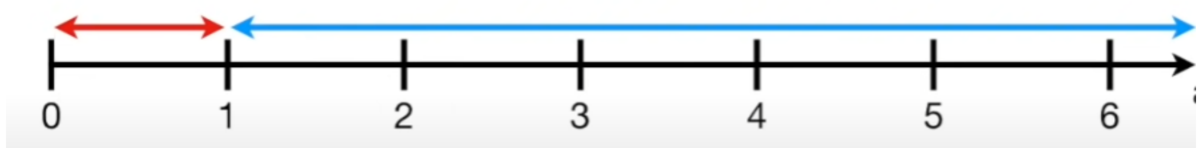
❖ odds

➢ 무엇인가 일어나는 비율

✓ 1번 이기고 4번 지는 경우 = ¼

✓ ¼=0.25, 1/8=0.125, 1/16=0.062, 1/32=0.031 -> 0 에 수렴

✓ 4/3=1.3, 8/3=2.7, 32/3=10.7 -> ∞ 에 수렴

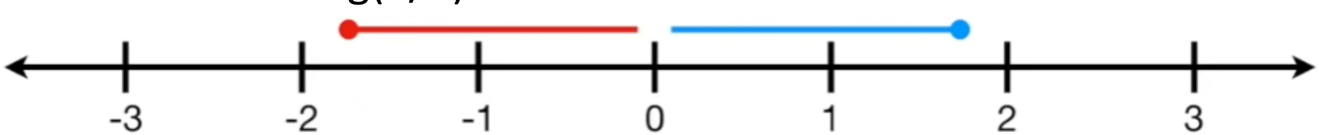


1/6=0.17

6/1=6

Log(1/6)= -1.79

Log(6)= 1.79



1. 로지스틱회귀

❖ odds

➢ 무엇인가 일어나는 비율

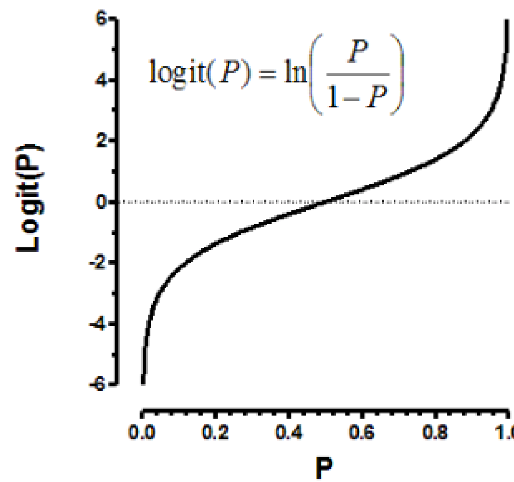
✓ 5번 이기고 3번 지는 경우, odds=$\frac{5}{3}$

✓ 이긴확률= $\frac{5}{8}$, 진확률= $\frac{3}{8}$ $\qquad \frac{p}{1-p} = \frac{5}{3}$

➢ 어떤 사건이 일어날 확률을 그 사건이 일어나지 않을 확률로 나눈 것

$$odds(p) = \frac{p}{1-p}$$

$$logit\ 함수 = log_e\left(\frac{p}{1-p}\right)$$



JEONJU UNIVERSITY

1. 로지스틱회귀

❖ 시그모이드 함수

$$log_e\left(\frac{p}{1-p}\right) = log_e\left(\frac{1}{1/p-1}\right) = log_e(1) - log_e(1/p-1)^{-1}$$

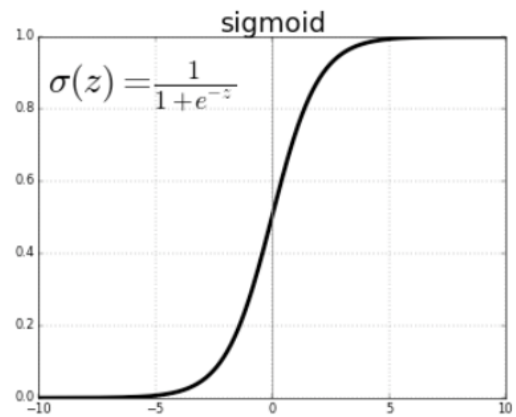
$$= log_e(\frac{1}{p} - 1)$$

$$z = -log_e(\frac{1}{p} - 1)$$

$$e^{-z} = \frac{1}{p} - 1$$

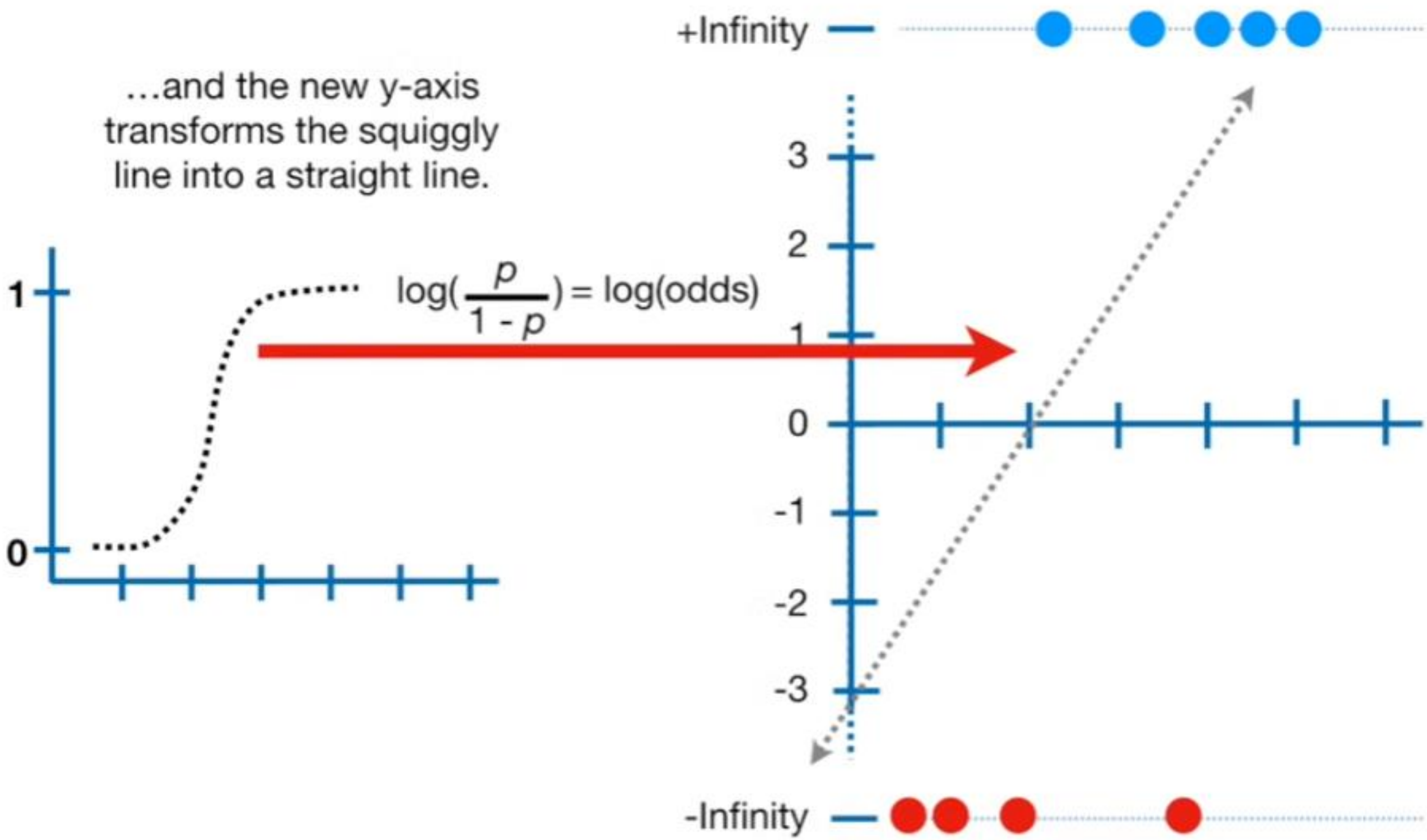
$$e^{-z} + 1 = \frac{1}{p}$$

$$p = \frac{1}{e^{-z} + 1}$$



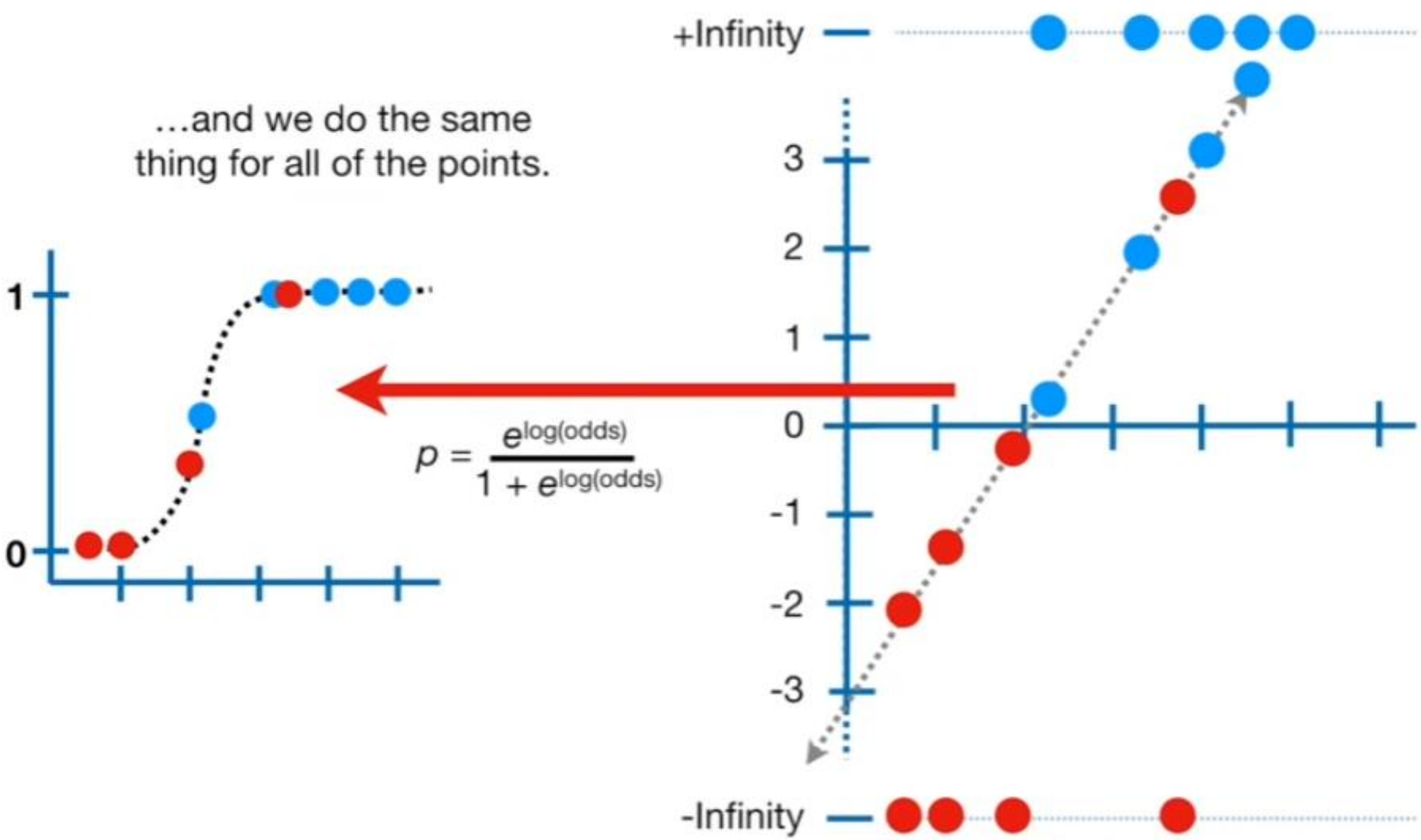
1. 로지스틱회귀

❖ 시그모이드 함수



1. 로지스틱회귀

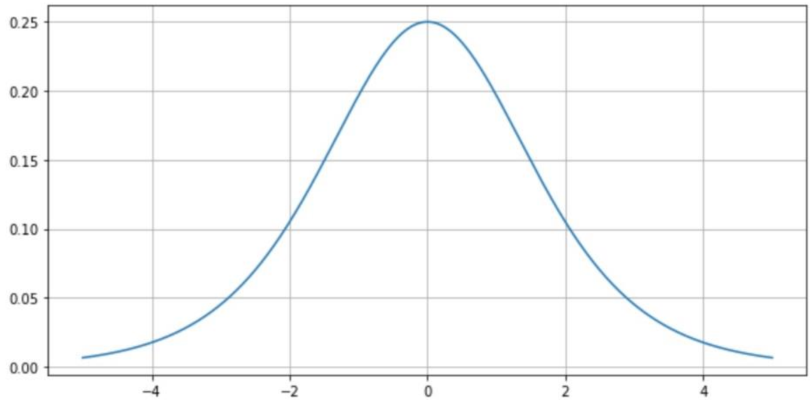
❖ 시그모이드 함수



1. 로지스틱회귀

❖ 시그모이드 함수

➢ 미분



$$
\begin{aligned}
\frac{d}{dx} sigmoid(x) &= \frac{d}{dx}(1+e^{-x})^{-1} \\
&= (-1)\frac{1}{(1+e^{-x})^2}\frac{d}{dx}(1+e^{-x}) \\
&= (-1)\frac{1}{(1+e^{-x})^2}(0+e^{-x})\frac{d}{dx}(-x) \\
&= (-1)\frac{1}{(1+e^{-x})^2}e^{-x}(-1) \\
&= \frac{e^{-x}}{(1+e^{-x})^2} \\
&= \frac{1+e^{-x}-1}{(1+e^{-x})^2} \\
&= \frac{(1+e^{-x})}{(1+e^{-x})^2} - \frac{1}{(1+e^{-x})^2} \\
&= \frac{1}{1+e^{-x}} - \frac{1}{(1+e^{-x})^2} \\
&= \frac{1}{1+e^{-x}}(1 - \frac{1}{1+e^{-x}}) \\
&= sigmoid(x)(1 - sigmoid(x))
\end{aligned}
$$

JEONJU UNIVERSITY

1. 로지스틱회귀

❖ 시그모이드 함수(신경망)

- 아무리 큰 값이 들어온다 하더라도 0-1 사이의 값만 반환하므로, 값이 일정비율로 줄어들어 값의 왜곡이라 할 수는 없으나, 값이 현저하게 줄어듬
- 출력값의 중앙값이 0이 아닌 0.5이며 모두 양수이기 때문에 출력의 가중치 합이 입력의 가중치 합보다 커지게 됨
- 신호가 각 레이어를 통과할 때마다 분산이 계속 커지게 되어, 활성화 함수의 출력이 최대값과 최소값인 0과 1에 수렴
- 시그모이드 함수의 도함수는 $\sigma(1-\sigma)$ 인데, 도함수에 들어가는 함수의 값이 0이나 1에 가까울수록 당연히 출력되는 값이 0에 가까워지게 됨

1. 로지스틱회귀

❖ 시그모이드 함수(신경망)

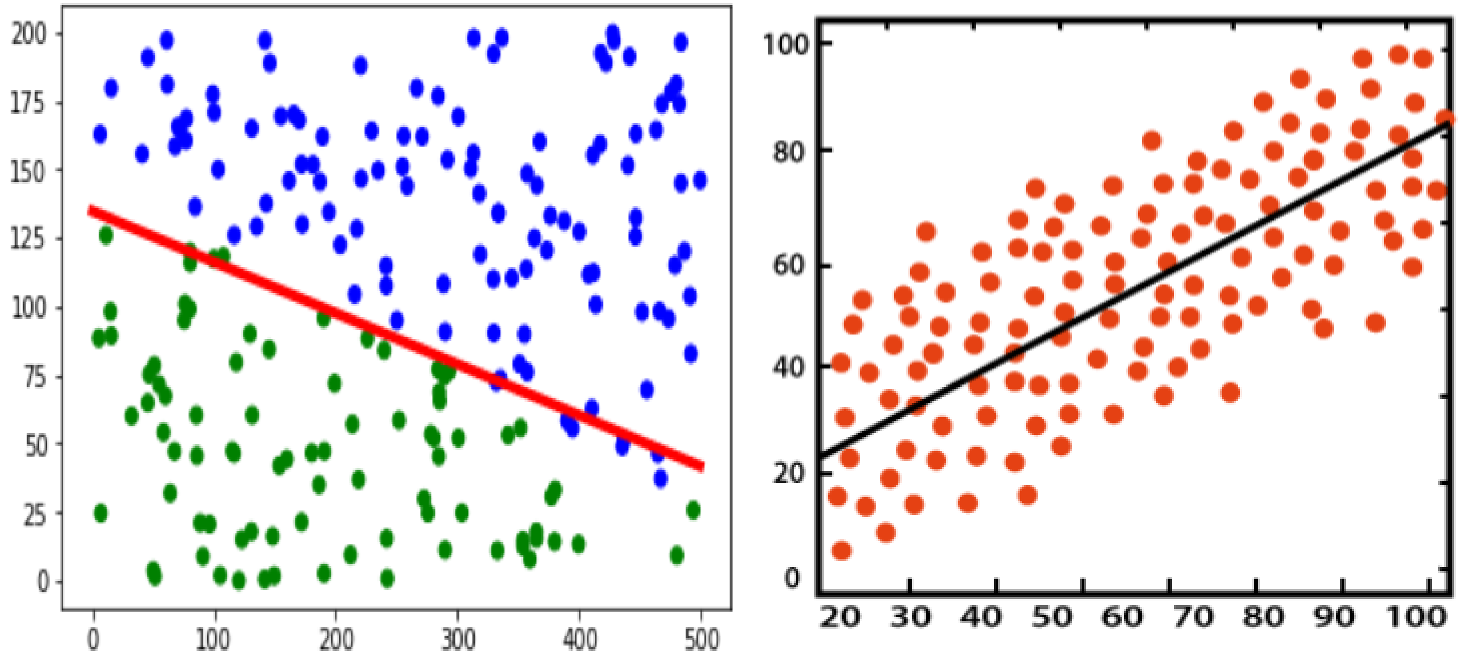
- 이로 인해 수렴되는 기울기 값이 0이되고, 역전파시 0이 곱해져서 기울기가 소멸되는 현상 발생

  -> 렐루함수 도입(1986-2006 해결되지 않은 문제)
- 출력값은 모두 양수이기 때문에 경사하강법을 진행할때, 그 기울기가 모두 양수이거나 음수가 됨
- 이는 기울기 업데이트가 지그재그로 변동하는 결과를 가져오고, 학습 효율성을 감소시킴
- 은닉층에서는 선형함수와 시그모이드 함수는 사용하지 않는 것이 좋음
- 시그모이드 함수는 이진 분류를 하고자 하는 경우 출력층에서만 사용하는 것을 권고

JEONJU UNIVERSITY

1. 로지스틱회귀

❖ 선형회귀와 분류의 차이



Classification Regression

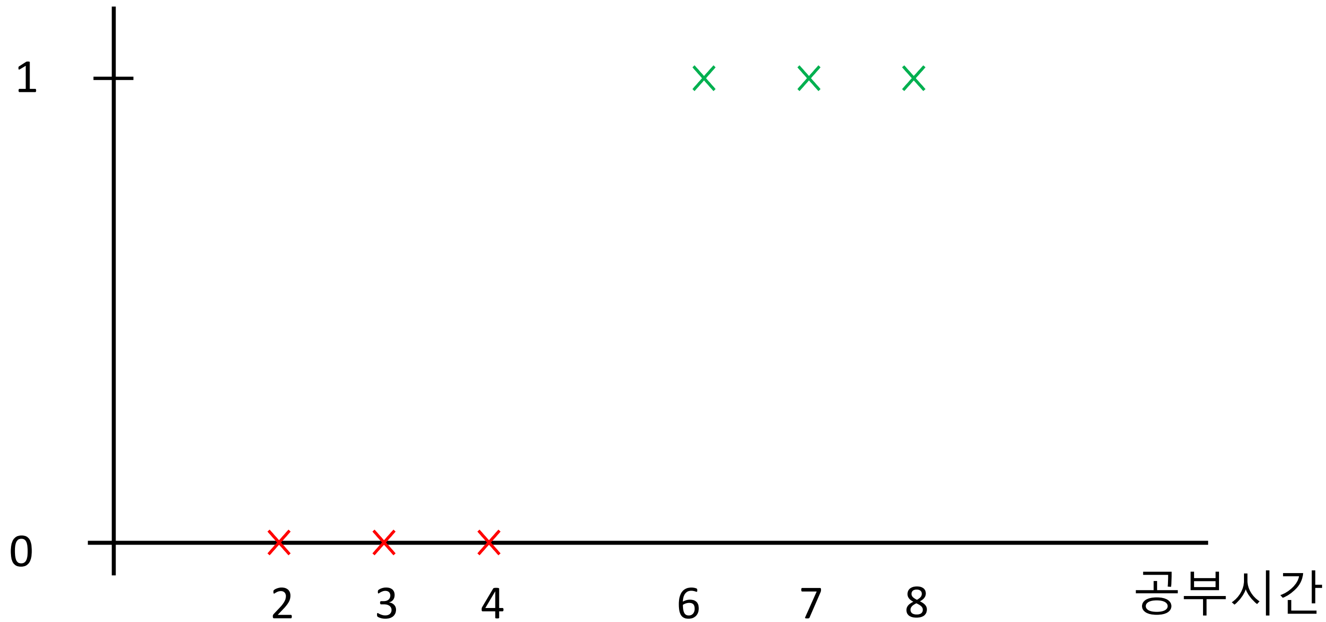
1. 로지스틱회귀

❖ 선형회귀를 이용한 이진 분류

➢ 공부시간에 따른 pass/fail 분류

✓ 공부시간 6, 7, 8시간 : pass(1)

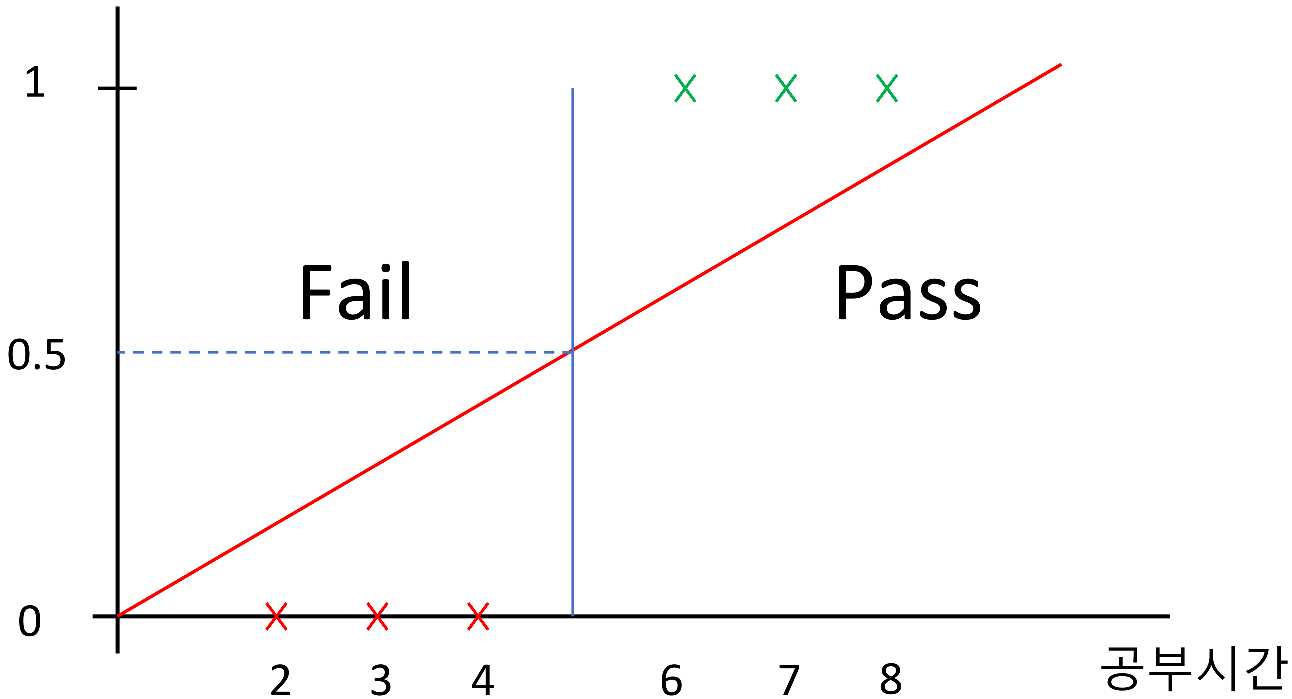
✓ 공부시간 2, 3, 4시간 : fail(0)



1. 로지스틱회귀

❖ 선형회귀를 이용한 이진 분류

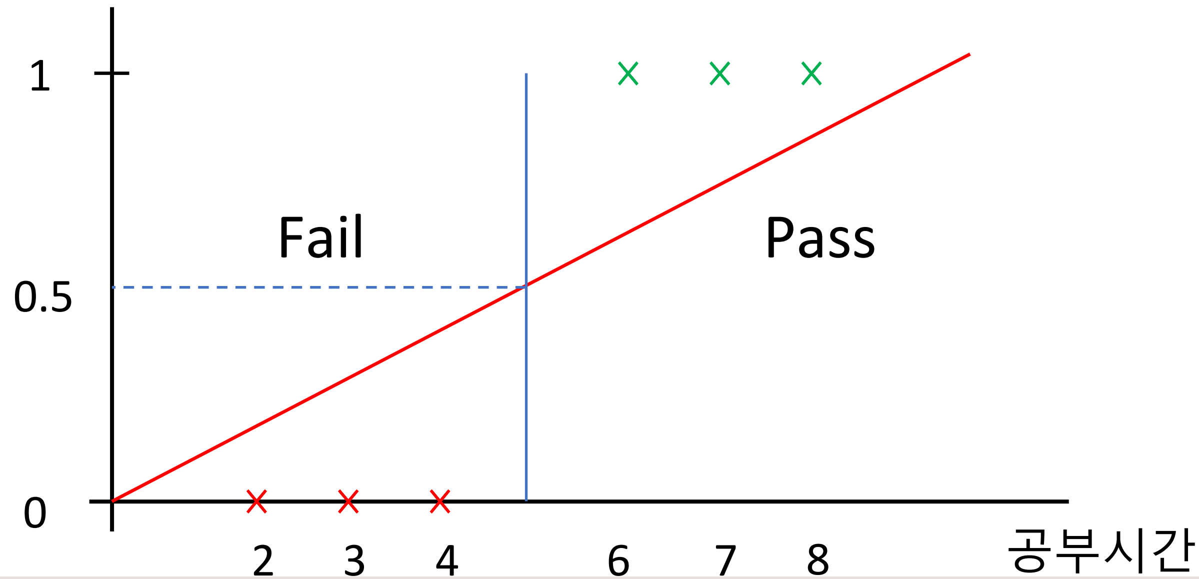
- ➢ 공부시간에 따른 pass/fail 분류
  - ✓ Label encoding : pass(1) , fail(0)



1. 로지스틱회귀

❖ 선형회귀를 이용한 이진 분류

- 공부시간에 따른 pass/fail 분류
  - ✓ 분류를 위한 문턱값을 0.5로 설정 가능
  - ✓ $h(X) \geq 0.5$ 이면 y=1 , pass로 예측
  - ✓ $h(X) < 0.5$ 이면 y=0 , fail로 예측

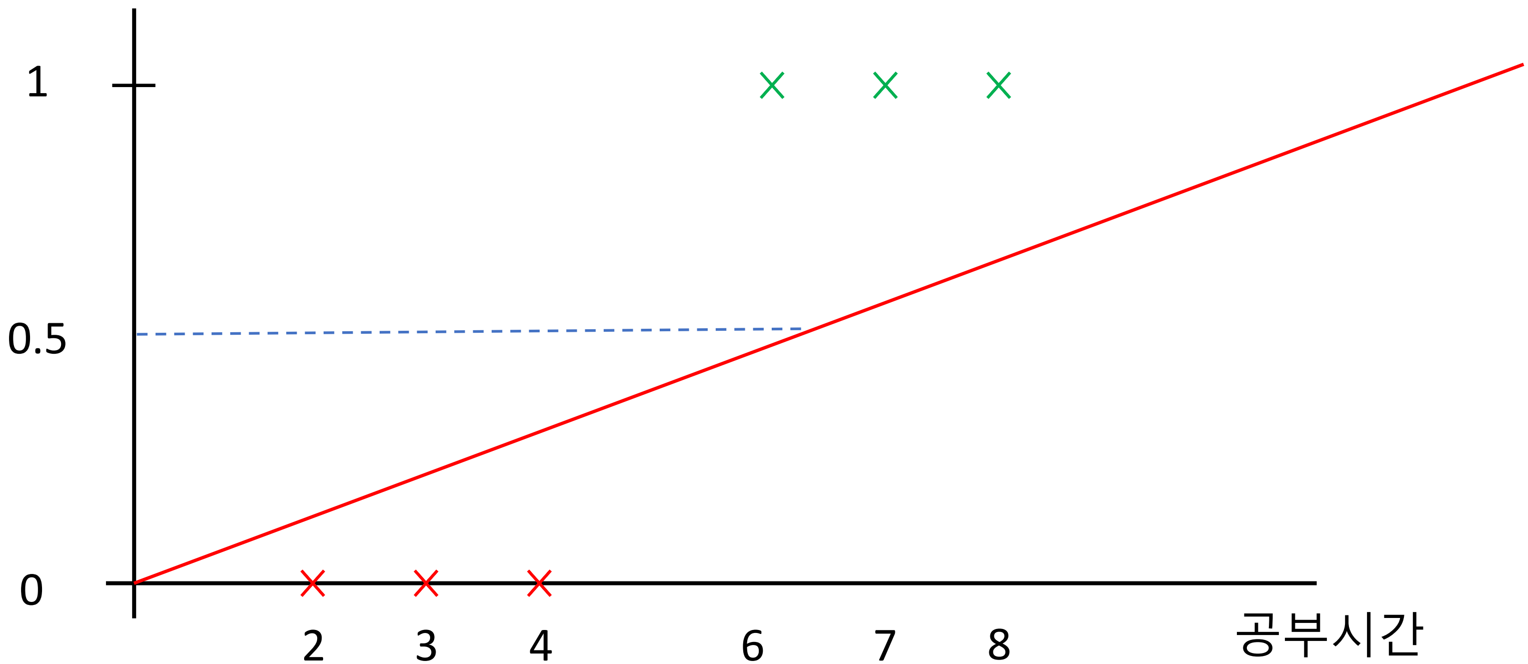


1. 로지스틱회귀

❖ 선형회귀를 이용한 분류의 한계

➢ 학습자료에 따른 문제점

✓ 공부시간 30시간 (pass) 추가

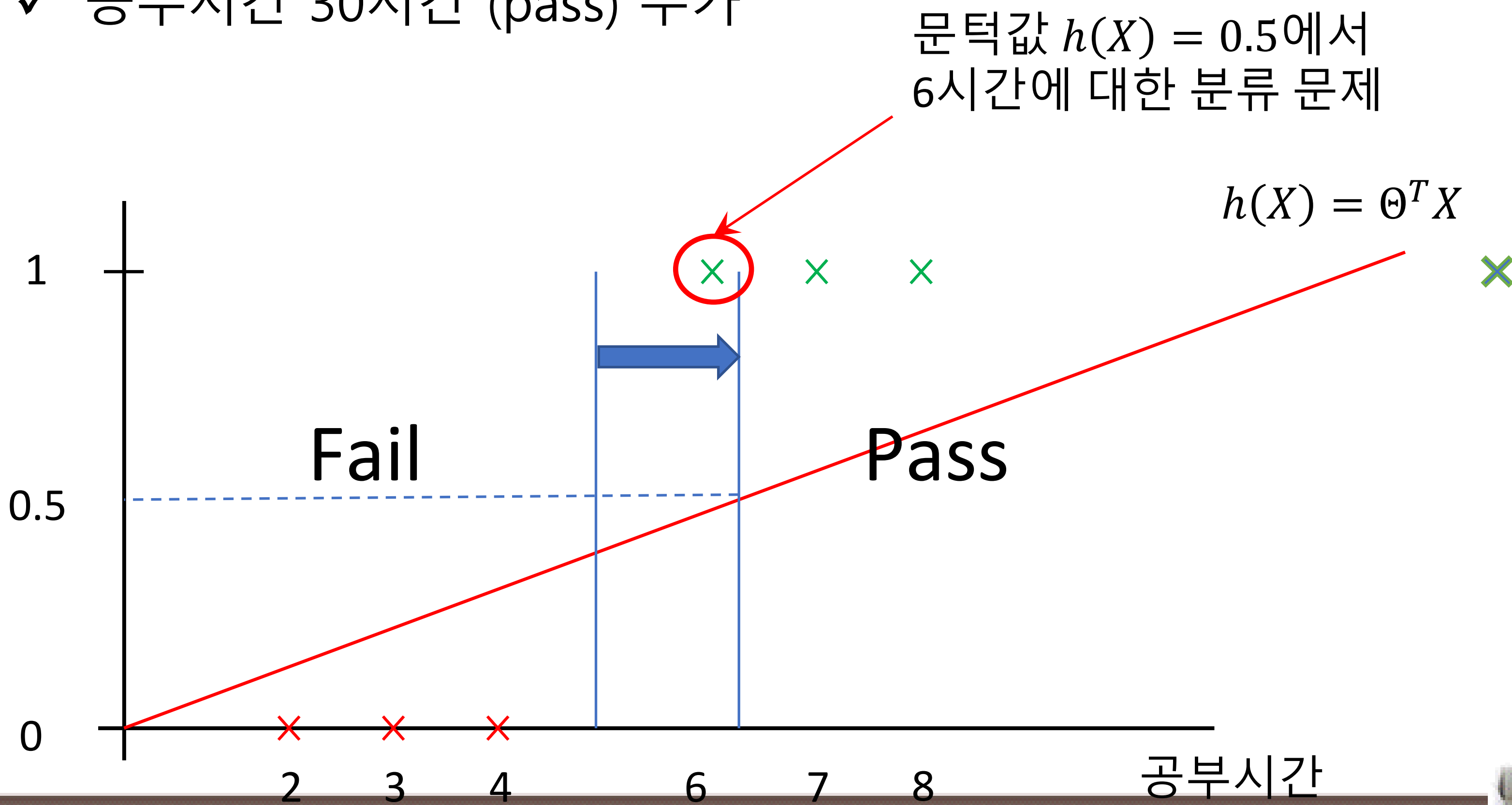


1. 로지스틱회귀

❖ 선형회귀를 이용한 분류의 한계

➢ 학습자료에 따른 문제점

✓ 공부시간 30시간 (pass) 추가



1. 로지스틱회귀

❖ 선형회귀를 이용한 분류의 한계

- 이진 분류의 경우
  - ✓ $y \in \{0,1\}$
- 실제 회귀 가설 $h(X) = w^T X$ 의 값
  - ✓ 1보다 클 수 있음
  - ✓ 0보다 작을 수 있음
- 따라서 아래의 조건을 만족하는 가설 함수 필요

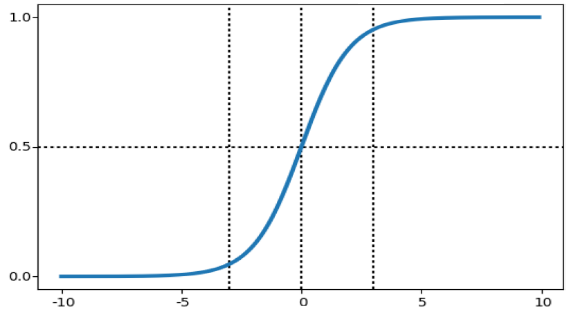
$$h(X) = w^T X$$

$$0 \le h(X) \le 1$$

1. 로지스틱회귀

❖ 로지스틱 함수

- ➢ 로지스틱 가설 표현
  - ✓ $y \in \{0,1\}$ 에 대해 $0 \leq h(X) \leq 1$
  - ✓ 로지스틱 회귀 모델
  - ✓ 분류문제에 적용
- ➢ 로지스틱 회귀 모델
  - ✓ 선형 가설표현을 로직스틱 함수를 적용하여 변환
  - ✓ 대표적인 로지스틱 함수 = sigmoid 함수



Sigmoid 함수

$$\sigma(z) = \frac{1}{1+e^{-z}}$$

$$z = w^T X$$

1. 로지스틱회귀

❖ 로지스틱 가설 표현의 의미

- $h(X)$: 입력 변수 X에 대해 결과가 1이 될 확률
- 예를 들어
  - ✓ 공부시간에 대한 pass 또는 fail의 분류에 있어
  - ✓ $h(X) = \sigma(w^T X) = 0.7$
  - ✓ $h(X) > 0.5$ 이므로 pass 예측하는 경우
  - ✓ 70%의 확률로 시험에 통과한다고 할 수 있음
- 로지스틱 가설표현은 출력이 1이 될 확률로 볼 수 있음
  - ✓ probability that y=1, given X, parameterized by w

$$h(X) = P(y = 1|X; w)$$

$$P(y = 0|X; w) = 1 - P(y = 1|X; w)$$

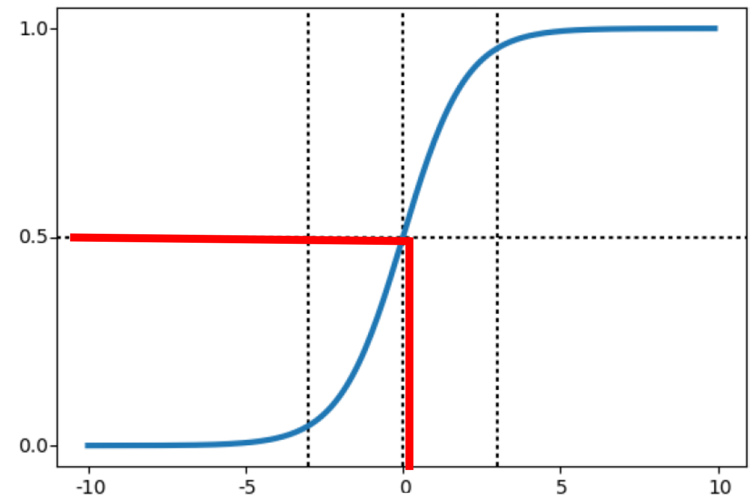
1. 로지스틱회귀

❖ Decision boundary

➢ 로지스틱 회귀 모델

$$h(X) = \sigma(w^T X)$$

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$



➢ 가정

✓ $h(X) \geq 0.5$ 이면 $y=1$ 로 분류

✓ $h(X) < 0.5$ 이면 $y=0$ 로 분류

➢ $h(X) \geq 0.5$ 조건

✓ Sigmoid 함수가 0.5 이상인 경우 : $w^T X \geq 0$

➢ $h(X) < 0.5$ 조건

✓ Sigmoid 함수가 0.5 미만인 경우 $w^T X < 0$

JEONJU UNIVERSITY

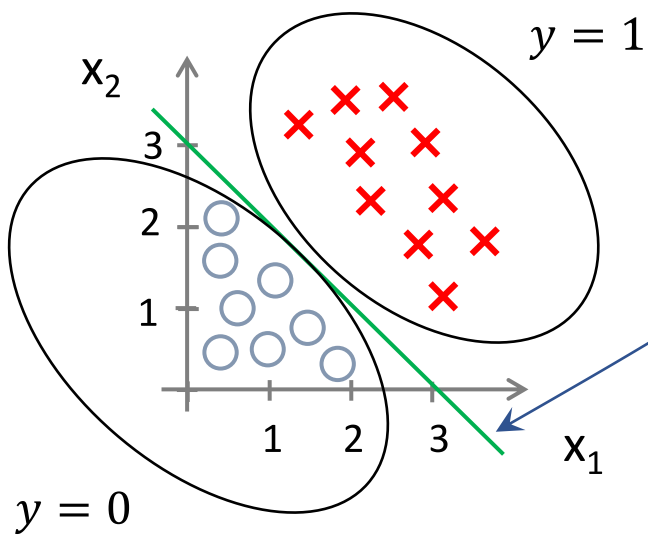
1. 로지스틱회귀

❖ Decision boundary

➢ 로지스틱 회귀 모델

✓ $w^T X \geq 0$ 이면 $y = 1$ 예측

✓ $w^T X < 0$ 이면 $y = 0$ 예측



$$h(x) = \sigma(w_0 + w_1 x_1 + w_2 x_2)$$
$$= \sigma(-3 + x_1 + x_2)$$

$x_1 + x_2 \geq 3$ 이면 $y = 1$ 예측
$x_1 + x_2 < 3$ 이면 $y = 0$ 예측

1. 로지스틱회귀

❖ 비용함수

- 학습자료 셋 s 에 대해 $S = \{(x^1, y^1), (x^2, y^2), \dots, (x^m, y^m)\}$

- n개의 입력 특징변수가 존재하는 로지스틱 회귀모델은

$$x^i = \begin{bmatrix} x_0^i \\ x_1^i \\ \cdots \\ x_n^i \end{bmatrix} \quad x_0^i = 1, y \in \{0,1\} \quad h(x^i) = \frac{1}{1 + e^{-w^T x^i}}$$

- 로지스틱 회귀 모델에서 회귀변수의 학습은?
  - ✓ 비용함수를 결정하고 경사하강 알고리즘을 이용
  - ✓ 비용함수의 최소값에 수렴하는 회귀변수 결정

JEONJU UNIVERSITY

1. 로지스틱회귀

❖ 비용함수

➢ 선형회귀모델에 사용한 비용함수 적용

$$J(w) = \frac{1}{m}\sum_{i=1}^{m}\boxed{\frac{1}{2}\left(h\left(x^{i}\right) - y^{i}\right)^{2}}$$

$$cost(h(x), y) = \frac{1}{2}(h(x) - y)^2 = \frac{1}{2}\left(\frac{1}{1 + e^{-w^T x}} - y\right)^2$$

➢ 로지스틱 회귀에서의 sigmoid 함수

✓ 선형회귀모델의 비용함수 적용하는 경우 non-convex 형태

✓ Local minima 로 수렴하는 문제 발생의 가능성이 높음

✓ 로지스틱 회귀모델을 위한 새로운 비용함수 설계 필요

JEONJU UNIVERSITY

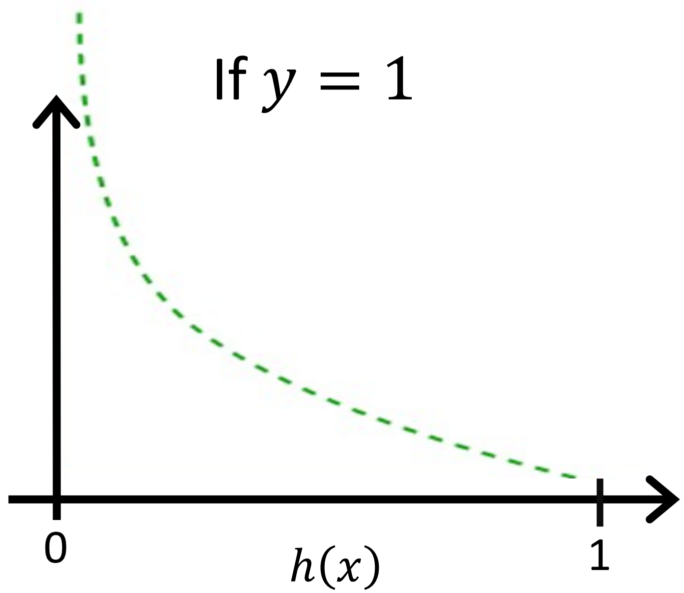
1. 로지스틱회귀

❖ 비용함수 설계

- 비용함수의 의미
  - ✓ 가설로부터 얻은 예측값이 정답에 얼마나 가까운가를 알려주는 척도
- 설계 방법
  - ✓ 예측값이 정답에 가까울수록 비용함수의 값을 작아지고
  - ✓ 예측값이 정답에 멀어질수록 비용함수의 값은 커짐
- 설계 예
  - ✓ $y=1$ 일때 $\hbar(x)=1$ 이면 $cost(\hbar(x),y)=0$
  - ✓ $y=1$ 일때 $\hbar(x)=0$ 이면
    - $cost(\hbar(x),y)=\infty$
    - $P(y=1|x;\mathrm{w})=0$ 을 의미하므로 학습 알고리즘에 페널티 적용

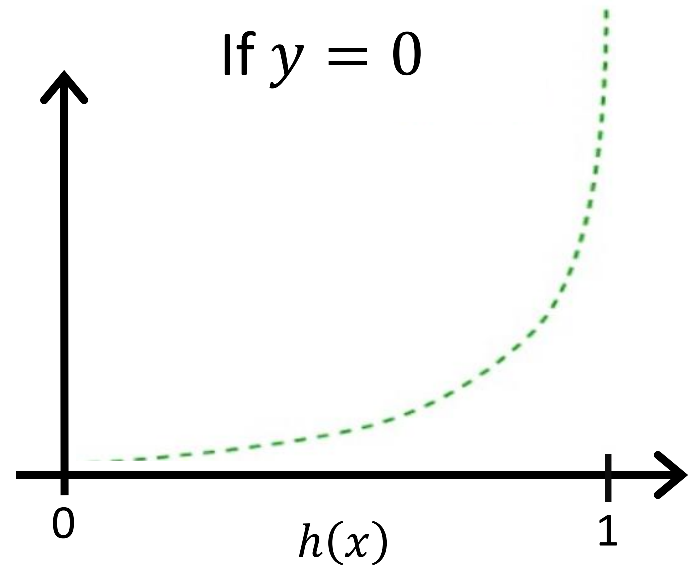
1. 로지스틱회귀

❖ 비용함수 설계



$$J(w) = \frac{1}{m}\sum_{i=1}^{m} cost\left(h\left(x^i\right), y^i\right)$$

$$cost\left(h\left(x^i\right), y^i\right) = -\log\left(h\left(x^i\right)\right)$$



$$J(w) = \frac{1}{m}\sum_{i=1}^{m} cost\left(h\left(x^i\right), y^i\right)$$

$$cost\left(h\left(x^i\right), y^i\right) = -\log\left(1 - h\left(x^i\right)\right)$$

1. 로지스틱회귀

❖ 비용함수 설계

$$J(w) = \frac{1}{m}\sum_{i=1}^{m} cost\left(h\left(x^{i}\right), y^{i}\right)$$

$$cost\left(h\left(x^{i}\right), y^{i}\right) = \begin{cases} -\log\left(h\left(x^{i}\right)\right) & if\ y = 1 \\ -\log\left(1 - h\left(x^{i}\right)\right) & if\ y = 0 \end{cases}$$

➢ 하나의 식으로 표현

$$cost(h(x), y) = -y\log\left(h(x)\right) - (1 - y)\log(1 - h(x))$$

1. 로지스틱회귀

❖ 로지스틱 회귀모델의 비용함수

$$J(w) = \frac{1}{m}\sum_{i=1}^{m} cost\left(h(x^i), y^i\right)$$

$$= -\frac{1}{m}\sum_{i=1}^{m}\left[y^i \log\left(h(x^i)\right) + (1 - y^i)\log(1 - h(x^i))\right]$$

❖회귀변수 w 학습

$$\hat{w} \leftarrow \min_{w} J(w)$$

❖입력에 대한 결과 예측

$$h(x) = \frac{1}{1 + e^{-\hat{w}^T x}}$$ → If $h(x) \geq 0.5, y = 1$

1. 로지스틱회귀

❖ 경사하강법을 이용한 학습

$$J(w) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^i \log\left(h(x^i)\right) + (1-y^i)\log(1-h(x^i))\right]$$

➢ 비용함수의 미분값 이용

$$w_j = w_j - \alpha \frac{\partial}{\partial w_j} J(w)$$

$\min_{w} J(w)$를 얻기 위해서

Repeat {

$$w_j = w_j - \alpha \sum_{i=1}^{m} \left(h(x^i) - y^i\right) x_j^i$$

} until convergence

- 선형회귀모델과 동일한 형태
- 가설 표현 함수 $h(x^i)$ 만 차이

1. 로지스틱회귀

❖ 경사하강법을 이용한 학습

$\frac{\partial Loss}{\partial \hat{\mathbf{w}}}$를 체인롤(Chain Rule)에 의하여 분해 $(z = \mathbf{x}^T\hat{\mathbf{w}})$

$$\frac{\partial Loss}{\partial \hat{\mathbf{w}}} = \frac{\partial Loss}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial z} \frac{\partial z}{\partial \hat{\mathbf{w}}}$$

$$\frac{\partial Loss}{\partial \hat{y}} = \frac{\partial}{\partial \hat{y}}(-y\log(\hat{y}) - (1-y)\log(1-\hat{y})) = -\frac{y}{\hat{y}} + \frac{1-y}{1-\hat{y}}$$

$$\frac{\partial \hat{y}}{\partial z} = \frac{\partial}{\partial z}\left(\frac{1}{1+e^{-z}}\right) = \frac{e^{-z}}{(1+e^{-z})^2} = \frac{1+e^{-z}-1}{(1+e^{-z})^2} = \frac{1}{1+e^{-z}}\left(1 - \frac{1}{1+e^{-z}}\right) = \hat{y}(1-\hat{y})$$

$$\frac{\partial z}{\partial \hat{\mathbf{w}}} = \frac{\partial}{\partial \hat{\mathbf{w}}}(\mathbf{x}^T\hat{\mathbf{w}}) = \mathbf{x}^T$$

$$\frac{\partial Loss}{\partial \hat{\mathbf{w}}} = \left(-\frac{y}{\hat{y}} + \frac{1-y}{1-\hat{y}}\right) * \hat{y}(1-\hat{y}) * \mathbf{x}^T$$

Thank You !

JEONJU UNIVERSITY
The Place for Superstars