Logistic Regression

POSTECH A.I

Seungbeom Lee

The slides is made based on moonjeong's note

Table of Contents

- Logistic Regression
  - Review the key points
  - Practice

- Multiclass Logistic Regression
  - Review the key points
  - Practice

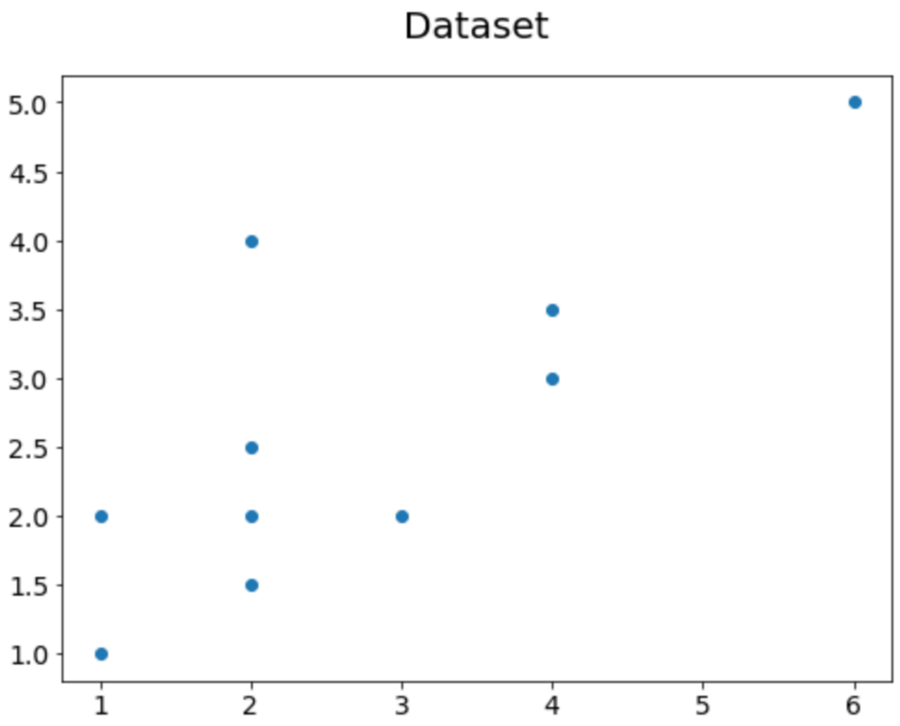
Table of Contents

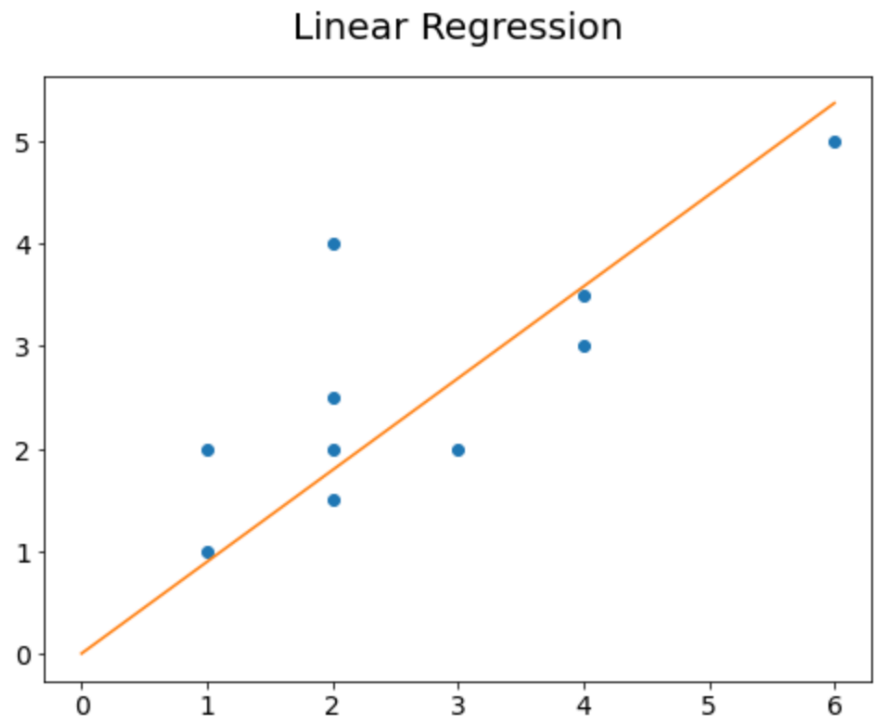
- Logistic Regression
  - Review the key points
  - Practice

- Multiclass Logistic Regression
  - Review the key points
  - Practice

Review the key points

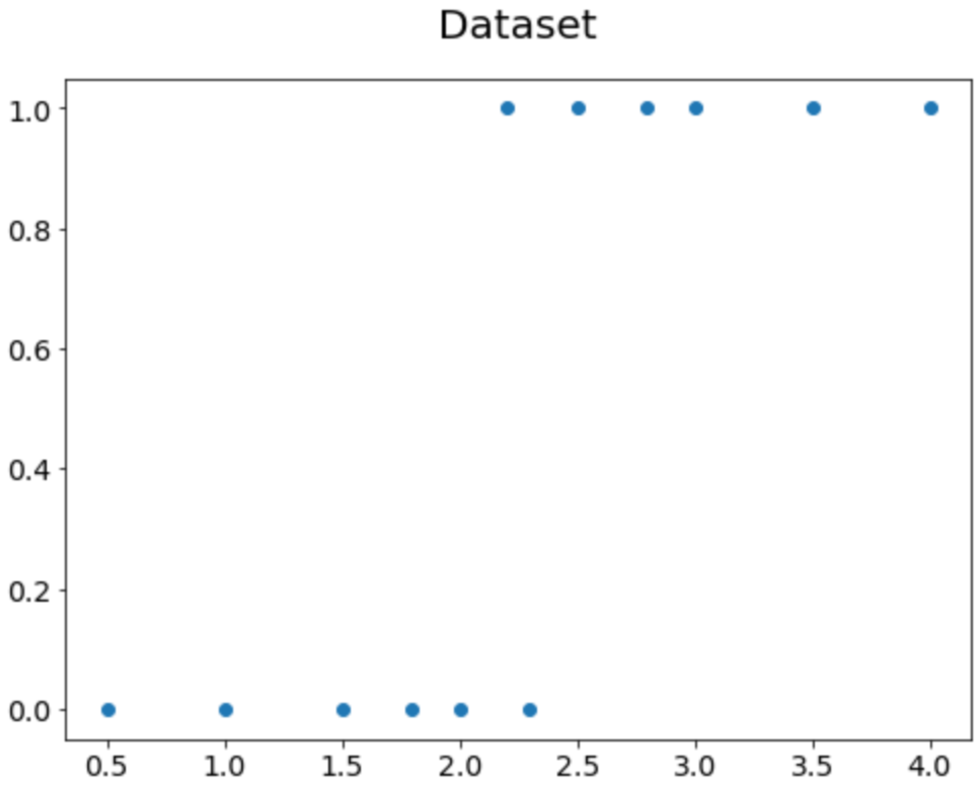
- Linear Regression: 연속적인 값을 근사할 때 효과적임





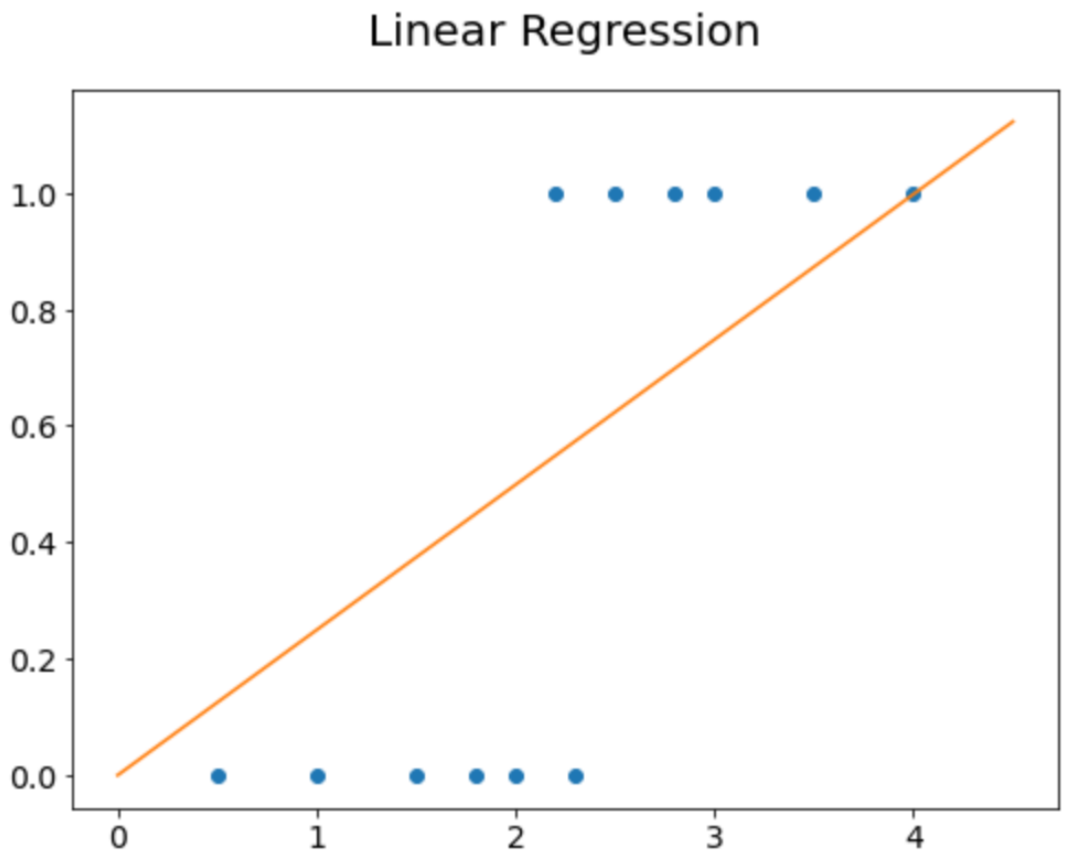
Review the key points

- y가 불연속적인 값 (ex. 합격(=1), 불합격(=0))을 가지는 데이터의 경우?
- 즉 분류 문제의 경우?



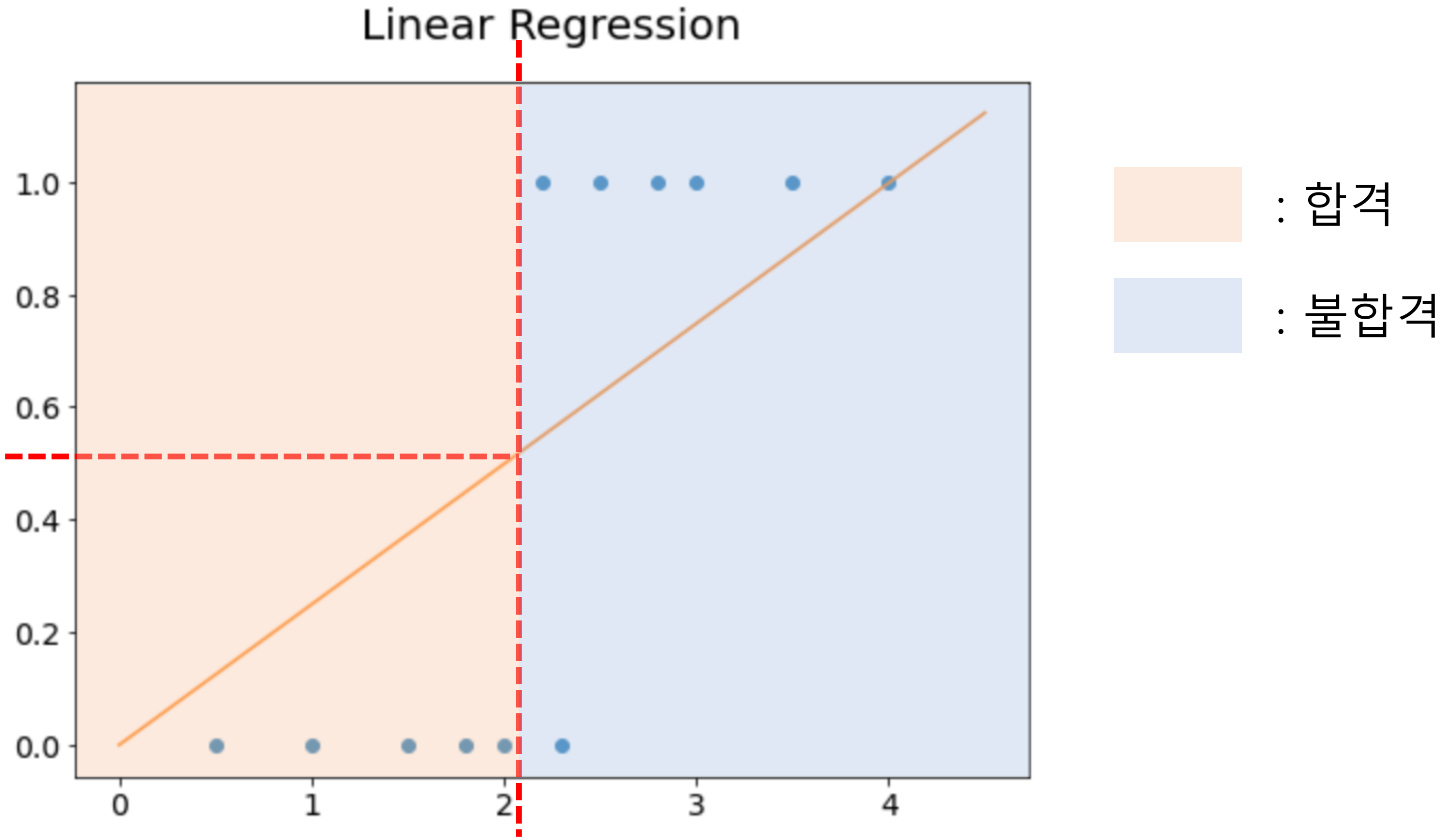
Review the key points

- Output값을 y=1일 확률로 봄



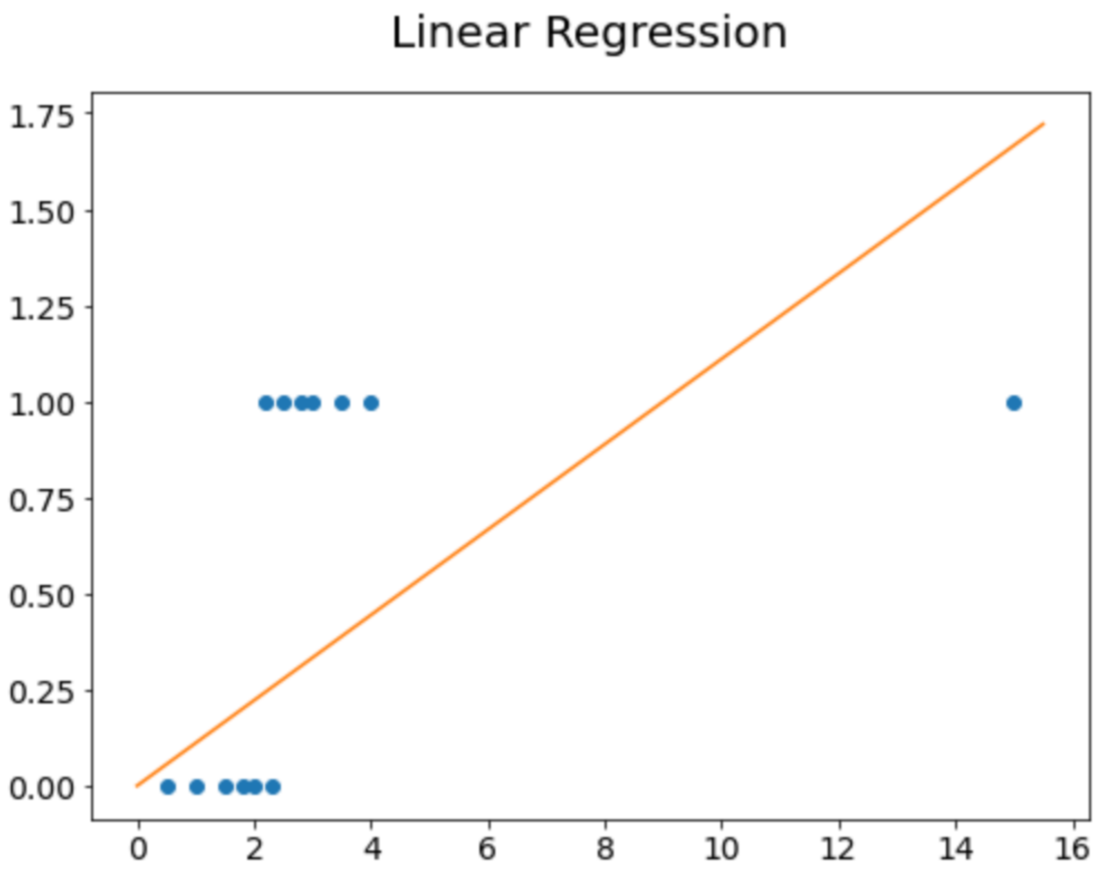
Review the key points

- Output값을 y=1일 확률로 봄 → 0.5 지점을 기준으로 구분



Review the key points

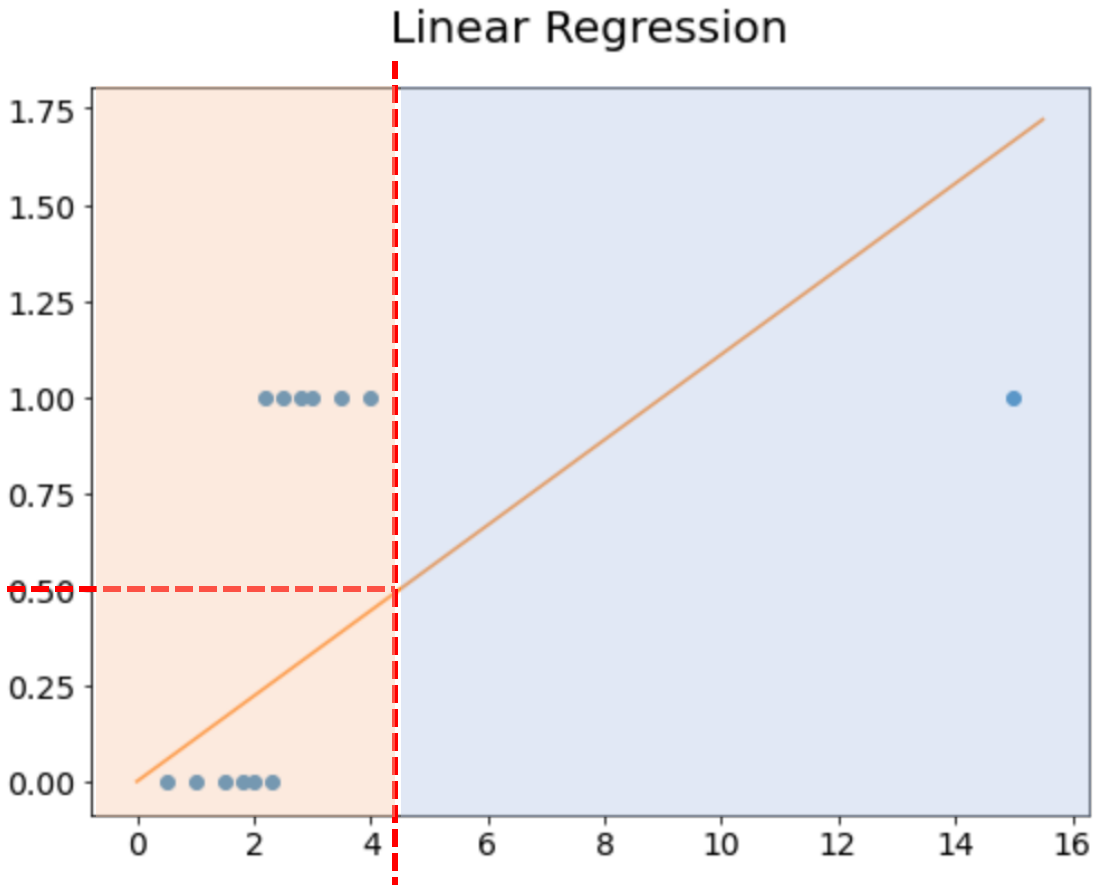
- 하지만 linear regression은 outlier에 망가지기 쉬움



POSTECH

Review the key points

- 하지만 linear regression은 outlier에 망가지기 쉬움

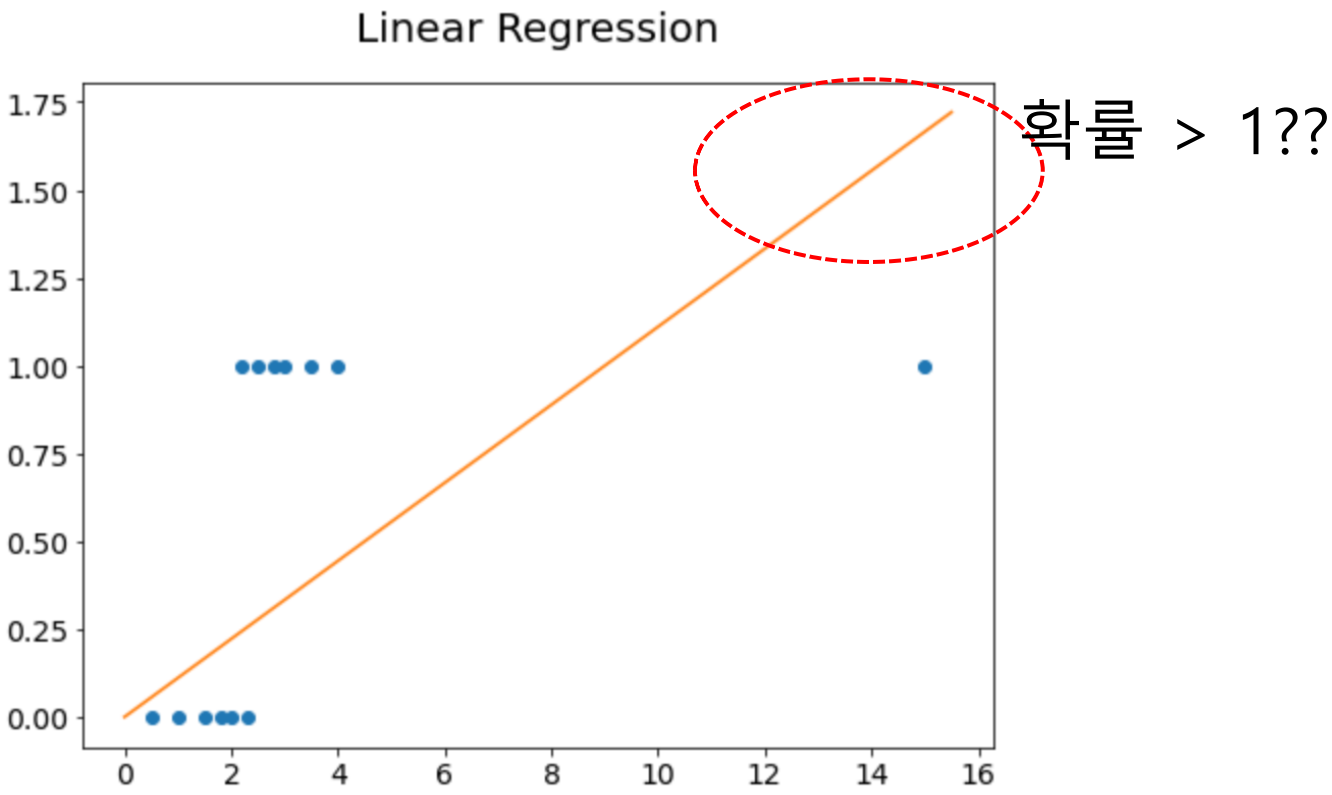


: 합격

: 불합격

Review the key points

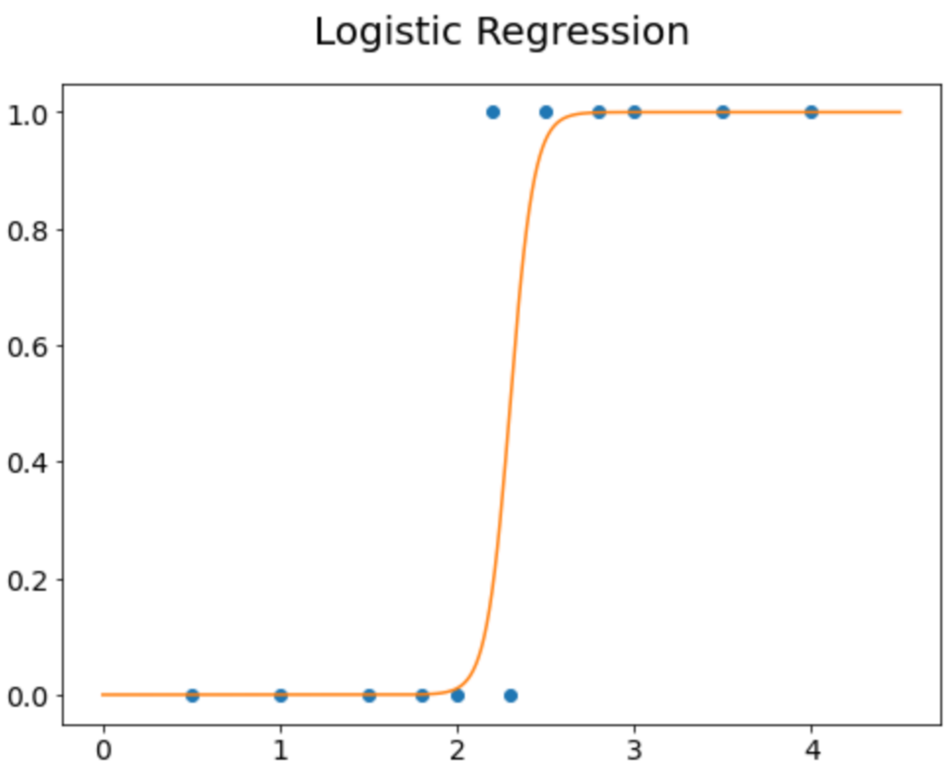
- 또한, output이 0보다 작거나 1보다 큰 경우, 확률로 모델링하기에 적절하지 않음.



Review the key points

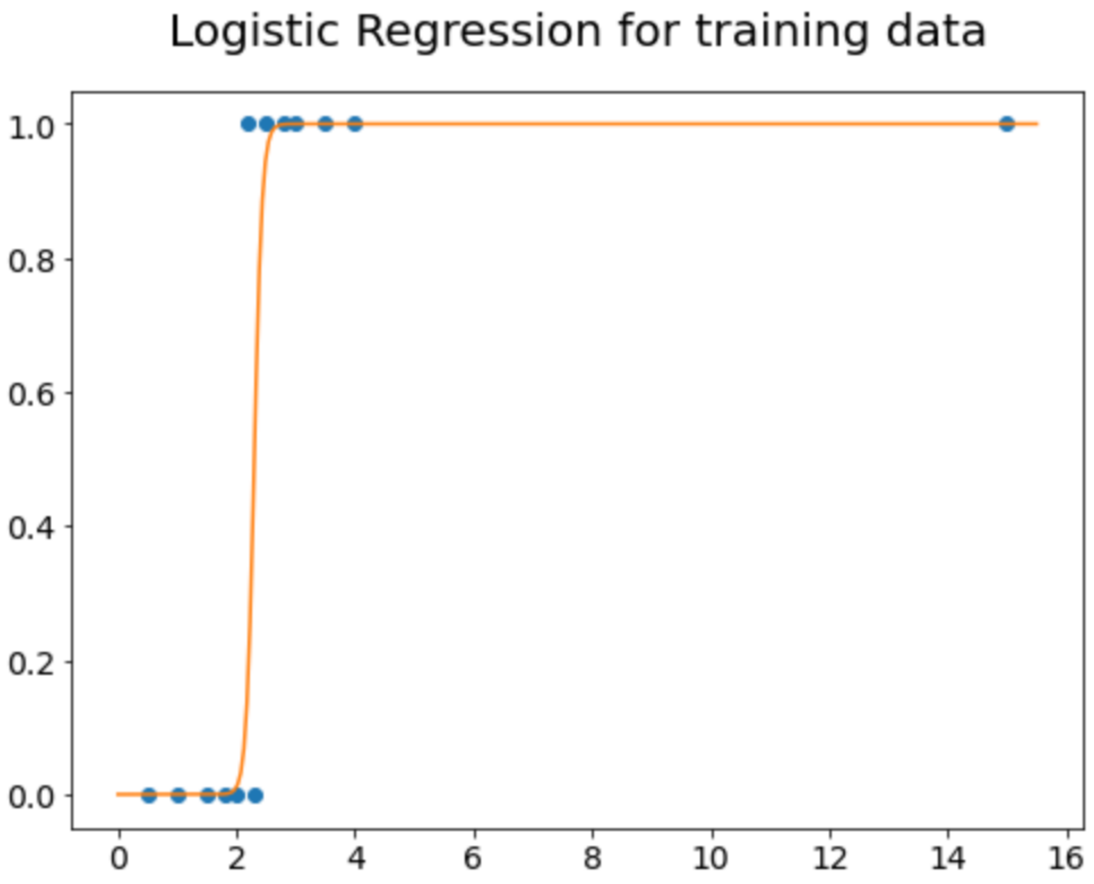
- Logistic regression
- Output값이 0 ~ 1 사이의 값을 가질 것
- Outlier에 덜 sensitive 할 것
- Output값이 0.5에서 급격히 변할 것
- Logistic function

$$\text{where } \sigma(t) = \frac{1}{1+e^{-t}} = \frac{e^t}{1+e^t}$$



Review the key points

- Outlier에도 robust한 편



Review the key points

- 풀고 싶은 문제
  - “다음과 같이 주어진 Dataset에서, $\boldsymbol{x}$와 y 사이에는 어떤 관계가 있을까?”

$$\text{Dataset} : \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \cdots, (\boldsymbol{x}_N, y_N)\}$$

$$(\boldsymbol{x} \in \mathbb{R}^d, \ y \in \{0, 1\})$$

Review the key points

$$P(y_n = 1|\mathbf{x}_n) = \sigma(\mathbf{w}^\top \mathbf{x}_n), \quad \text{where } \sigma(t) = \frac{1}{1+e^{-t}} = \frac{e^t}{1+e^t}$$

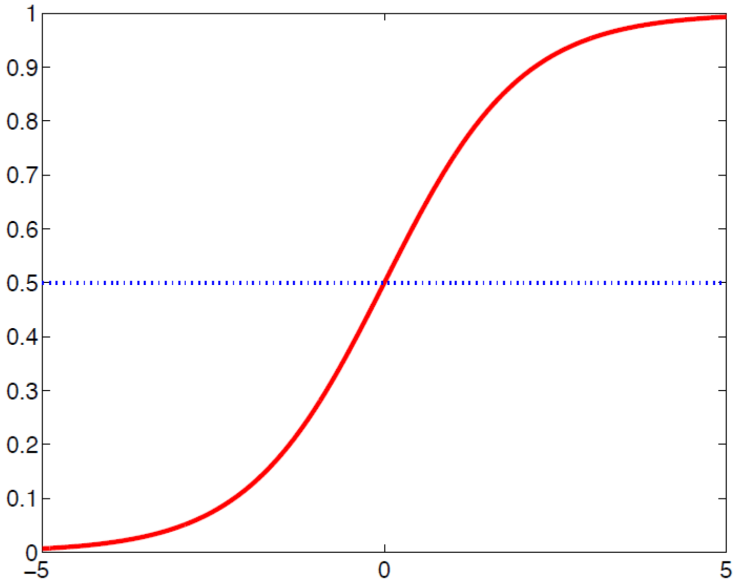


Figure: Logistic function $\sigma(t) = \frac{1}{1+e^{-t}}$

- $\sigma(t) \to 0$ as $t \to -\infty$
- $\sigma(t) \to 1$ as $t \to \infty$
- $\sigma(-t) = 1 - \sigma(t)$
- $\frac{d}{dt}\sigma(t) = \sigma(t)\sigma(-t) = \sigma(t)(1 - \sigma(t))$

Review the key points

- MLE

Given $\{(\mathbf{x}_n, y_n) | n = 1, \dots, N\}$, the likelihood is given by

$$p(\mathbf{y}|\mathbf{X}, \mathbf{w}) = \prod_{n=1}^{N} p(y_n = 1|\mathbf{x}_n)^{y_n}(1 - p(y_n = 1|\mathbf{x}_n))^{1-y_n}$$

$$\boxed{y_n \in \{0, 1\}} \qquad = \prod_{n=1}^{N} \sigma(\mathbf{w}^\top \mathbf{x}_n)^{y_n}(1 - \sigma(\mathbf{w}^\top \mathbf{x}_n))^{1-y_n}$$

POSTECH

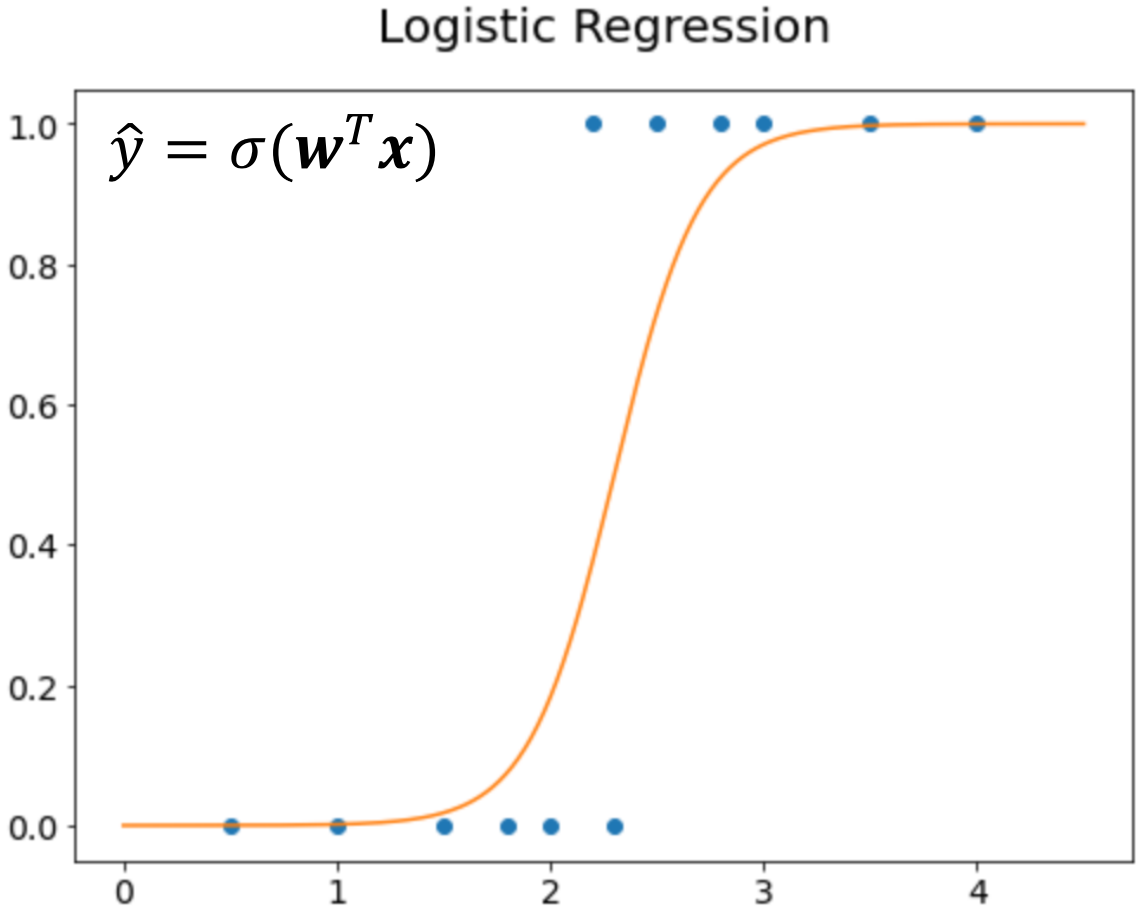
Review the key points

Given $\{(\mathbf{x}_n, y_n) | n = 1, \ldots, N\}$, the likelihood is given by

$$p(\mathbf{y}|\mathbf{X}, \mathbf{w}) = \prod_{n=1}^{N} p(y_n = 1|\mathbf{x}_n)^{y_n}(1 - p(y_n = 1|\mathbf{x}_n))^{1-y_n}$$

$$= \prod_{n=1}^{N} \sigma(\mathbf{w}^\top \mathbf{x}_n)^{y_n}(1 - \sigma(\mathbf{w}^\top \mathbf{x}_n))^{1-y_n}$$

- MLE



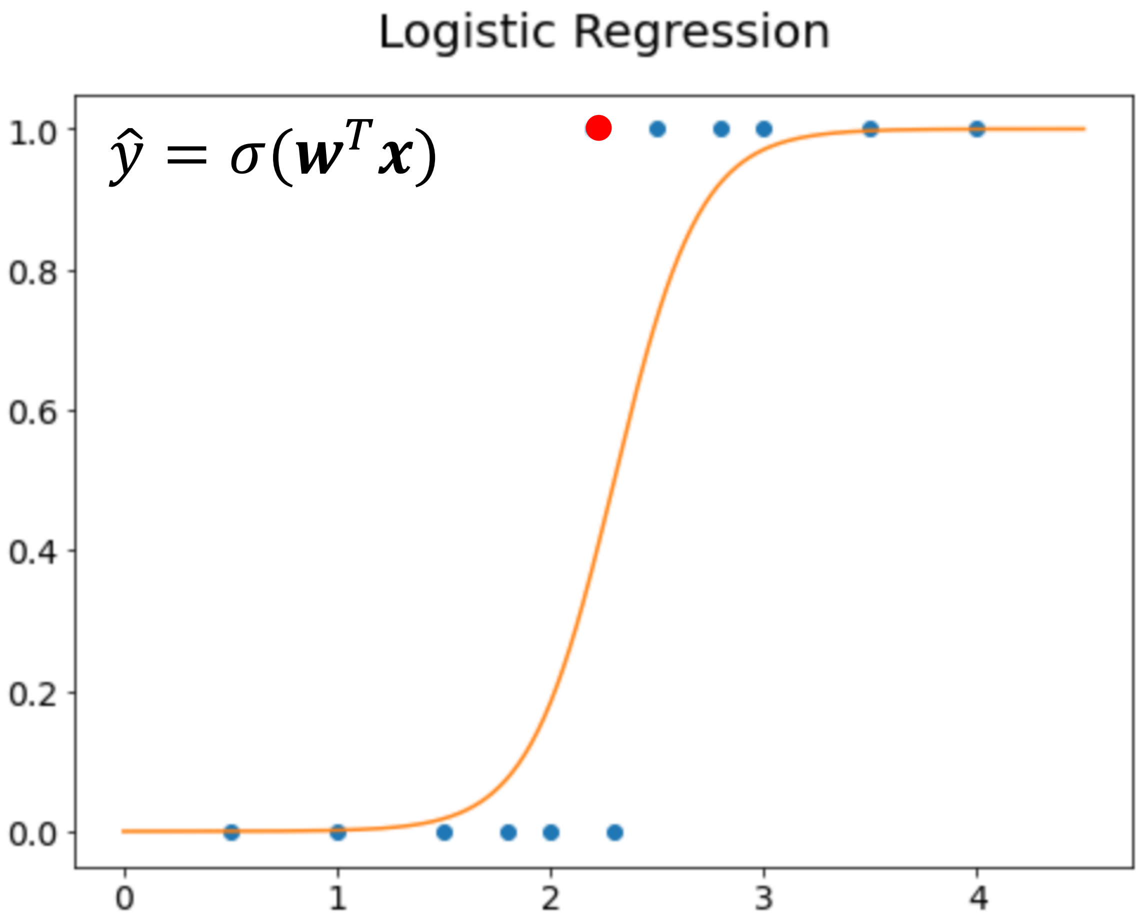
Review the key points

Given $\{(\mathbf{x}_n, y_n) | n = 1, \ldots, N\}$, the likelihood is given by

$$p(\mathbf{y}|\mathbf{X}, \mathbf{w}) = \prod_{n=1}^{N} p(y_n = 1|\mathbf{x}_n)^{y_n}(1 - p(y_n = 1|\mathbf{x}_n))^{1-y_n}$$

$$= \prod_{n=1}^{N} \sigma(\mathbf{w}^\top \mathbf{x}_n)^{y_n}(1 - \sigma(\mathbf{w}^\top \mathbf{x}_n))^{1-y_n}$$

- MLE

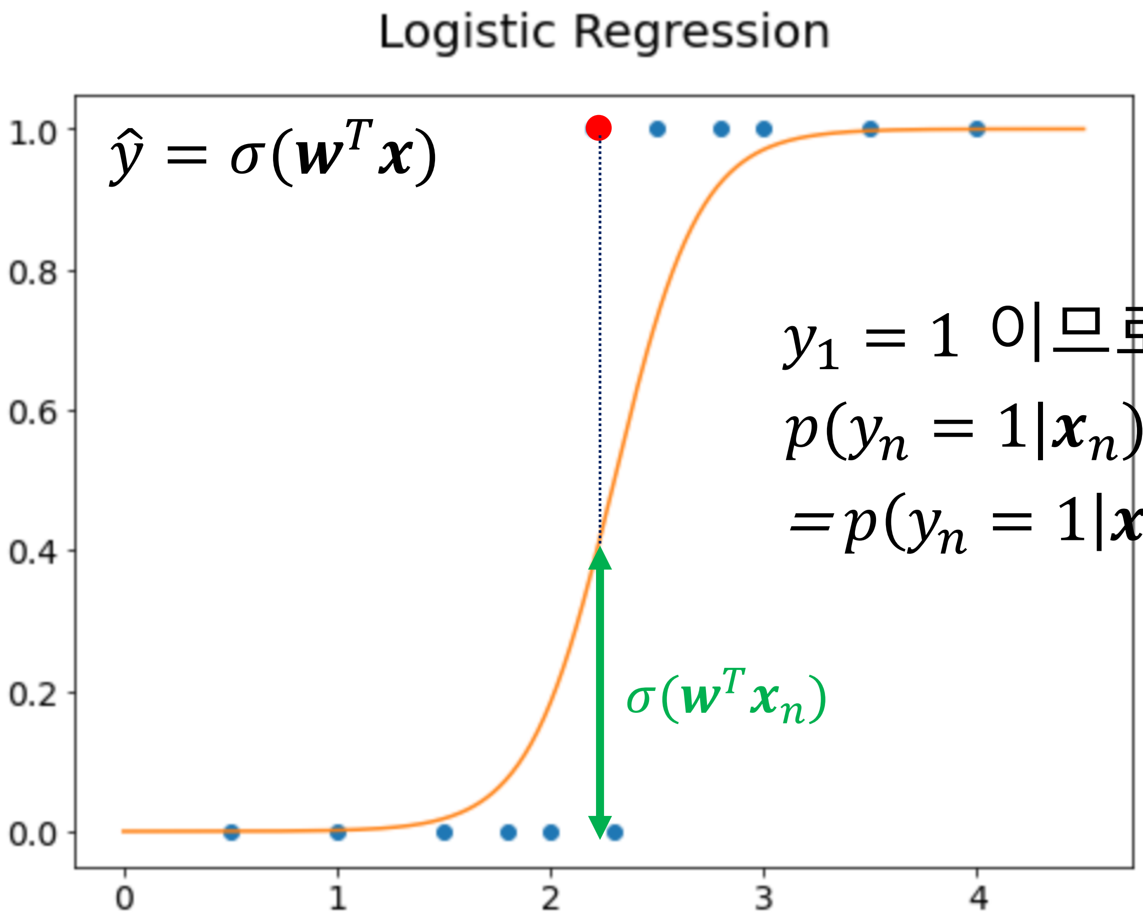


Review the key points

- MLE

Given $\{(\mathbf{x}_n, y_n) | n = 1, \ldots, N\}$, the likelihood is given by

$$
\begin{aligned}
p(\mathbf{y}|\mathbf{X}, \mathbf{w}) &= \prod_{n=1}^{N} p(y_n = 1|\mathbf{x}_n)^{y_n}(1 - p(y_n = 1|\mathbf{x}_n))^{1-y_n} \\
&= \prod_{n=1}^{N} \sigma(\mathbf{w}^\top \mathbf{x}_n)^{y_n}(1 - \sigma(\mathbf{w}^\top \mathbf{x}_n))^{1-y_n}
\end{aligned}
$$



$y_1 = 1$ 이므로

$p(y_n = 1|\boldsymbol{x}_n)^{y_n}(1 - p(y_n = 1|\boldsymbol{x}_n))^{1-y_n}$

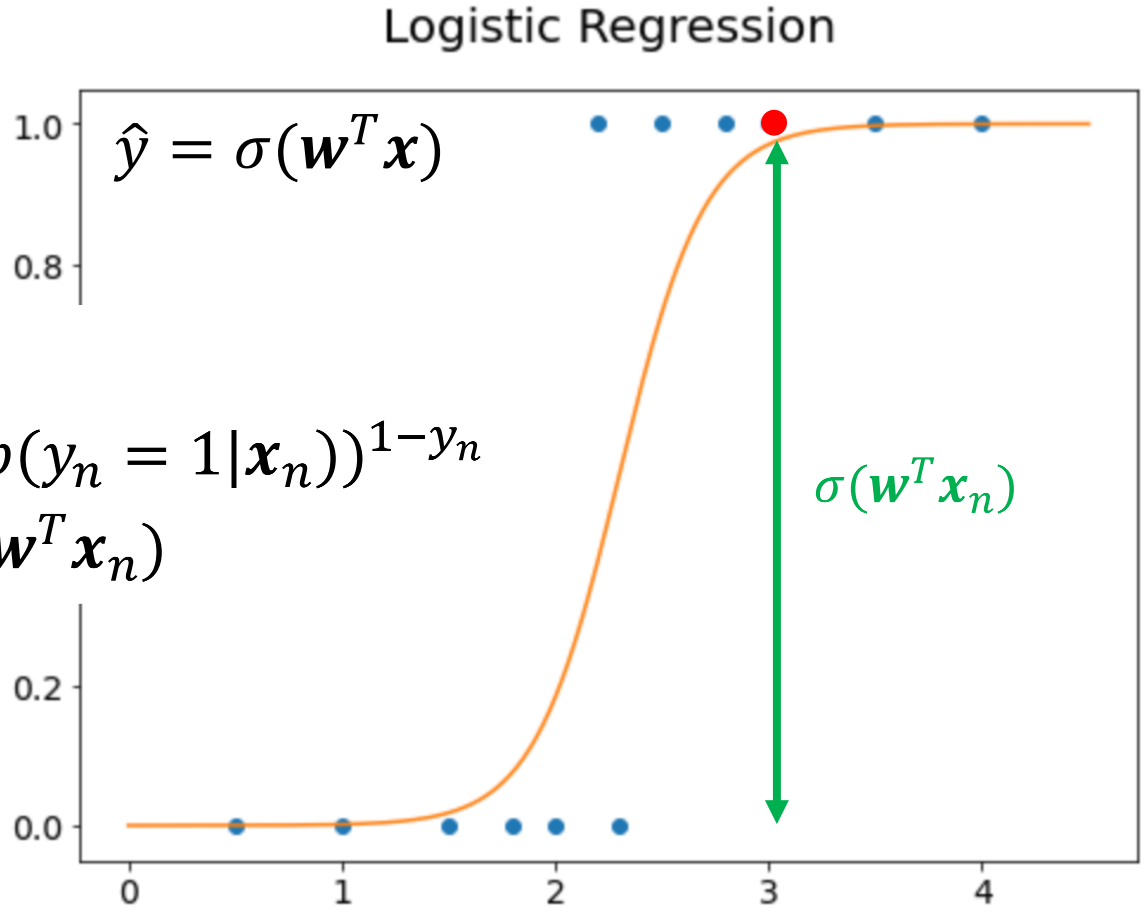
$= p(y_n = 1|\boldsymbol{x}_n) = \sigma(\boldsymbol{w}^T \boldsymbol{x}_n)$

Given $\{(\mathbf{x}_n, y_n) | n = 1, \ldots, N\}$, the likelihood is given by

$$p(\mathbf{y}|\mathbf{X}, \mathbf{w}) = \prod_{n=1}^{N} p(y_n = 1|\mathbf{x}_n)^{y_n}(1 - p(y_n = 1|\mathbf{x}_n))^{1-y_n}$$
$$= \prod_{n=1}^{N} \sigma(\mathbf{w}^\top \mathbf{x}_n)^{y_n}(1 - \sigma(\mathbf{w}^\top \mathbf{x}_n))^{1-y_n}$$

Review the key points

- MLE



$y_2 = 1$ 이므로

$p(y_n = 1|\boldsymbol{x}_n)^{y_n}(1 - p(y_n = 1|\boldsymbol{x}_n))^{1-y_n}$

$= p(y_n = 1|\boldsymbol{x}_n) = \sigma(\boldsymbol{w}^T \boldsymbol{x}_n)$

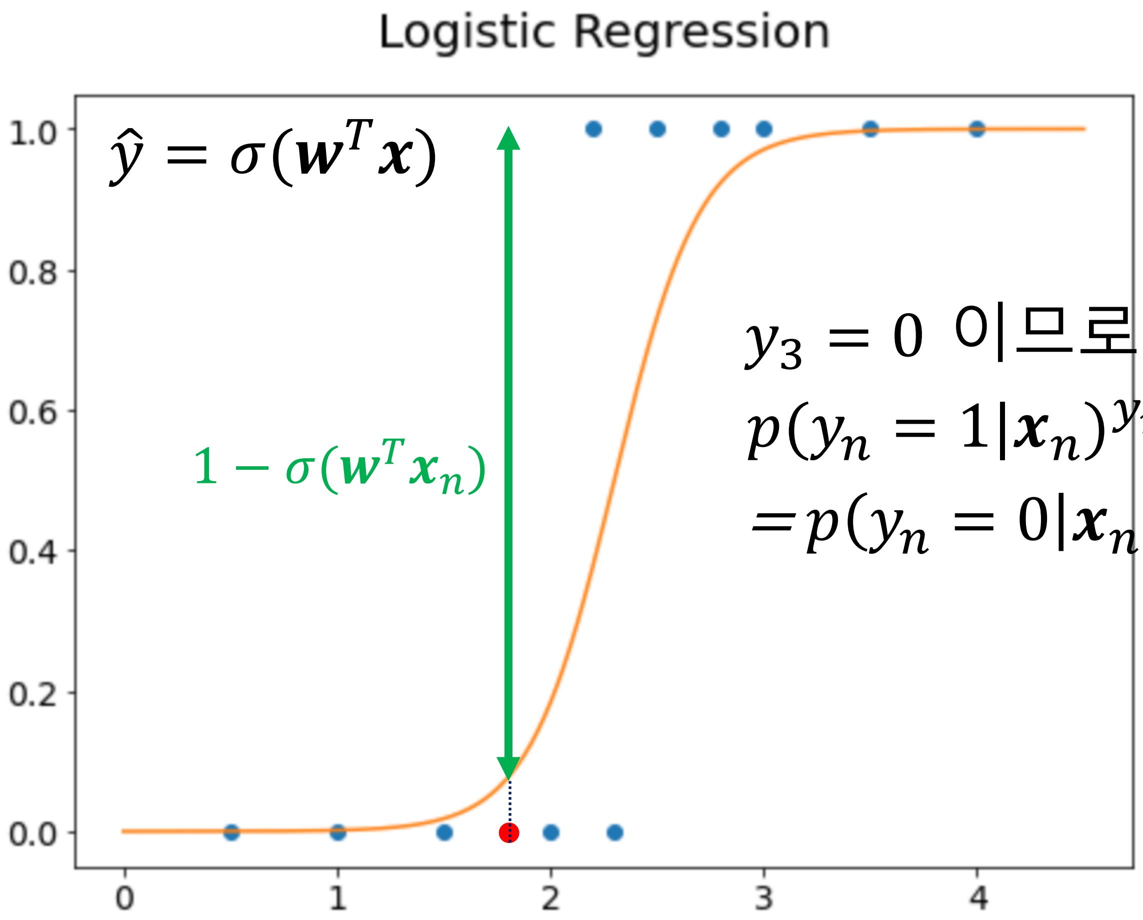
Review the key points

Given $\{(\mathbf{x}_n, y_n) | n = 1, \dots, N\}$, the likelihood is given by

$$p(\mathbf{y}|\mathbf{X}, \mathbf{w}) = \prod_{n=1}^{N} p(y_n = 1|\mathbf{x}_n)^{y_n}(1 - p(y_n = 1|\mathbf{x}_n))^{1-y_n}$$

$$= \prod_{n=1}^{N} \sigma(\mathbf{w}^\top \mathbf{x}_n)^{y_n}(1 - \sigma(\mathbf{w}^\top \mathbf{x}_n))^{1-y_n}$$

- MLE



$y_3 = 0$ 이므로

$p(y_n = 1|\boldsymbol{x}_n)^{y_n}(1 - p(y_n = 1|\boldsymbol{x}_n))^{1-y_n}$

$= p(y_n = 0|\boldsymbol{x}_n) = 1 - \sigma(\boldsymbol{w}^T \boldsymbol{x}_n)$

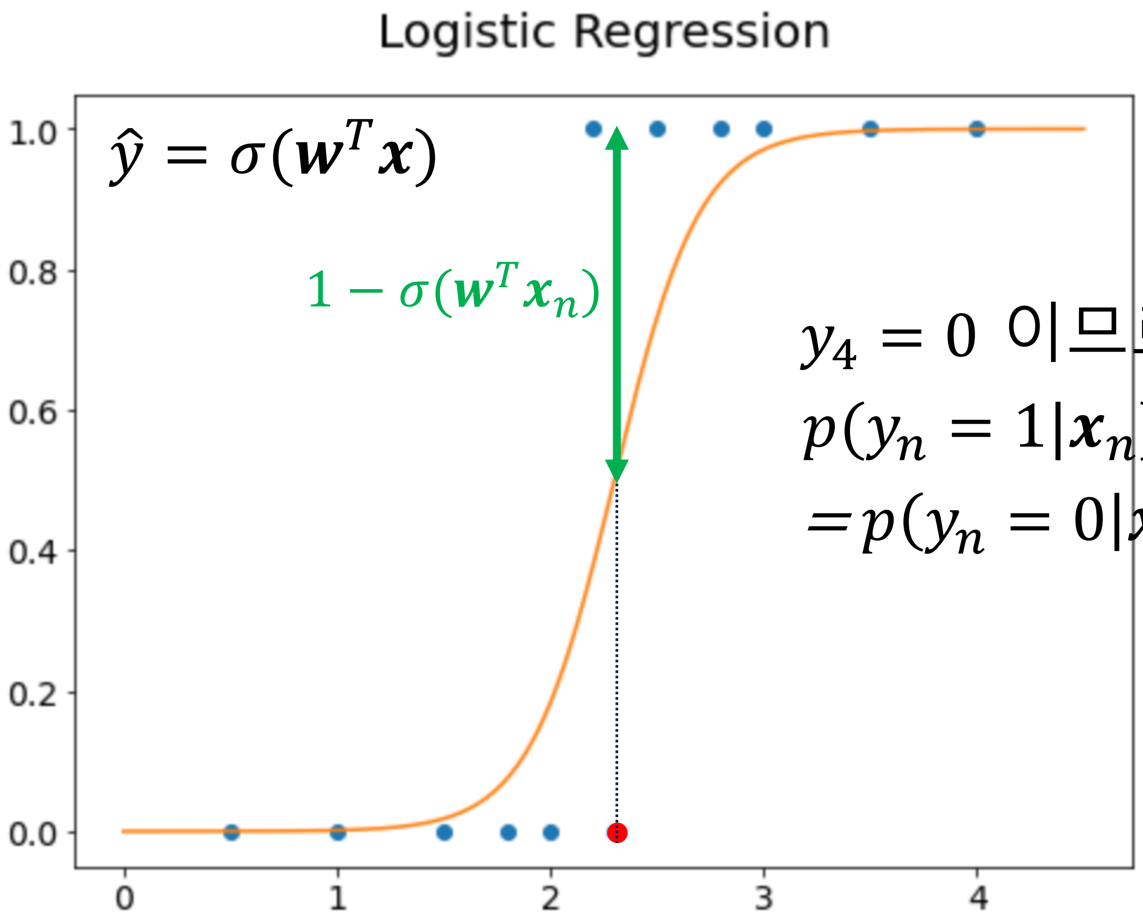
Given $\{(\mathbf{x}_n, y_n) | n = 1, \ldots, N\}$, the likelihood is given by

$$p(\mathbf{y}|\mathbf{X}, \mathbf{w}) = \prod_{n=1}^{N} p(y_n = 1|\mathbf{x}_n)^{y_n}(1 - p(y_n = 1|\mathbf{x}_n))^{1-y_n}$$

$$= \prod_{n=1}^{N} \sigma(\mathbf{w}^\top \mathbf{x}_n)^{y_n}(1 - \sigma(\mathbf{w}^\top \mathbf{x}_n))^{1-y_n}$$

Review the key points

- MLE



$y_4 = 0$ 이므로

$$p(y_n = 1|\boldsymbol{x}_n)^{y_n}(1 - p(y_n = 1|\boldsymbol{x}_n))^{1-y_n}$$

$$= p(y_n = 0|\boldsymbol{x}_n) = 1 - \sigma(\boldsymbol{w}^T \boldsymbol{x}_n)$$

Review the key points

- MLE

$$p(y_n=1|x_n)^{y_n}(1-p(y_n=1|x_n))^{1-y_n}$$

$\Rightarrow$

if $y_n=1$) $p(y_n=1|x_n)(1-p(y_n=1|x_n))^{1-1}$

$= p(y_n=1|x_n)$

if $y_n=0$) $p(y_n=0|x_n)^0(1-p(y_n=1|x_n))^{1-0}$

$= 1-p(y_n=1|x_n)$

$= p(y_n=0|x_n)$

$$\boxed{y_n \in \{0, 1\}}$$

Given $\{(\mathbf{x}_n, y_n)|n=1,\ldots,N\}$, the likelihood is given by

$$\begin{aligned} p(\mathbf{y}|\mathbf{X},\mathbf{w}) &= \prod_{n=1}^{N} p(y_n=1|\mathbf{x}_n)^{y_n}(1-p(y_n=1|\mathbf{x}_n))^{1-y_n} \\ &= \prod_{n=1}^{N} \sigma(\mathbf{w}^\top\mathbf{x}_n)^{y_n}(1-\sigma(\mathbf{w}^\top\mathbf{x}_n))^{1-y_n} \end{aligned}$$

Review the key points

- MLE
  - Then log-likelihood function is given by

$$\mathcal{L} = \sum_{n=1}^{N} \log p(y_n|\mathbf{x}_n) = \sum_{n=1}^{N} \{y_n \log \hat{y}_n + (1 - y_n) \log(1 - \hat{y}_n)\},$$

    where $\hat{y}_n = \sigma(\mathbf{w}^\top \mathbf{x}_n)$.
  - This is a nonlinear function of $\mathbf{w}$ whose maximum cannot be computed in a closed form.

- Then log-likelihood function is given by

$$\mathcal{L} = \sum_{n=1}^{N} \log p(y_n|\mathbf{x}_n) = \sum_{n=1}^{N} \{y_n \log \hat{y}_n + (1-y_n)\log(1-\hat{y}_n)\},$$

where $\hat{y}_n = \sigma(\mathbf{w}^\top \mathbf{x}_n)$.

Review the key points

- MLE $\quad y_n \log \hat{y}_n + (1-y_n)\log(1-\hat{y}_n)$

- Loss Function - $\quad BCE = -\frac{1}{N}\sum_{i=0}^{N} y_i \cdot log(\hat{y}_i) + (1-y_i) \cdot log(1-\hat{y}_i)$

POSTECH

Table of Contents

- Logistic Regression
  - Review the key points
  - Practice

- Multiclass Logistic Regression
  - Review the key points
  - Practice

Practice

- Output $y \in \{0, 1\}$로 하는 synthetic dataset을 생성하고 이에 대한 Logistic Regression 구현하기

  - Loss function 설정 (Cross entropy)
  - Optimization 설정
  - 함수 정의
  - 유용한 tensor flow 활용

- Any question ?

Thank you

slee2020@postech.ac.kr