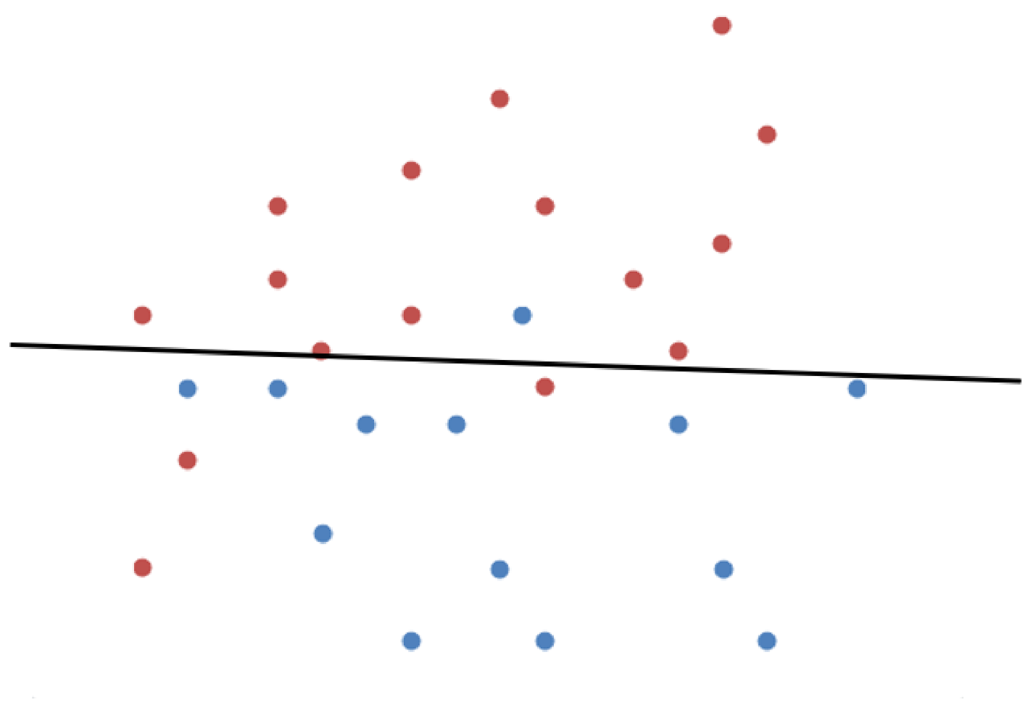
Logistic Regression

Linear Classifier

- Can we find a linear boundary between two classes?

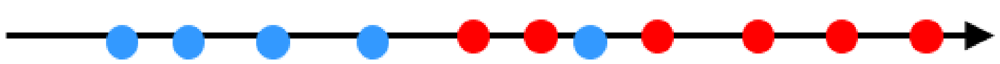
  maybe yes..



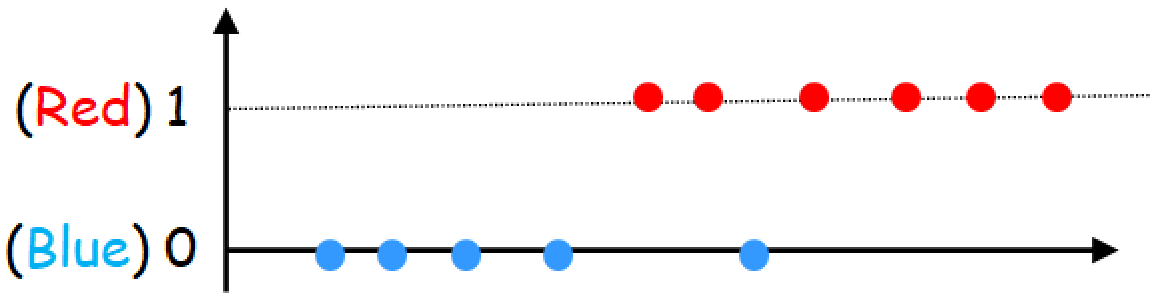
Let's Consider a Simple case

2-Dimensional Space

>>입력 1차원, 출력 1차원(red or blue>>1 or 0)

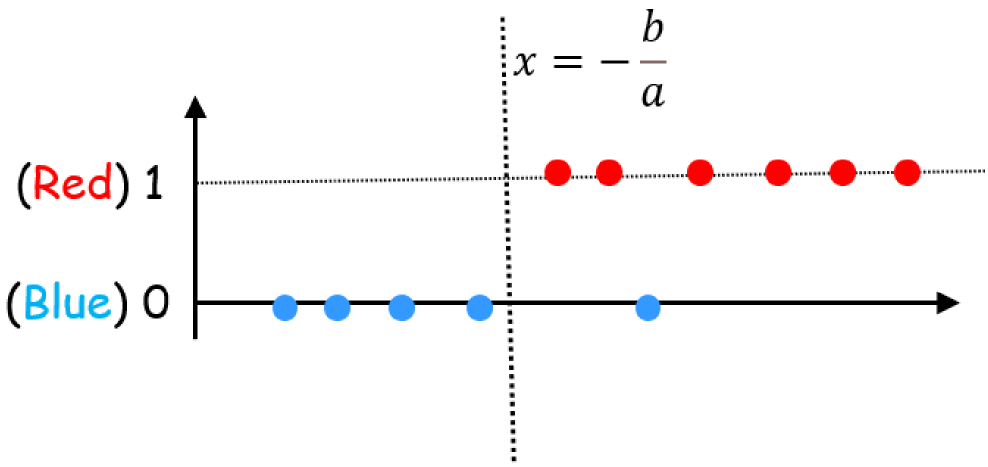


- Map Red to 1 and Blue to 0

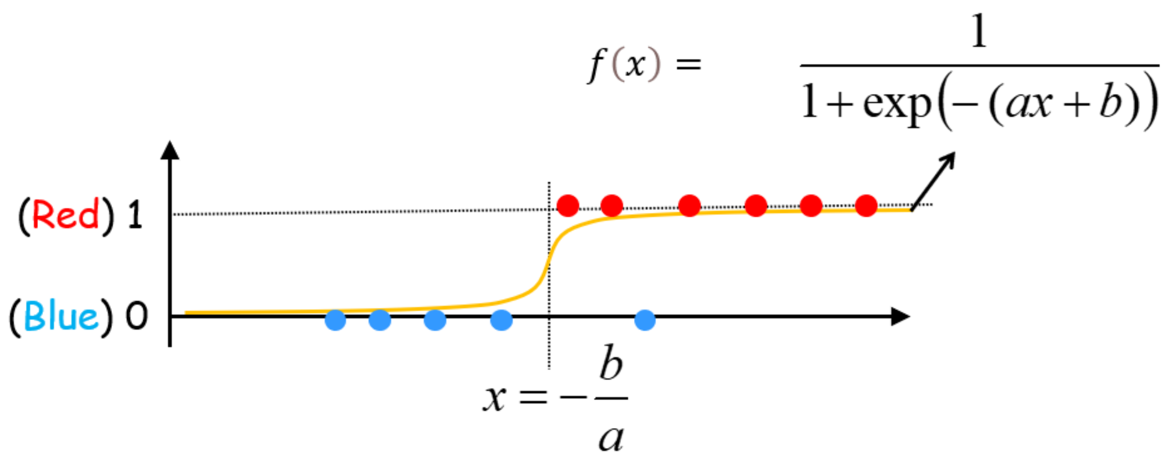


- Which model is good for this case?
    - A simple line?

- Find a simple line which best fits the data

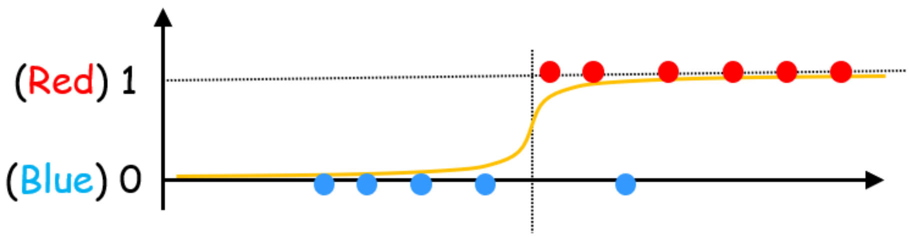


- Instead of a simple line, let's use a **logistic function(sigmoid function)**



input 값이 경계선 x=-b/a보다 오른쪽으로 멀 경우 f(x)는1, 왼쪽으로 멀 경우 f(x)는 0

∵sigmoid function(logistic function)을 사용했기 때문에



$$h(\mathbf{x}) = \frac{1}{1+\exp(-\mathbf{w}\cdot\mathbf{x})}$$

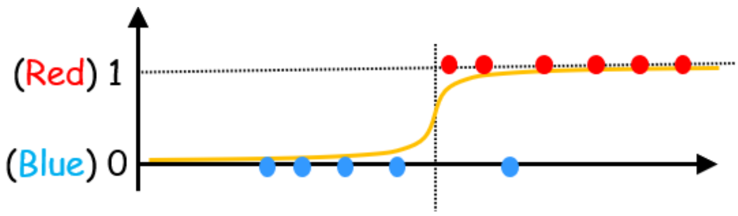
- For unknown data ***x***

$$\text{Label}(\mathbf{x}) = \begin{cases} 1\,(\text{Red}) & \text{if } h(\mathbf{x}) \geq 0.5 \\ 0\,(\text{Blue}) & \text{if } h(\mathbf{x}) < 0.5 \end{cases}$$

Find w which best fits given data

- Given

$$\mathbf{D} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \cdots, (\mathbf{x}_n, y_n)\} \text{ where } y_i \in \{0,1\}$$



$$h(\mathbf{x}) = \frac{1}{1+\exp(-\mathbf{w}\cdot\mathbf{x})}$$

- Approach 1
  - Define an error function
  - Find **w** which minimizes the error function

- Approach 2
  - Evaluate the probability of **D**
  - Find **w** which maximizes the probability

Approach with Error Function

- Define an error function

$$E(\mathbf{w}) = \sum_{i=1}^{n} \left(y_i - h(\mathbf{x})\right)^2$$

$$\text{where } h(\mathbf{x}) = \frac{1}{1+\exp(-\mathbf{w}\cdot\mathbf{x})}$$

- How to minimize the error function?
  - Solving Linear Equations?
  - Gradient Descent Method?

>>사실상 불가능하다(매우 힘들다)

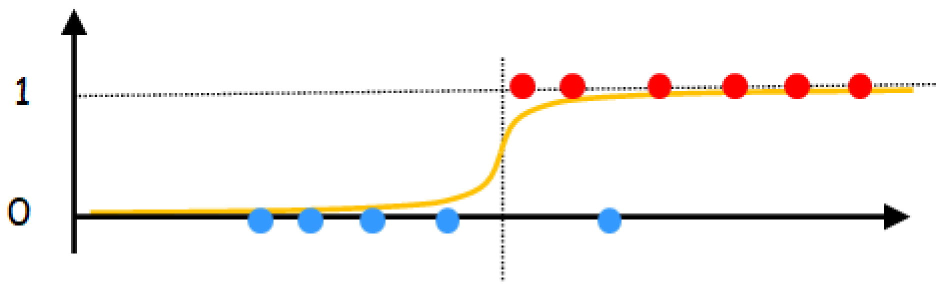
Approach with Probability

- Define a probability that yi is observed when xi and w are given

이 식을 보면 MLE다.

$$P(y_i \mid \mathbf{x}_i, \mathbf{w}) = \begin{cases} h(\mathbf{x}_i) & \text{if } y_i = 1 \\ 1 - h(\mathbf{x}_i) & \text{if } y_i = 0 \end{cases}$$

You have data



$$\mathbf{D} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \cdots, (\mathbf{x}_n, y_n)\} \text{ where } y_i \in \{0,1\}$$

- Probability that D is observed

$$\begin{aligned} P(\mathbf{D} \mid \mathbf{w}) &= \prod_{i=1}^{n} P(y_i, \mathbf{x}_i \mid \mathbf{w}) \\ &= \prod_{i=1}^{n} P(y_i \mid \mathbf{x}_i, \mathbf{w}) P(\mathbf{x}_i \mid \mathbf{w}) \\ &= \left(\prod_{i=1}^{n} \begin{cases} h(\mathbf{x}_i) & \text{if } y_i = 1 \\ 1 - h(\mathbf{x}_i) & \text{if } y_i = 0 \end{cases} \right) \left(\prod_{i=1}^{n} P(\mathbf{x}_i \mid \mathbf{w}) \right) \\ &= \left(\prod_{i \in \{i \mid y_i = 1\}} h(\mathbf{x}_i) \right) \left(\prod_{i \in \{i \mid y_i = 0\}} 1 - h(\mathbf{x}_i) \right) \left(\prod_{i=1}^{n} P(\mathbf{x}_i \mid \mathbf{w}) \right) \end{aligned}$$

$$P(y_i \mid \mathbf{x}_i, \mathbf{w}) = \begin{cases} h(\mathbf{x}_i) & \text{if } y_i = 1 \\ 1 - h(\mathbf{x}_i) & \text{if } y_i = 0 \end{cases}$$

Likelihood of **w**

Assume constant

- Maximum likelihood estimator of W

$$
\begin{aligned}
&\arg\max_{\mathbf{w}} P(\mathbf{D} \mid \mathbf{w}) \\
&= \arg\max_{\mathbf{w}} \log P(\mathbf{D} \mid \mathbf{w}) \\
&= \arg\max_{\mathbf{w}} \log \left(\prod_{i \in \{i \mid y_i = 1\}} h(\mathbf{x}_i) \right) \left(\prod_{i \in \{i \mid y_i = 0\}} 1 - h(\mathbf{x}_i) \right) \left(\prod_{i=1}^{n} P(\mathbf{x}_i \mid \mathbf{w}) \right) \\
&= \arg\max_{\mathbf{w}} \left(\sum_{i \in \{i \mid y_i = 1\}} \log h(\mathbf{x}_i) + \sum_{i \in \{i \mid y_i = 0\}} \log \left(1 - h(\mathbf{x}_i)\right) \right) \\
&= \arg\max_{\mathbf{w}} \left(\sum_{i \in \{i \mid y_i = 1\}} \log h(\mathbf{x}_i) + \sum_{i \in \{i \mid y_i = 0\}} \log \left(1 - h(\mathbf{x}_i)\right) \right) \\
&= \arg\max_{\mathbf{w}} \left(\sum_{i \in \{i \mid y_i = 1\}} \left(y_i \log h(\mathbf{x}_i) + (1 - y_i) \log\left(1 - h(\mathbf{x}_i)\right)\right) + \sum_{i \in \{i \mid y_i = 0\}} \left(y_i \log h(\mathbf{x}_i) + (1 - y_i) \log\left(1 - h(\mathbf{x}_i)\right)\right) \right) \\
&= \arg\max_{\mathbf{w}} \left(\sum_{i=1}^{n} \left(y_i \log h(\mathbf{x}_i) + (1 - y_i) \log\left(1 - h(\mathbf{x}_i)\right)\right) \right)
\end{aligned}
$$

- **Find MLE**

$$
\begin{aligned}
&\arg\max_{\mathbf{w}} \left(\sum_{i=1}^{n} \left(y_i \log h(\mathbf{x}_i) + (1 - y_i) \log\left(1 - h(\mathbf{x}_i)\right)\right) \right) \\
&= \arg\min_{\mathbf{w}} \left(-\sum_{i=1}^{n} \left(y_i \log h(\mathbf{x}_i) + (1 - y_i) \log\left(1 - h(\mathbf{x}_i)\right)\right) \right)
\end{aligned}
$$

$$
\text{where} \quad h(\mathbf{x}) = \frac{1}{1 + \exp(-\mathbf{w} \cdot \mathbf{x})}
$$

  - How to solve it?
    - Gradient Descent Method!!

- Find MLE by Gradient Descent

$$E(\mathbf{w}) = -\sum_{i=1}^{n}\left(y_i \log h(\mathbf{x}_i) + (1-y_i)\log\left(1-h(\mathbf{x}_i)\right)\right)$$

where $h(\mathbf{x}) = g\left(f(\mathbf{x})\right) = \dfrac{1}{1+\exp\left(-f(\mathbf{x})\right)} = \dfrac{1}{1+\exp\left(-(w_0x_0 + w_1x_1 + \cdots + w_dx_d)\right)}$

$$\boxed{\frac{\partial}{\partial w_j} E(\mathbf{w}) = \sum_{i=1}^{n}\left(h(\mathbf{x}_i) - y_i\right)\mathbf{x}_{ij}}$$

$$\begin{aligned} y &= -\mathbf{w}^T\mathbf{x} \\ &= -(w_0x_0 + w_1x_1 + \cdots + w_dx_d) \end{aligned}$$

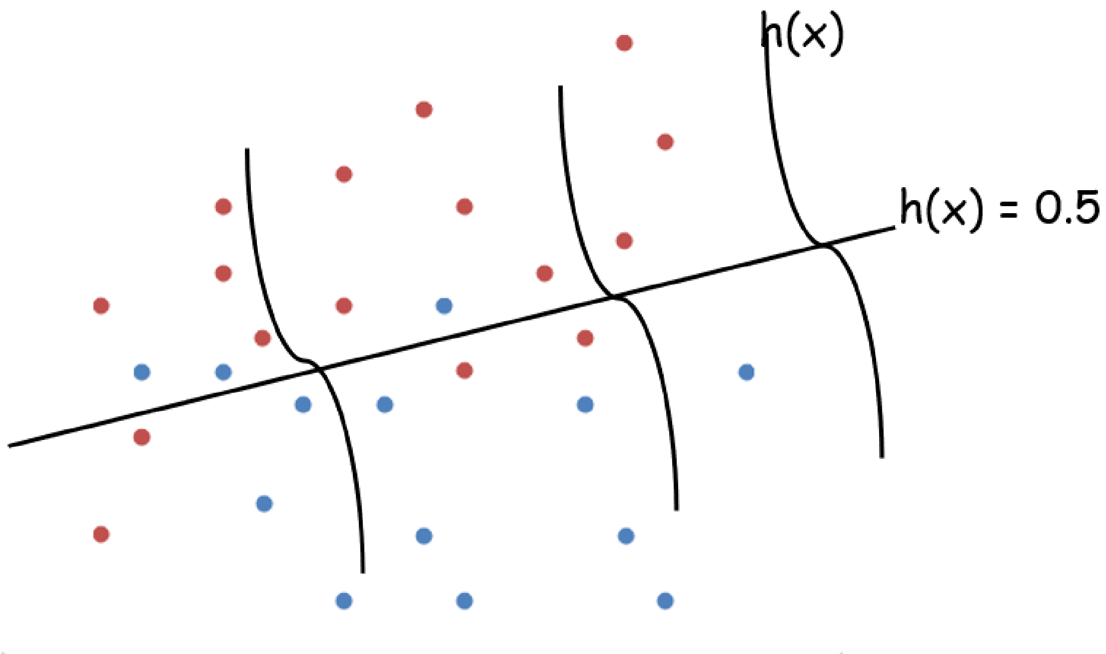
$$\frac{dy}{dw_j} = -x_j$$

$$h(x) = \frac{1}{1+\exp(-x)}$$

$$\frac{d}{dx}h(x) = h(x)\left(1-h(x)\right)$$

$$
\begin{aligned}
\frac{\partial}{\partial w_j}E(\mathbf{w}) &= -\sum_{i=1}^{n}\frac{\partial}{\partial w_j}\left(y_i \log g\big(f(\mathbf{x}_i)\big)+(1-y_i)\log\big(1-g\big(f(\mathbf{x}_i)\big)\big)\right) \\
&= -\sum_{i=1}^{n}\frac{\partial f(\mathbf{x}_i)}{\partial w_j}\frac{\partial g\big(f(\mathbf{x}_i)\big)}{\partial f(\mathbf{x}_i)}\frac{\partial}{\partial g\big(f(\mathbf{x}_i)\big)}\left(y_i \log g\big(f(\mathbf{x}_i)\big)+(1-y_i)\log\big(1-g\big(f(\mathbf{x}_i)\big)\big)\right) \\
&= -\sum_{i=1}^{n}\frac{\partial f(\mathbf{x}_i)}{\partial w_j}\frac{\partial g\big(f(\mathbf{x}_i)\big)}{\partial f(\mathbf{x}_i)}\left(y_i\frac{1}{g\big(f(\mathbf{x}_i)\big)}-(1-y_i)\frac{1}{1-g\big(f(\mathbf{x}_i)\big)}\right) \\
&= -\sum_{i=1}^{n}\frac{\partial f(\mathbf{x}_i)}{\partial w_j}g\big(f(\mathbf{x}_i)\big)\big(1-g\big(f(\mathbf{x}_i)\big)\big)\left(y_i\frac{1}{g\big(f(\mathbf{x}_i)\big)}-(1-y_i)\frac{1}{1-g\big(f(\mathbf{x}_i)\big)}\right) \\
&= -\sum_{i=1}^{n}x_{ij}\,g\big(f(\mathbf{x}_i)\big)\big(1-g\big(f(\mathbf{x}_i)\big)\big)\left(y_i\frac{1}{g\big(f(\mathbf{x}_i)\big)}-(1-y_i)\frac{1}{1-g\big(f(\mathbf{x}_i)\big)}\right) \\
&= -\sum_{i=1}^{n}x_{ij}\left(y_i\big(1-g\big(f(\mathbf{x}_i)\big)\big)-(1-y_i)\,g\big(f(\mathbf{x}_i)\big)\right) \\
&= -\sum_{i=1}^{n}x_{ij}\left(y_i-g\big(f(\mathbf{x}_i)\big)\right)=\sum_{i=1}^{n}x_{ij}\left(h(\mathbf{x}_i)-y_i\right)
\end{aligned}
$$

Find the Boundary



Discussion

- Logistic Regression
    - No closed form solution
        - optimized by Gradient Descent method
    - A linear boundary
    - Binary classifier
    - 종속변수가 범주형 데이터이며 이진형일때 logistic regression (분류)