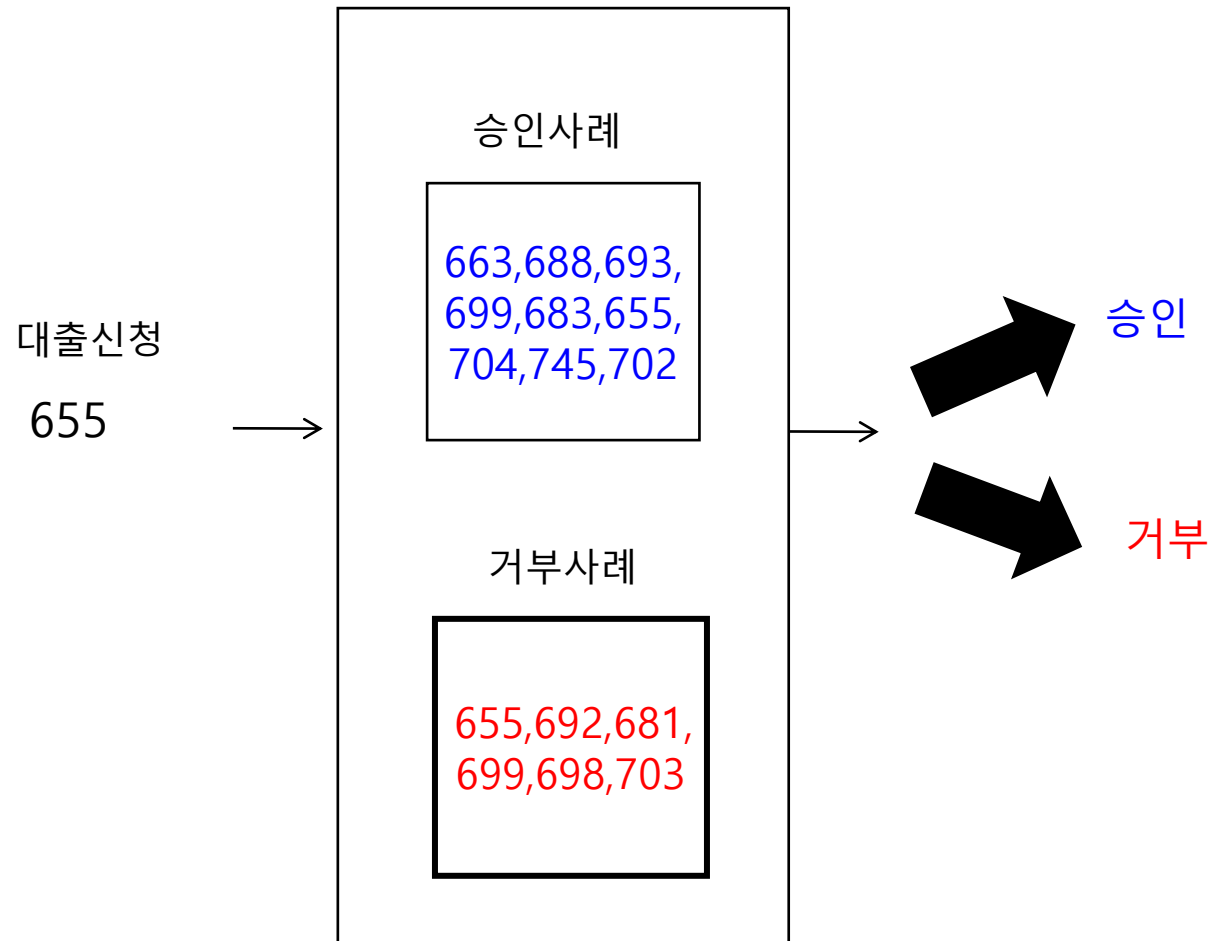


이진 분류



신용도에 따른 대출 승인



당뇨병 진단 예측

| Pregnancies | Glucose | BloodPressure | SkinThickness | Insulin | BMI | DiabetesPedigreeFunction | Age |
|-------------|---------|---------------|---------------|---------|------|--------------------------|-----|
| 6 | 148 | 72 | 35 | 0 | 33.6 | 0.627 | 50 |
| 1 | 85 | 66 | 29 | 0 | 26.6 | 0.351 | 31 |
| 8 | 183 | 64 | 0 | 0 | 23.3 | 0.672 | 32 |
| 1 | 89 | 66 | 23 | 94 | 28.1 | 0.167 | 21 |
| 0 | 137 | 40 | 35 | 168 | 43.1 | 2.288 | 33 |

Pregnancies - 임신횟수

Glucose - 글루코스 내성

BloodPressure - 확장기 혈압

SkinThickness - 상완심두근 피부 두께

Insulin - 혈액내 인슐린 수치

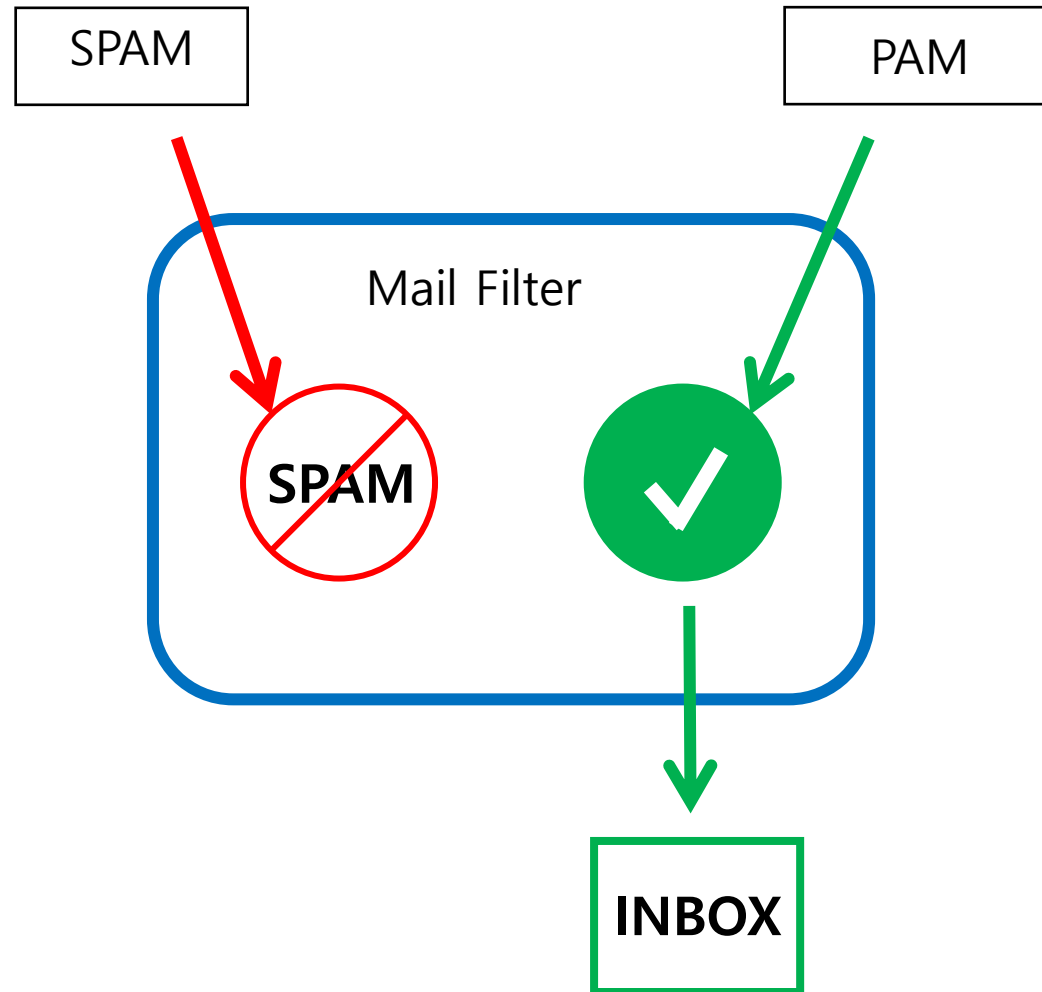
BMI - BMI 지수

DiabetesPedigreeFunction - 당뇨병 가족력

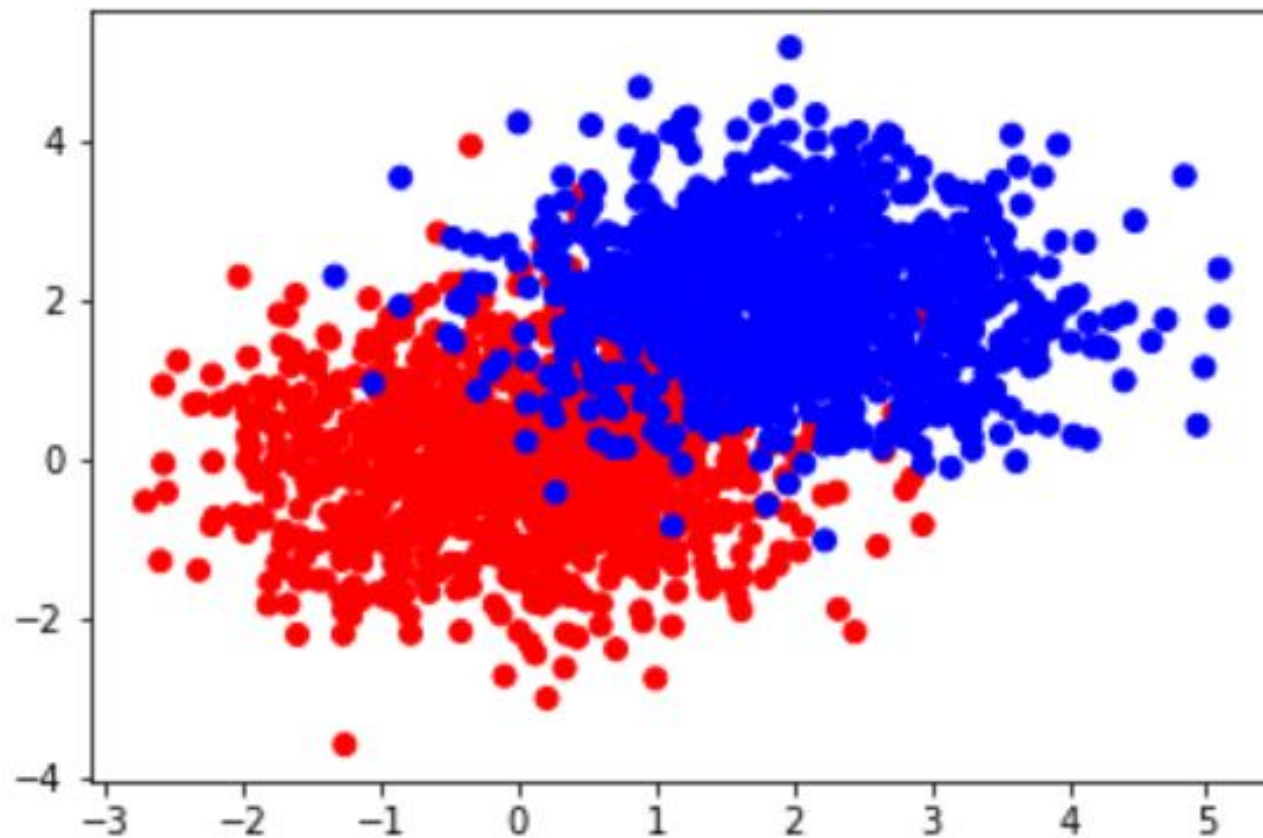
Age - 나이

SPAM 필터링

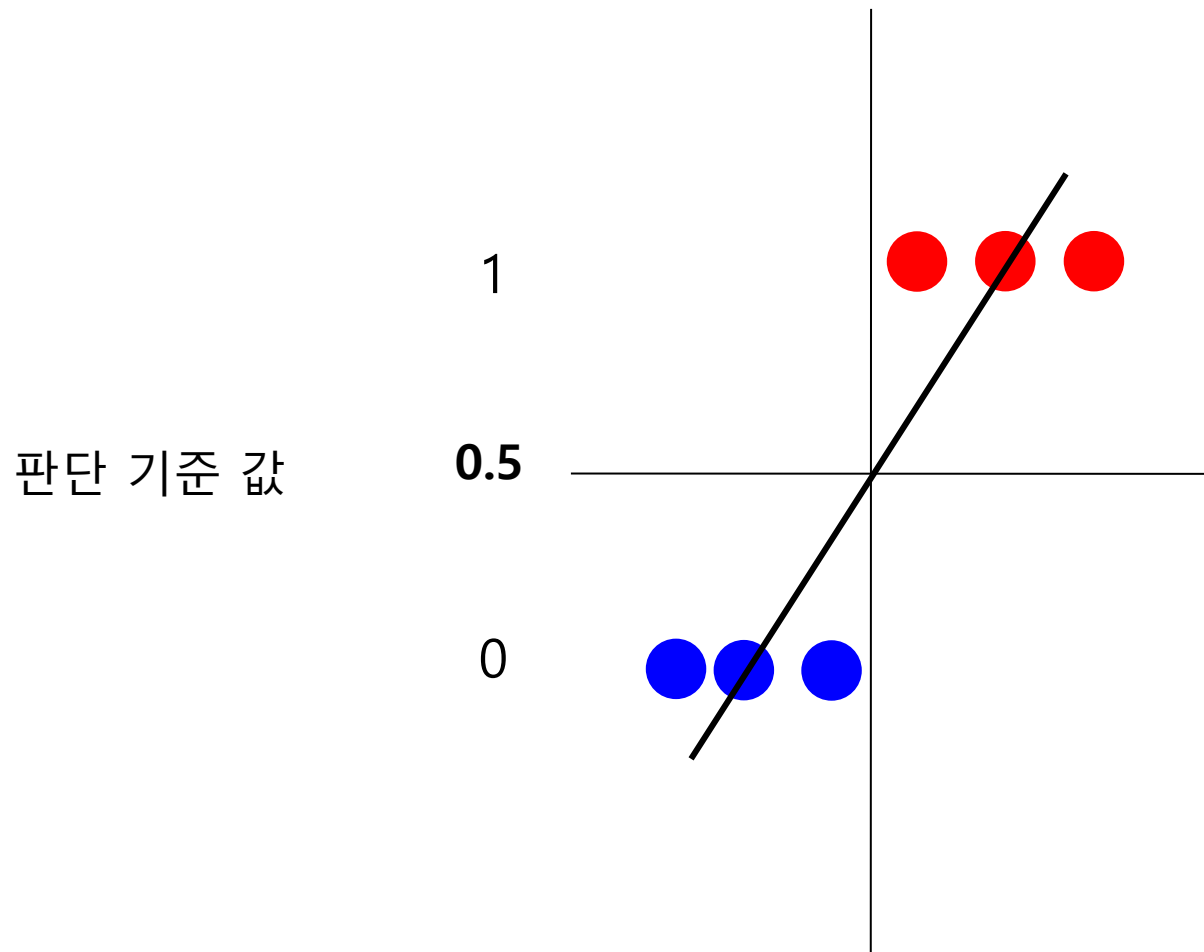
| | SPAM | PAM |
|--------|------|-----|
| Free | 8 | 2 |
| Cash | 8 | 1 |
| Can | 2 | 9 |
| Won | 7 | 2 |
| Never | 10 | 1 |
| Chance | 9 | 3 |
| Sorry | 0 | 8 |
| Need | 2 | 9 |
| Click | 10 | 0 |
| Can | 3 | 8 |



2진 분류

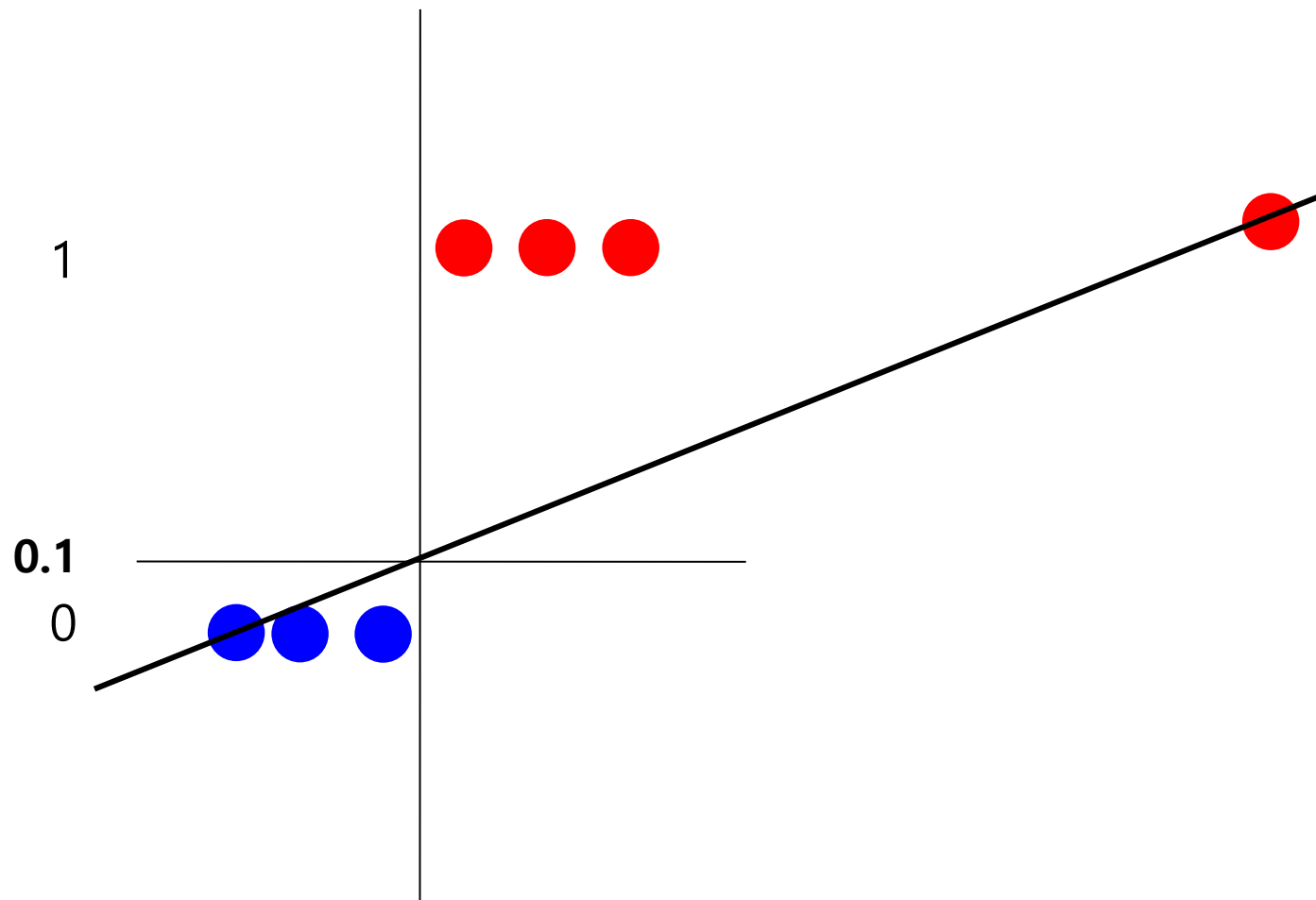


직선을 이용한 이항 분포 모델링

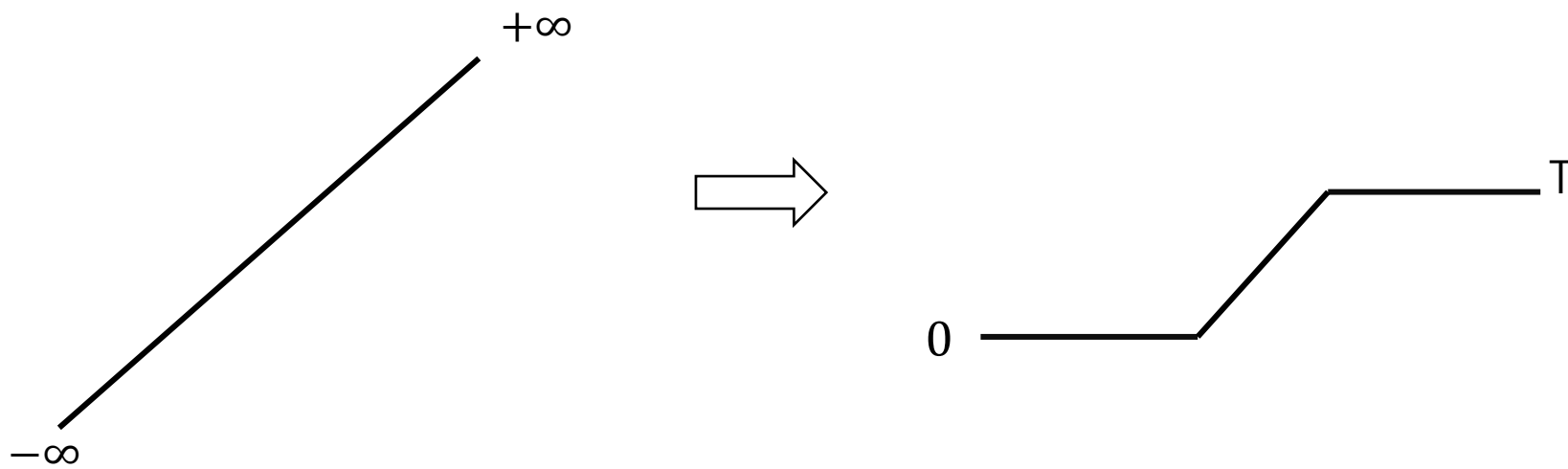


직선을 이용한 이항 분포 모델링

판단 기준 값

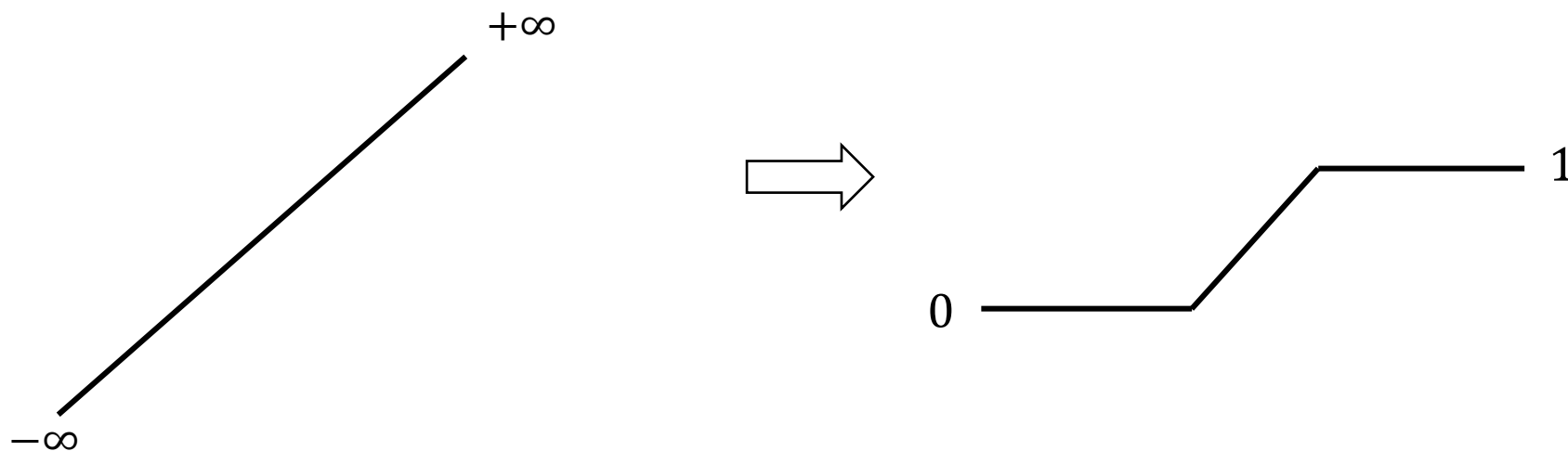


출력 값 제한



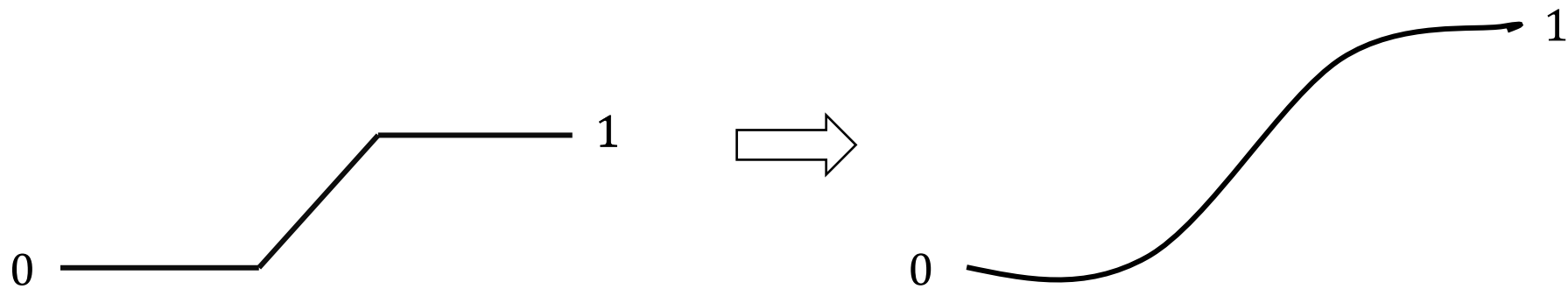
$$y=wx+b, \quad \begin{array}{ll} \text{if } y > T, & y = T \\ \text{if } y < 0, & y = 0 \end{array}$$

확률 함수로 변환



$$y=wx+b, \quad \begin{array}{l} \text{if } y > 1, y = 1 \\ \text{if } y < 0, y = 0 \end{array}$$

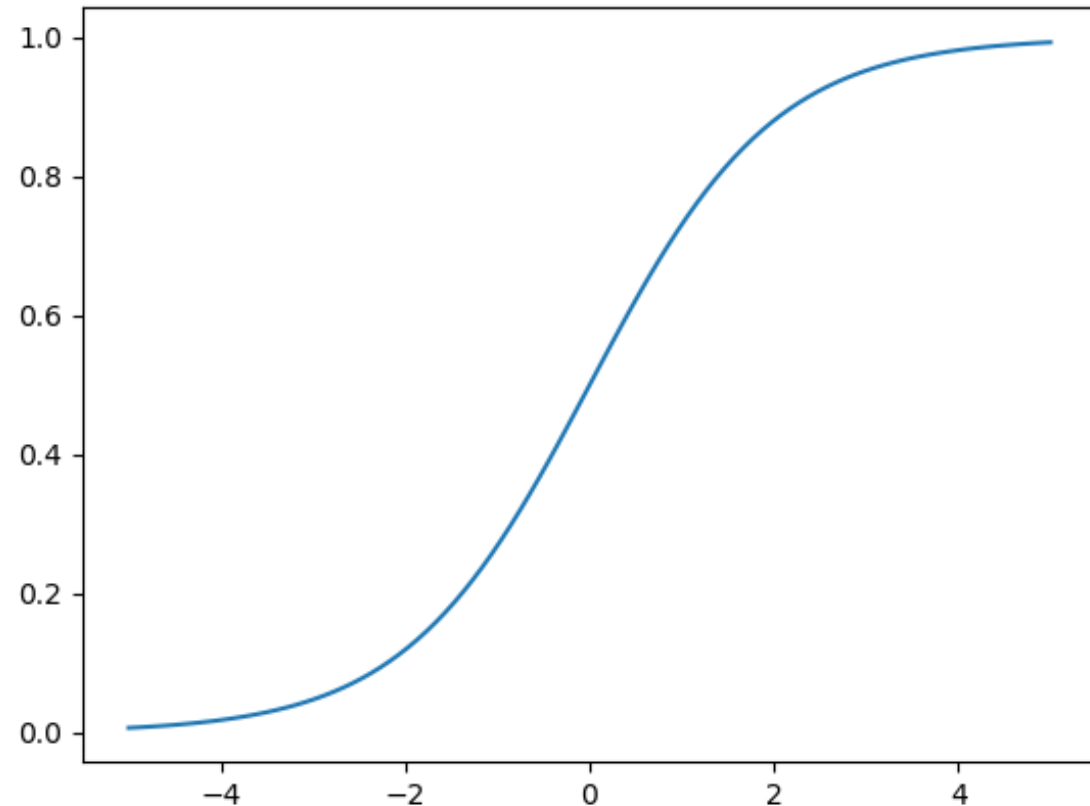
미분 가능한 연속 함수로 변환



시그모이드 함수(sigmoid function)

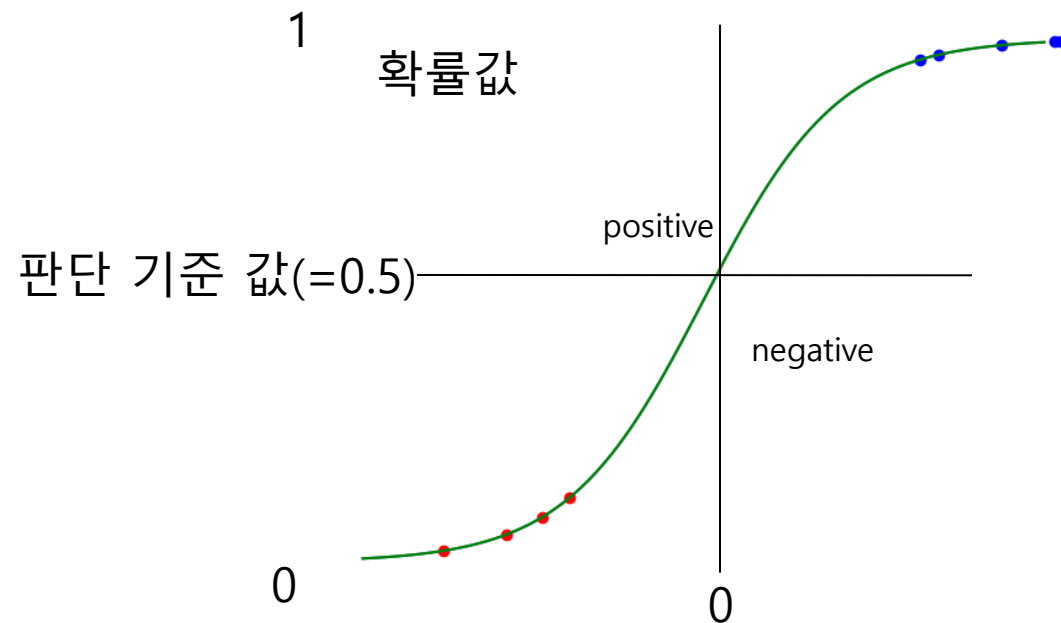
- S자 모양을 갖는 함수
- 중심축 0을 중심으로 좌측은 0으로 수렴, 우측은 1로 수렴

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



로지스틱 회귀

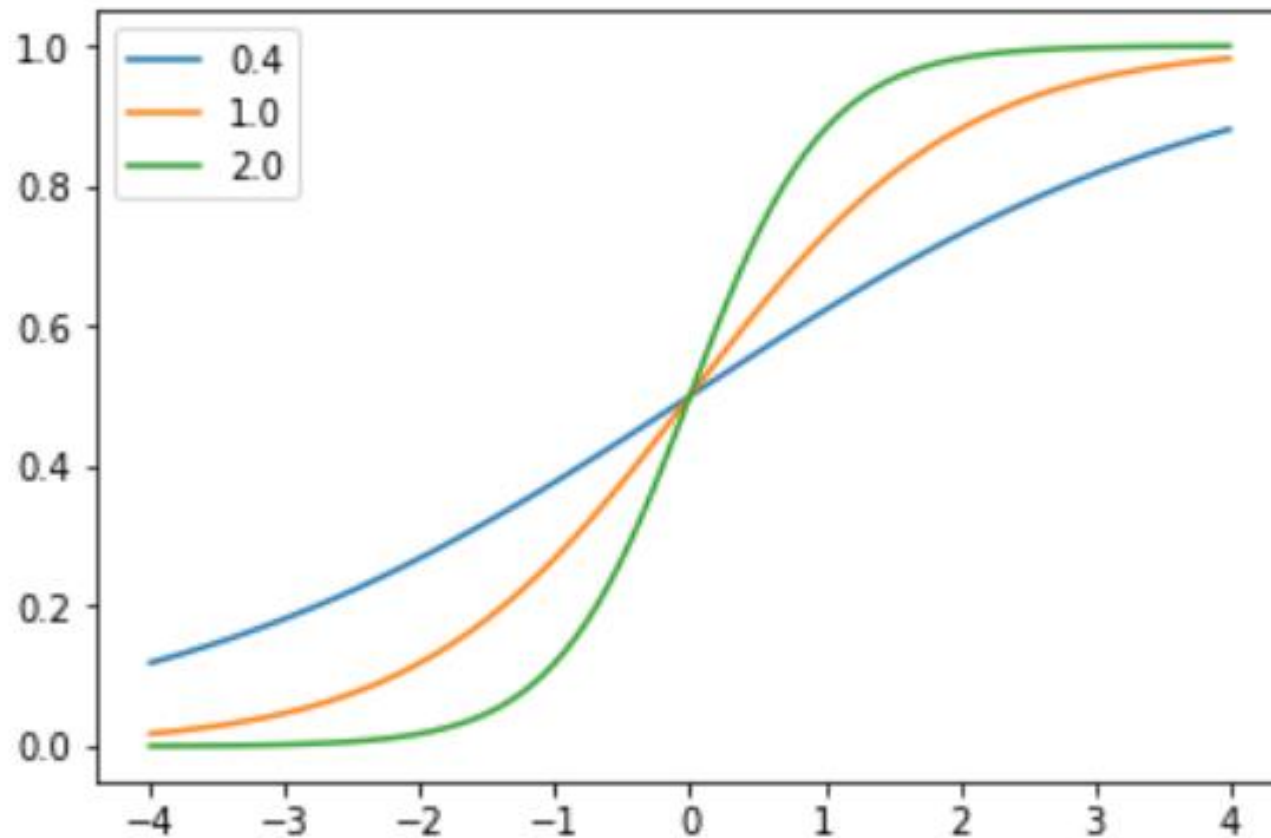
$$y = H(x) = \frac{1}{1 + e^{-(wx+b)}}$$



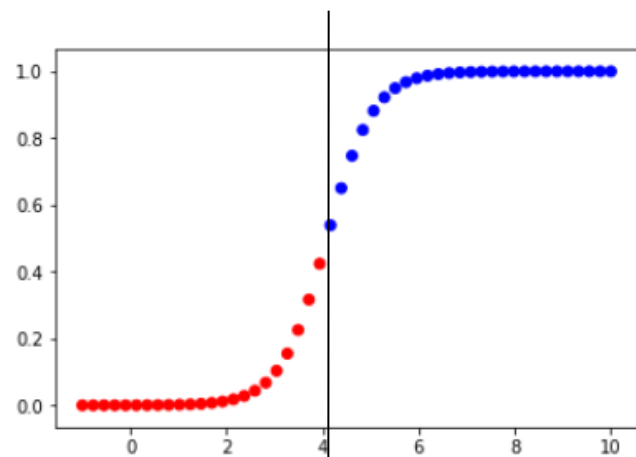
로지스틱 회귀

- $w = [0.4, 1.0, 2.0]$
- $b=0$

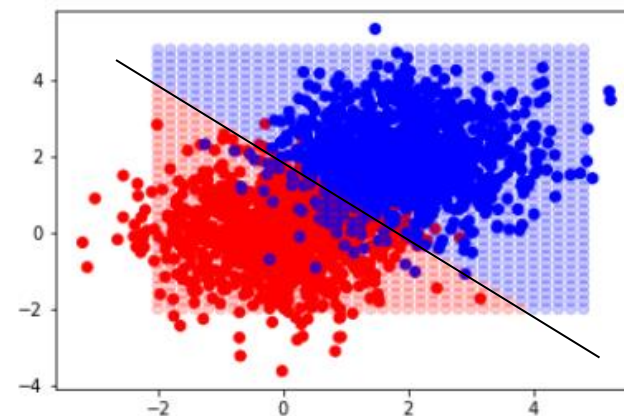
$$y = H(x) = \frac{1}{1 + e^{-(wx+b)}}$$



로지스트 회귀



1차원 데이터



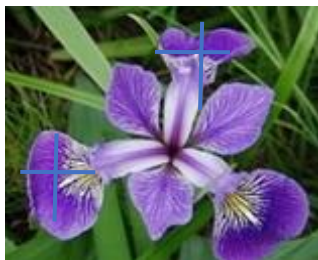
2차원 데이터

다중 분류

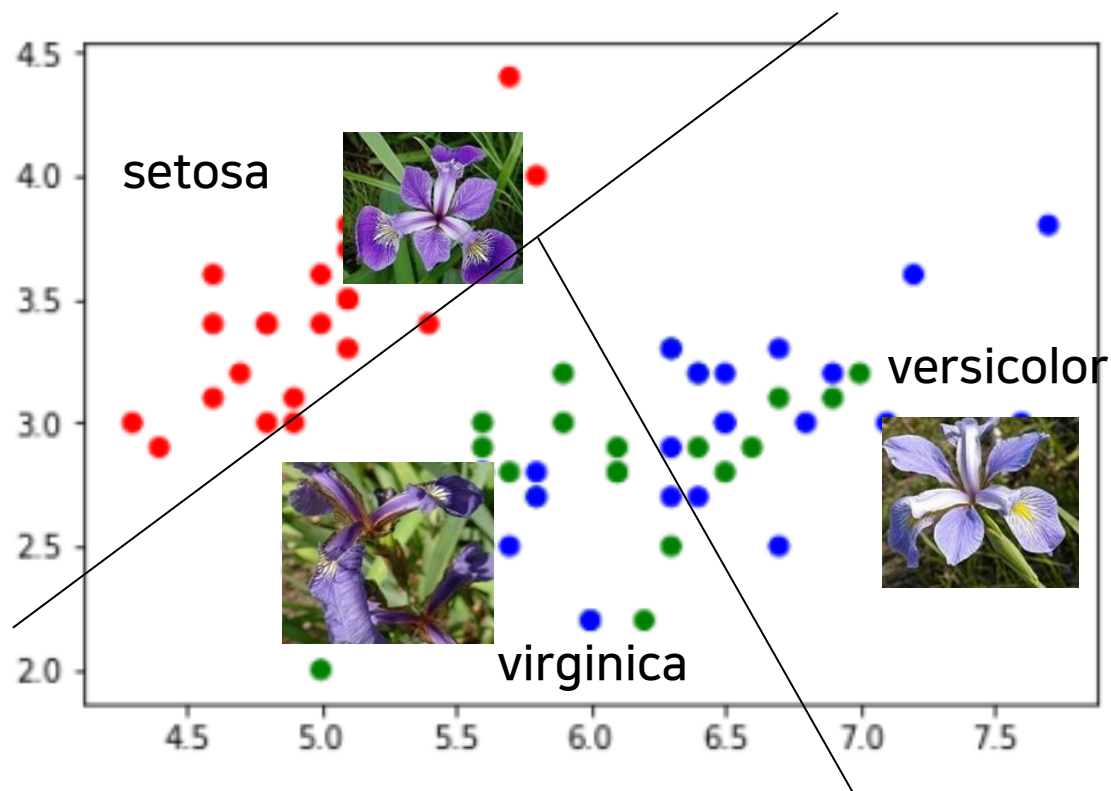


붓꽃 분류

- 3개의 영역을 잘 구분할 수 있는 직선 학습 : softmax 알고리즘

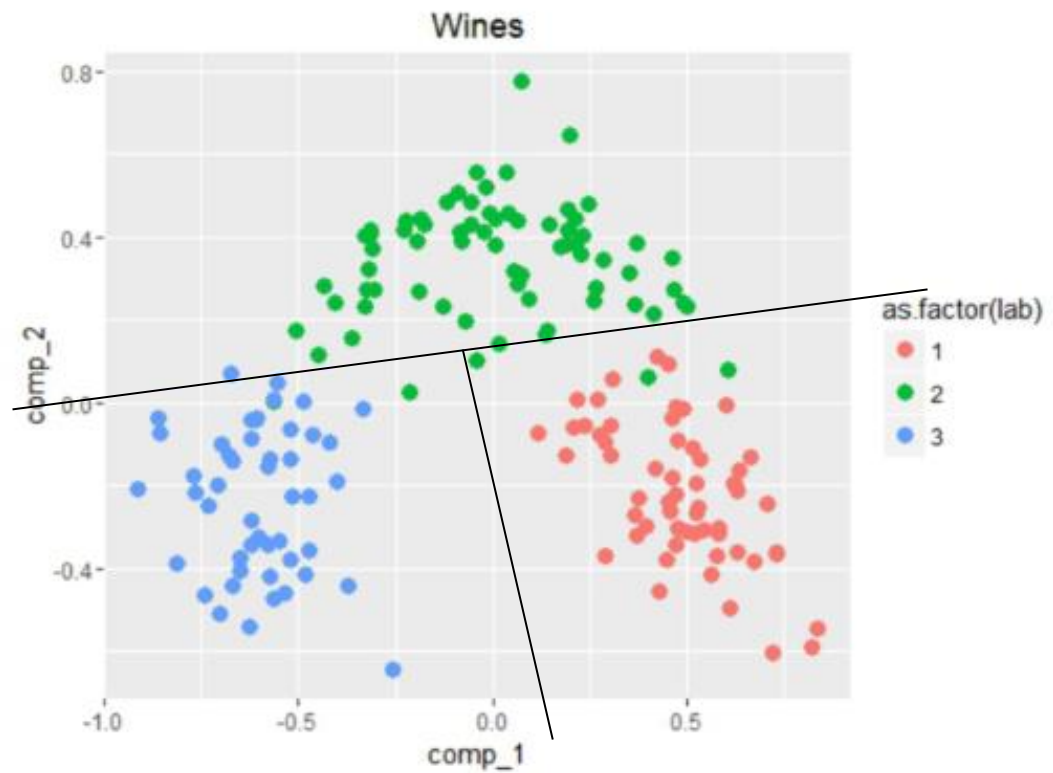


꽃잎의 폭, 길이
꽃받침의 폭, 길이



와인 분류

13개의 화학적 분석
- 알코올
- 사과산
- 마그네슘
- 색 강도
3가지 와인 범주로 분류



숫자 인식

- 186(28x28)개의 특징, 10(0~9)개의 class로 구성

- 28x28영상
- 0 ~ 9 필기 숫자



영상 인식

- 28x28영상
- T-shirt, Coat, Bag 등 10개 범주

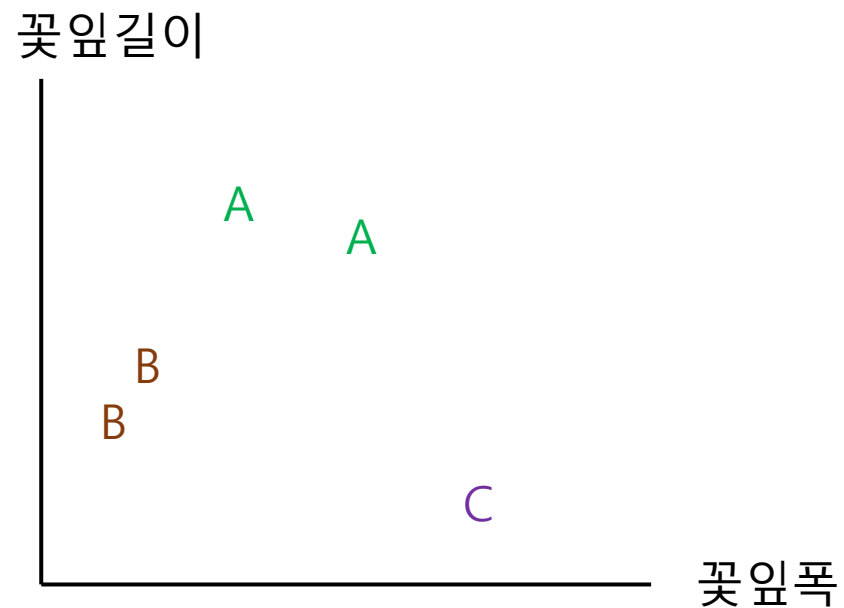


<https://github.com/zalandoresearch/fashion-mnist>

다중 클래스(Multinomial Classification)

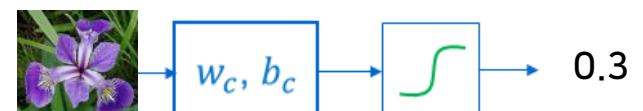
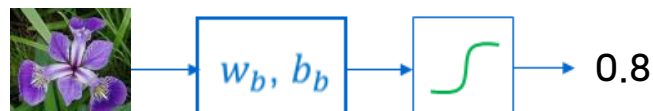
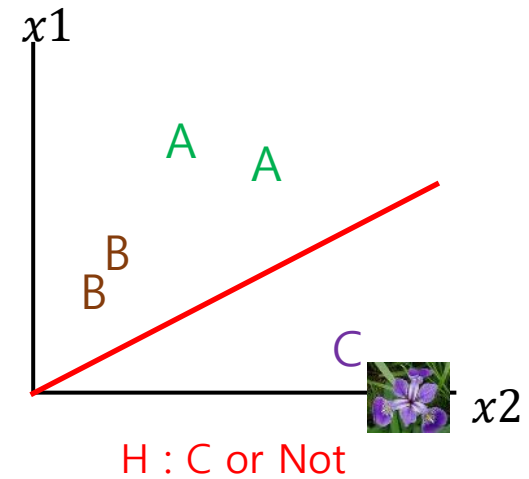
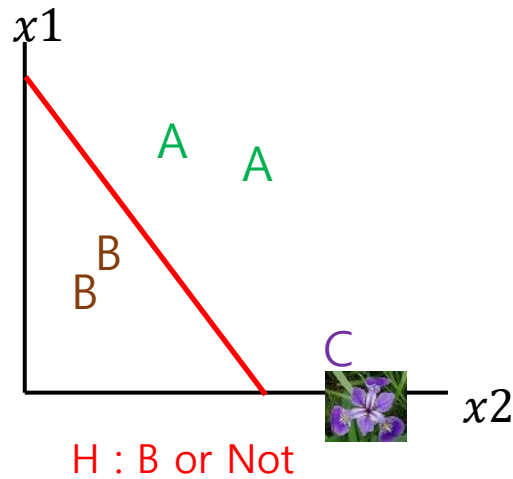
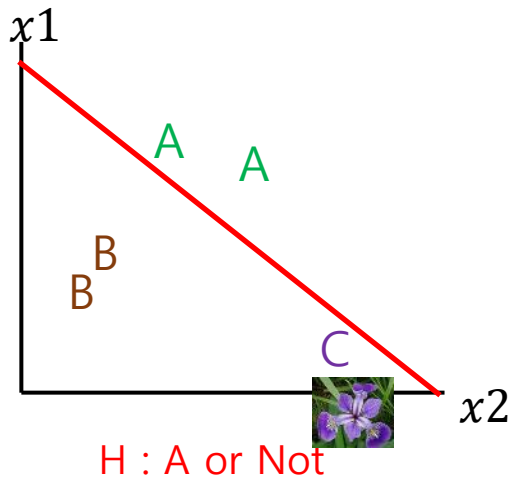
- 주어진 5개 데이터의 붓꽃 3종류 예측

| 꽃잎폭 | 꽃잎길이 | 붓꽃종류 |
|-----|------|------|
| 3 | 8 | A |
| 4 | 7 | A |
| 3 | 4 | B |
| 2 | 2 | B |
| 8 | 2 | C |



다중 클래스(Multinomial Classification)

- 3개의 class 분류를 위해 3개의 분류기 사용
 - 임의의 붓꽃 data에 대해 각각의 w 와 b 를 학습한 분류기에서 확률값 추정

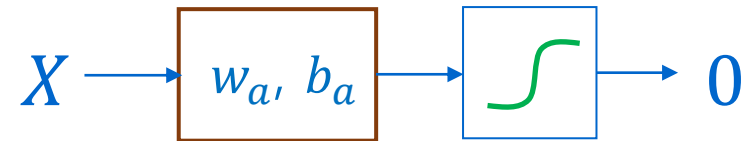


다중 클래스(Multinomial Classification)

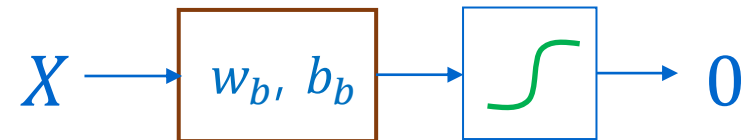
- data를 행렬로 변형하면 계산을 쉽고 빠르게 할 수 있음

$$\begin{array}{ccc} X & & Y \\ (x_1, x_2) & \longrightarrow & (0, 0, 1) \\ & & y_1, y_2, y_3 \end{array}$$

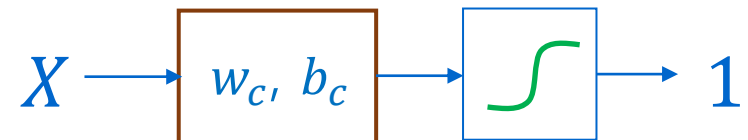
$$[w_{a_1} \ w_{a_2}] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + b_a \xrightarrow{f} y_1$$



$$[w_{b_1} \ w_{b_2}] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + b_b \xrightarrow{f} y_2$$



$$[w_{c_1} \ w_{c_2}] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + b_c \xrightarrow{f} y_3$$



행렬 표현으로 바꾸기 ($y=wx+b$)

- 3개의 분류기를 하나의 수식으로 묶어서 간단하게 표현
 - > w 들을 열벡터로 표현해 class 수 \times 특징 차원 수 행렬로 표현 (3x2 행렬)

$$\begin{bmatrix} w_{a_1} & w_{a_2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + b_a = w_{a_1}x_1 + w_{a_2}x_2 + b_a$$

$$\begin{bmatrix} w_{b_1} & w_{b_2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + b_b = w_{b_1}x_1 + w_{b_2}x_2 + b_b$$

$$\begin{bmatrix} w_{c_1} & w_{c_2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + b_c = w_{c_1}x_1 + w_{c_2}x_2 + b_c$$

$$\begin{bmatrix} w_{a_1} & w_{a_2} \\ w_{b_1} & w_{b_2} \\ w_{c_1} & w_{c_2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} b_a \\ b_b \\ b_c \end{bmatrix} = \begin{bmatrix} w_{a_1}x_1 + w_{a_2}x_2 + b_a \\ w_{b_1}x_1 + w_{b_2}x_2 + b_b \\ w_{c_1}x_1 + w_{c_2}x_2 + b_c \end{bmatrix}$$

행렬 표현으로 바꾸기 ($y=xw+b$)

- 클래스 3, 데이터 4개인 경우 → 데이터를 행벡터로 표기
- 열벡터로 정의된 w 는 샘플 수에 따라서 행렬의 크기가 변화되지 않고 고정됨

$$\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \\ x_{41} & x_{42} \end{bmatrix} \begin{bmatrix} w_{a1} & w_{b1} & w_{c1} \\ w_{a2} & w_{b2} & w_{c2} \end{bmatrix} + \begin{bmatrix} b_1 & b_2 & b_3 \\ b_1 & b_2 & b_3 \\ b_1 & b_2 & b_3 \\ b_1 & b_2 & b_3 \end{bmatrix}$$

$$[4 \times 2] \times [2 \times 3] + [4 \times 3] = [4 \times 3]$$

$$= \begin{bmatrix} x_{11}w_{a1} + x_{12}w_{a2} + b_1 & x_{11}w_{b1} + x_{12}w_{b2} + b_2 & x_{11}w_{c1} + x_{12}w_{c2} + b_3 \\ x_{21}w_{a1} + x_{22}w_{a2} + b_1 & x_{21}w_{b1} + x_{22}w_{b2} + b_2 & x_{21}w_{c1} + x_{22}w_{c2} + b_3 \\ x_{31}w_{a1} + x_{32}w_{a2} + b_1 & x_{31}w_{b1} + x_{32}w_{b2} + b_2 & x_{31}w_{c1} + x_{32}w_{c2} + b_3 \\ x_{41}w_{a1} + x_{42}w_{a2} + b_1 & x_{41}w_{b1} + x_{42}w_{b2} + b_2 & x_{41}w_{c1} + x_{42}w_{c2} + b_3 \end{bmatrix}$$

행렬 표현으로 바꾸기 ($y=xw+b$)

$$\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \\ x_{41} & x_{42} \end{bmatrix} \begin{bmatrix} w_{a_1} & w_{b_1} & w_{c_1} \\ w_{a_2} & w_{b_2} & w_{c_2} \end{bmatrix} + \begin{bmatrix} b_1 & b_2 & b_3 \\ b_1 & b_2 & b_3 \\ b_1 & b_2 & b_3 \\ b_1 & b_2 & b_3 \end{bmatrix} \longrightarrow \begin{bmatrix} x_{11} & x_{12} & 1 \\ x_{21} & x_{22} & 1 \\ x_{31} & x_{32} & 1 \\ x_{41} & x_{42} & 1 \end{bmatrix} \begin{bmatrix} w_{a_1} & w_{b_1} & w_{c_1} \\ w_{a_2} & w_{b_2} & w_{c_2} \\ b_1 & b_2 & b_3 \end{bmatrix}$$

One-hot 인코딩

- Sample 하나에 대해 3개의 y값 필요

2클래스 분류



[1]



[0]

1개의 분류기 필요

3클래스 분류



3개의 분류기 필요

One-hot 인코딩으로 표현할 때
행렬 크기는 $n \times \text{class 수}$

One-hot 인코딩

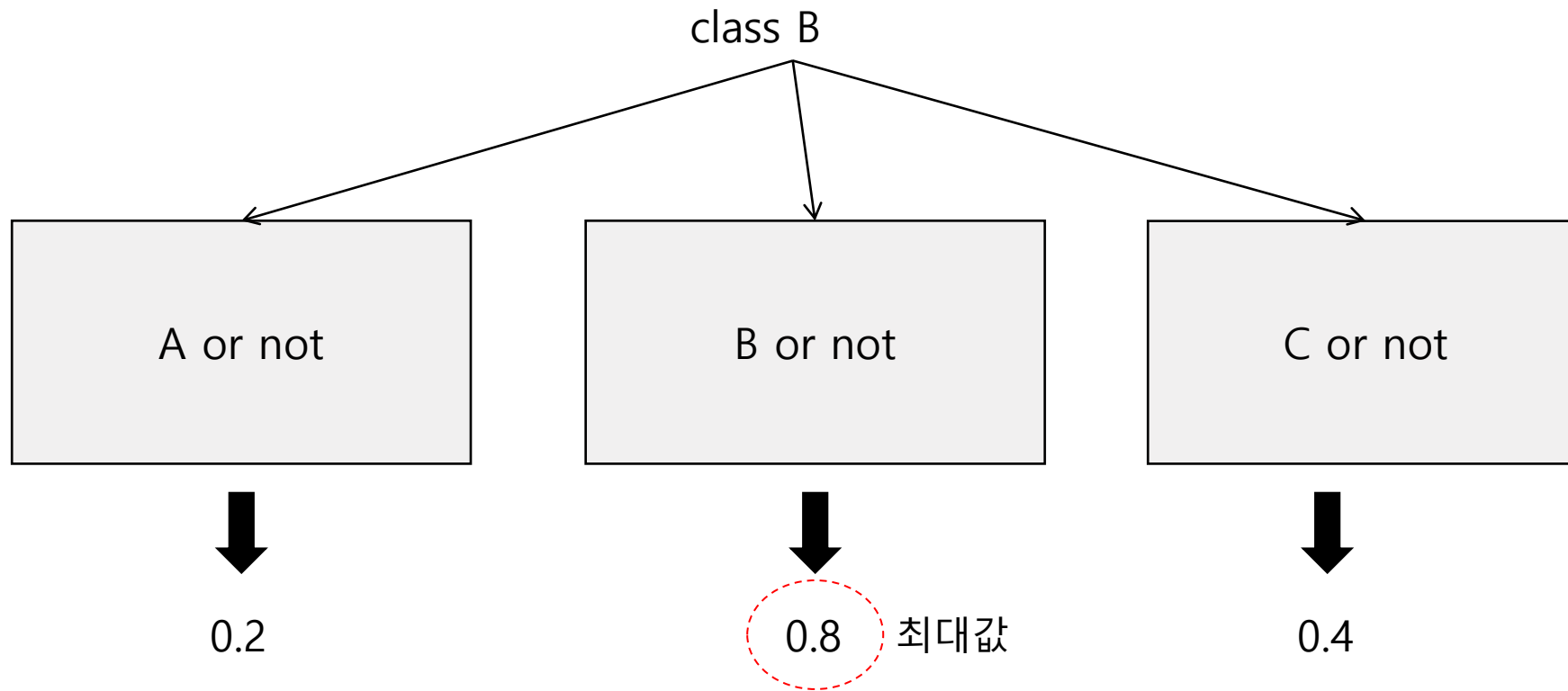
[1 0 0]

[0 1 0]

[0 0 1]

다중 클래스 예측

- 주어진 data는 최대값을 갖는 class로 분류



클래스 레이블 예측

- 예측값은 세 개 중에서 가장 확률값이 높은 class의 위치 값

$$h = wx + b$$

$$y = \operatorname{argmax}(h)$$

Ex)

$$h = [0.05 \quad 0.8 \quad 0.15]$$

$$y = 1$$

클래스 레이블 예측

- N 개의 샘플, C 개의 클래스, d 특징 차원의 경우 행렬 크기

$$H = XW$$
$$Y = \operatorname{argmax}(H)$$



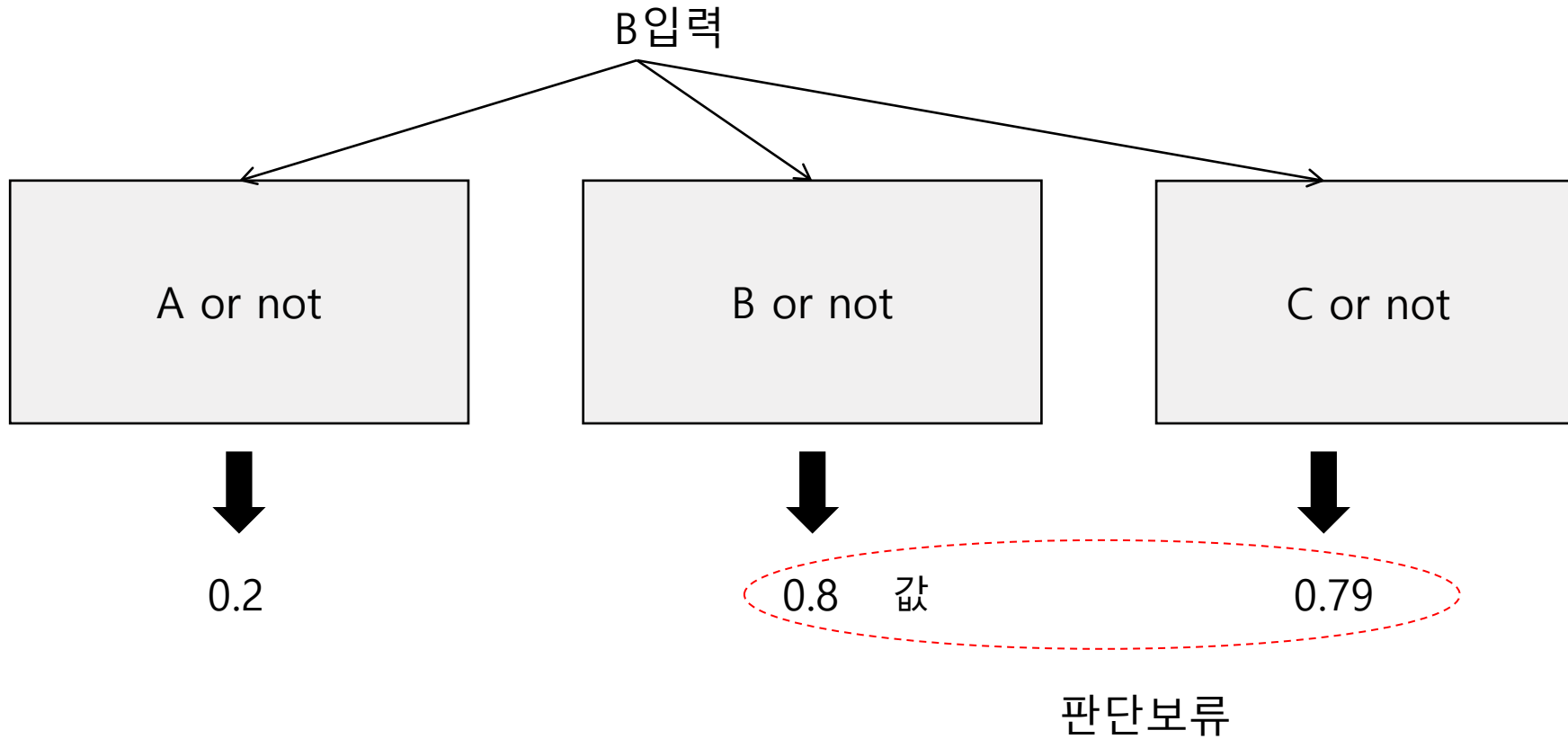
$$X = N \times (d + 1)$$
$$W = (d + 1) \times c$$
$$H = N \times c$$
$$Y = N$$

Ex) 클래스 3개, 샘플 2개

$$H = \begin{bmatrix} 0.9 & 0.02 & 0.08 \\ 0.05 & 0.8 & 0.15 \end{bmatrix}$$

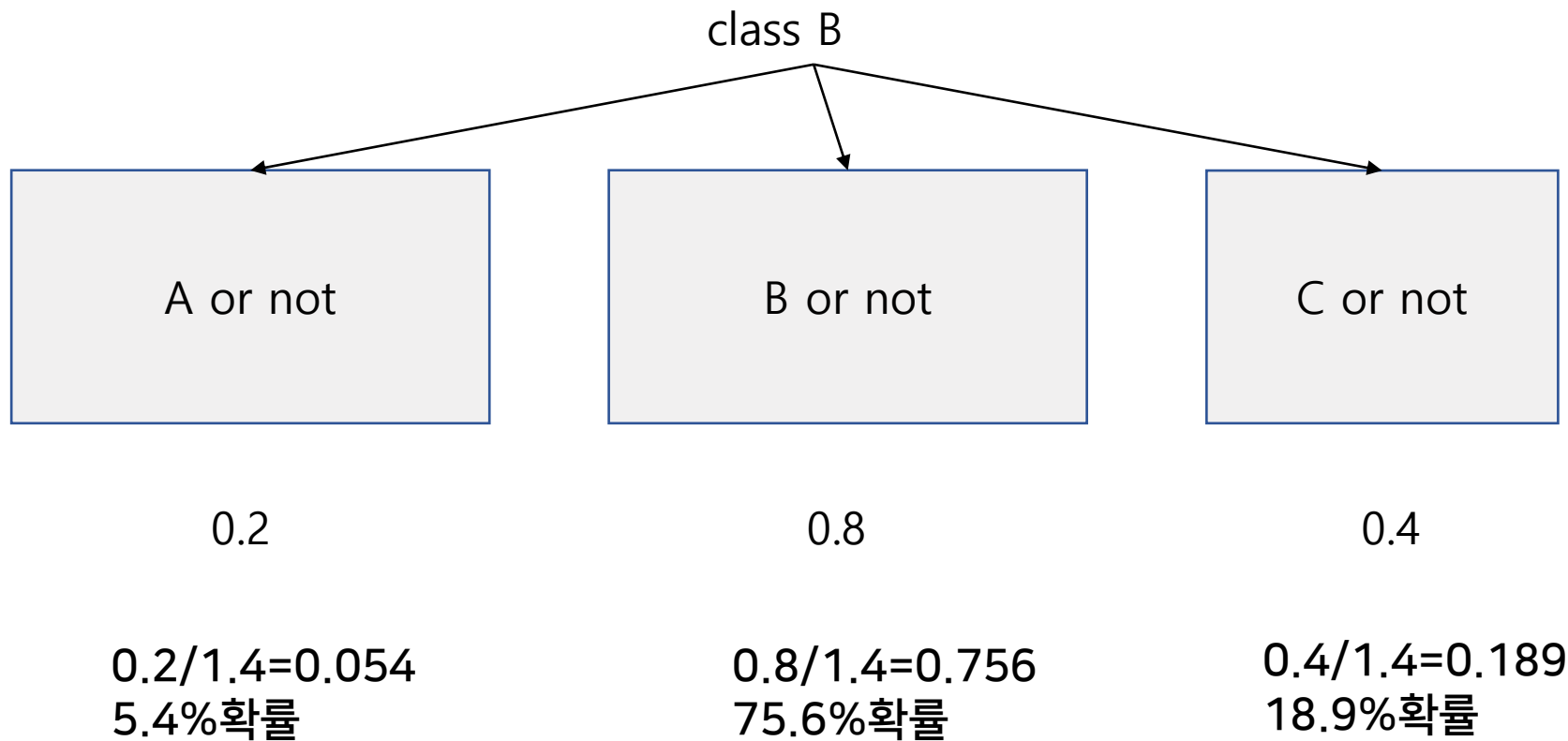
$$Y = \begin{bmatrix} 0 & 1 \end{bmatrix}$$

다중 클래스 예측



다중 클래스 예측

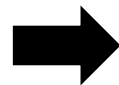
- 확률의 합이 1을 초과하므로 정규화 필요



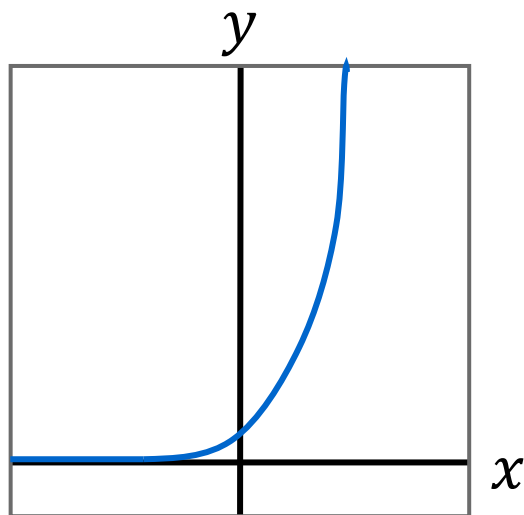
Softmax 정규화

- 전체 data 합으로 자신의 출력값을 나누어줄 때 지수 함수 사용
 - > 지수함수를 사용하는 이유는 값이 급격하게 변하기 때문

$$\sigma(x_j) = \frac{x_j}{\sum x_i}$$



$$\sigma(x_j) = \frac{\exp(x_j)}{\sum \exp(x_i)}$$



$$y = \exp(x) = e^x$$

Softmax 정규화

- $[0.2 \ 0.8 \ 0.4], \text{sum} = 1.2$
- 정규화
 - > $[0.2 \ 0.8 \ 0.4] / 1.2 = [0.142, 0.5714, 0.2857]$
- Softmax 정규화
 - > $[1.22, 2.22, 1.49], \text{sum}=4.93$
 - > $[1.22, 2.22, 1.49] / 4.93 = [0.24, 0.45, 0.30]$

클래스 레이블 예측

- 예측값은 세 개 중에서 가장 확률값이 높은 class의 위치 값

$$h = \text{softmax}(wx + b)$$

$$\hat{y} = \text{argmax}(h)$$

Ex)

$$h = [0.05 \quad 0.8 \quad 0.15]$$

$$\hat{y} = 1$$

다중 클래스 학습

- 학습 데이터

> 학습데이터 x 와 one-hot 코딩 레이블

$(x, [1\ 0\ 0\ 0\ 0\ 0])$

- 학습 모델 정의

> softmax

$s(x) = \text{softmax}(wx + b)$

- 예측

> w, b 를 사용해 함수의 최대값 위치

$h = \text{softmax}(wx + b)$
 $y = \text{argmax}(h)$