

데이터 과학

L05: Softmax Regression

Kookmin University

지도학습 Supervised Learning

훈련 데이터(Training Data)로부터 하나의 함수를 유추해내기 위한 기계 학습(Machine Learning)의 한 방법

Training Data

[1.2, 3.8, -1.4, ..., 4.1]	→	1.1
[3.2, -1.2, -0.2, ..., 2.1]	→	2.7
[2.8, -1.4, -0.3, ..., 2.3]	→	2.8
[1.2, 3.4, -1.5, ..., 4.2]	→	0.9
[4.2, 2.1, 2.8, ..., -0.5]	→	-0.1
...		
[3.2, 2.2, 2.2, ..., -0.4]	→	-0.2

Test

[1.3, 3.2, -1.5, ..., 4.1]	→	?
----------------------------	---	---

이진 분류 문제 Binary Classification

종속 변수 y 가 0 또는 1인 경우의 회귀 분석

Training Data

[1.2, 3.8, -1.4, ..., 4.1]	→	0
[3.2, -1.2, -0.2, ..., 2.1]	→	0
[2.8, -1.4, -0.3, ..., 2.3]	→	1
[1.2, 3.4, -1.5, ..., 4.2]	→	0
[4.2, 2.1, 2.8, ..., -0.5]	→	1
...		
[3.2, 2.2, 2.2, ..., -0.4]	→	1

Test

[1.3, 3.2, -1.5, ..., 4.1]	→	?
----------------------------	---	---

Logistic Regression

가설함수:

$$H(x) = \frac{1}{1 + e^{-(\mathbf{w}^\top \mathbf{x} + b)}}$$

비용:

$$cost(\mathbf{w}, b) = \frac{1}{n} \sum_{i=0}^n C(H(\mathbf{x}_i), y_i)$$

$$C(h, y) = \begin{cases} -\log(1 - h) & \text{if } y = 0 \\ -\log(h) & \text{if } y = 1 \end{cases}$$

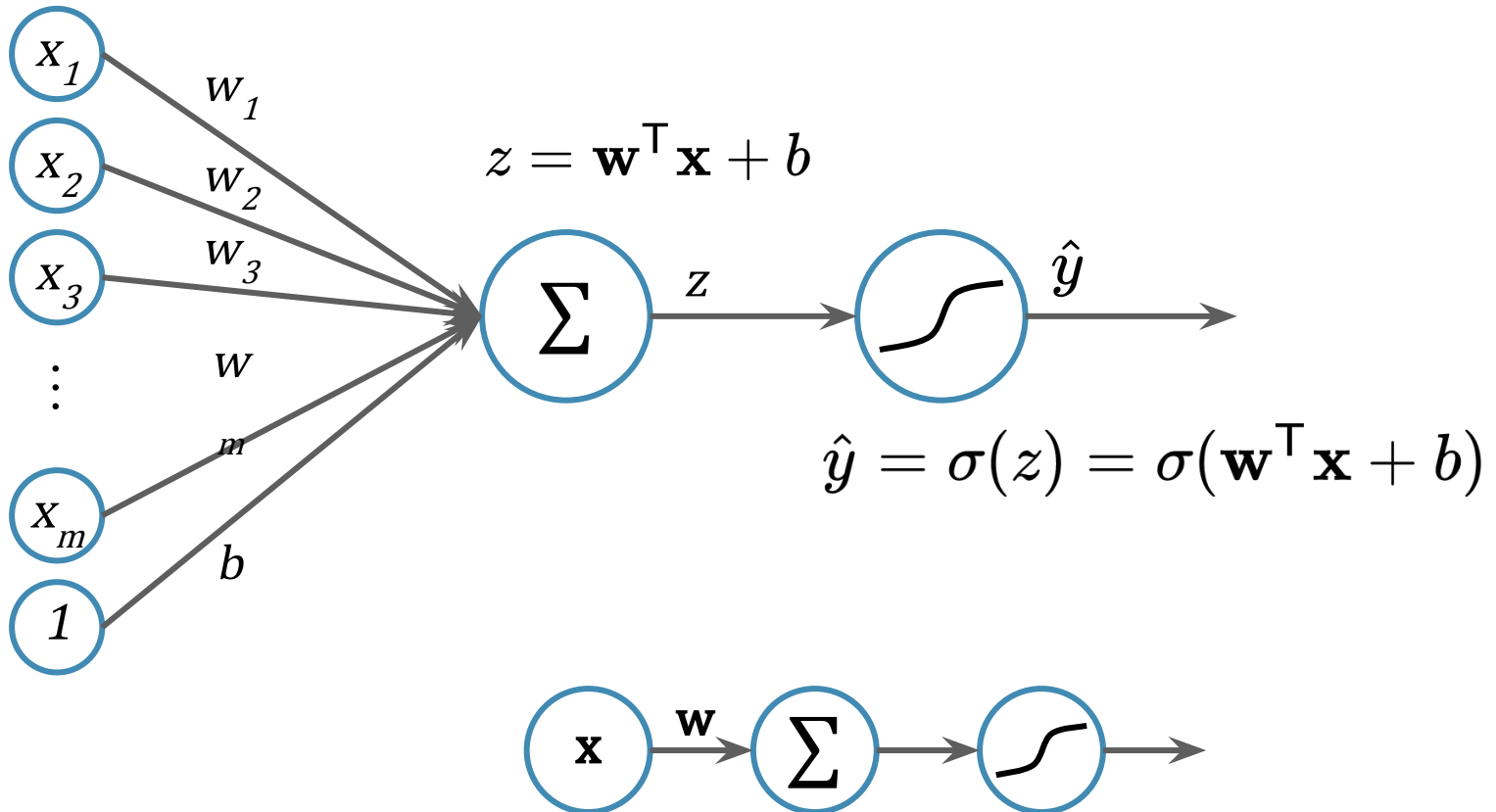
업데이트:

$$\mathbf{w} = \mathbf{w} - \alpha \frac{\partial cost(\mathbf{w}, b)}{\partial \mathbf{w}}$$

$$b = b - \alpha \frac{\partial cost(\mathbf{w}, b)}{\partial b}$$

Logistic Regression

가설함수: $H(x) = \frac{1}{1 + e^{-(\mathbf{w}^T \mathbf{x} + b)}} = \sigma(\mathbf{w}^T \mathbf{x} + b)$



다중 분류 문제 Multinomial Classification

종속 변수 y 가 두 가지 이상의 값을 가지는 경우의 회귀분석

활용

- 동물 사진 분류 (이 사진은 고양이? 멧멍이? 토끼?)



- 자율주행 (지금 신호는 직진? 정지? 좌회전?)



- 글자인식 (이 숫자는 무엇일까?)



다중 분류 문제 Multinomial Classification

종속 변수 y 가 두 가지 이상의 값을 가지는 경우의 회귀분석

Training Data

[1.2, 3.8, -1.4, ..., 4.1]	→	A
[3.2, -1.2, -0.2, ..., 2.1]	→	B
[2.8, -1.4, -0.3, ..., 2.3]	→	C
[1.2, 3.4, -1.5, ..., 4.2]	→	B
[4.2, 2.1, 2.8, ..., -0.5]	→	A
...		
[3.2, 2.2, 2.2, ..., -0.4]	→	A

Test

[1.3, 3.2, -1.5, ..., 4.1]	→	?
----------------------------	---	----------

다중 분류 문제 Multinomial Classification

이진 분류에서는 두 종류를 0과 1로 나타내면 됐는데, 3개 이상일 때는?
"One-hot encoding"

Training Data

[1.2, 3.8, -1.4, ..., 4.1]	→	A	[1, 0, 0]
[3.2, -1.2, -0.2, ..., 2.1]	→	B	[0, 1, 0]
[2.8, -1.4, -0.3, ..., 2.3]	→	C	[0, 0, 1]
[1.2, 3.4, -1.5, ..., 4.2]	→	B	[0, 1, 0]
[4.2, 2.1, 2.8, ..., -0.5]	→	A	[1, 0, 0]
...			
[3.2, 2.2, 2.2, ..., -0.4]	→	A	[1, 0, 0]

Test

[1.3, 3.2, -1.5, ..., 4.1]	→	?
----------------------------	---	---

이진 분류 → 다중 분류

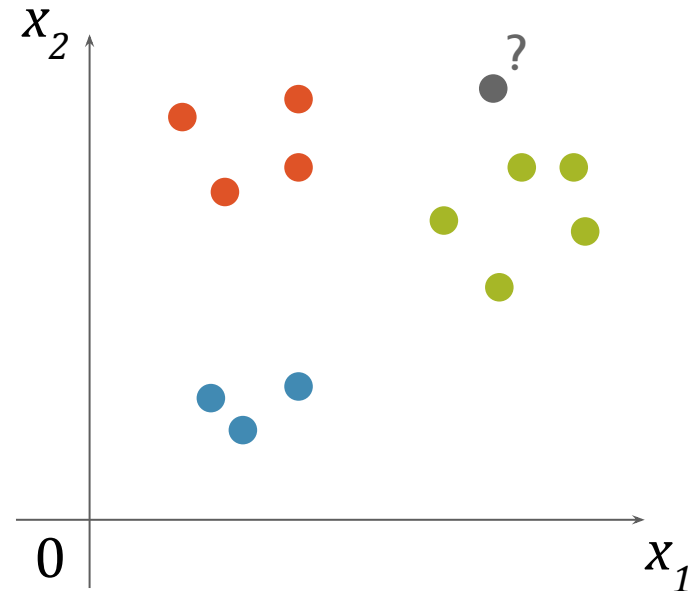
분류할 class가 k개 일 때? → **k 개의 이진 분류 수행!**

Training Data

[1.2, 3.8]	→	[1, 0, 0]
[3.2, -1.2]	→	[0, 1, 0]
[2.8, -1.4]	→	[0, 0, 1]
[1.2, 3.4]	→	[0, 1, 0]
[4.2, 2.1]	→	[1, 0, 0]
...		
[3.2, 2.2]	→	[1, 0, 0]

Test

[1.3, 3.2]	→	?
------------	---	---



이진 분류 → 다중 분류

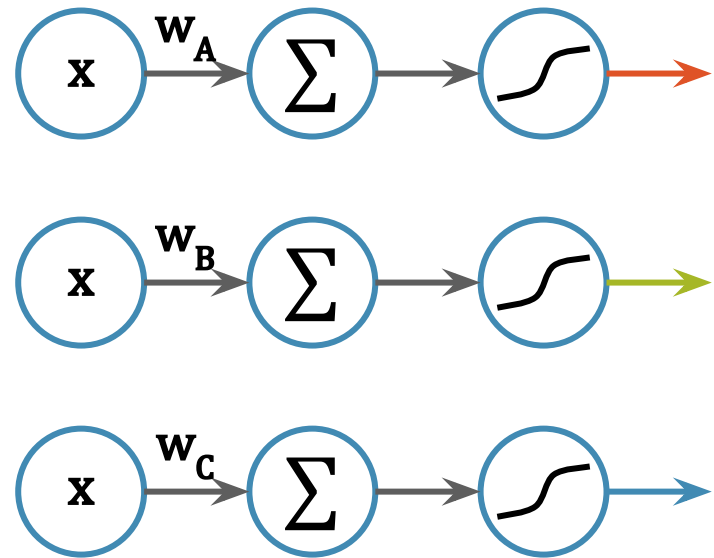
분류할 class가 k 개 일 때? → k 개의 이진 분류 수행!

Training Data

$[1.2, 3.8]$	→	$[1, 0, 0]$
$[3.2, -1.2]$	→	$[0, 1, 0]$
$[2.8, -1.4]$	→	$[0, 0, 1]$
$[1.2, 3.4]$	→	$[0, 1, 0]$
$[4.2, 2.1]$	→	$[1, 0, 0]$
...		
$[3.2, 2.2]$	→	$[1, 0, 0]$

Test

$[1.3, 3.2]$	→	?
--------------	---	---

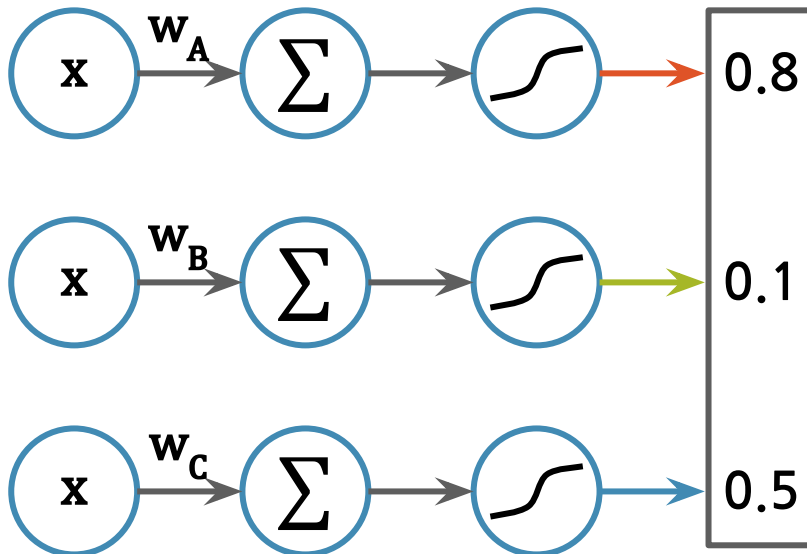


각각의 이진 분류에서 계산된 출력을
어떻게 해석해야 할까?

이진 분류 → 다중 분류

출력을 어떻게 해석해야 할까? → 확률로 해석

합이 1이
아닌데..
?



$\hat{y}?$

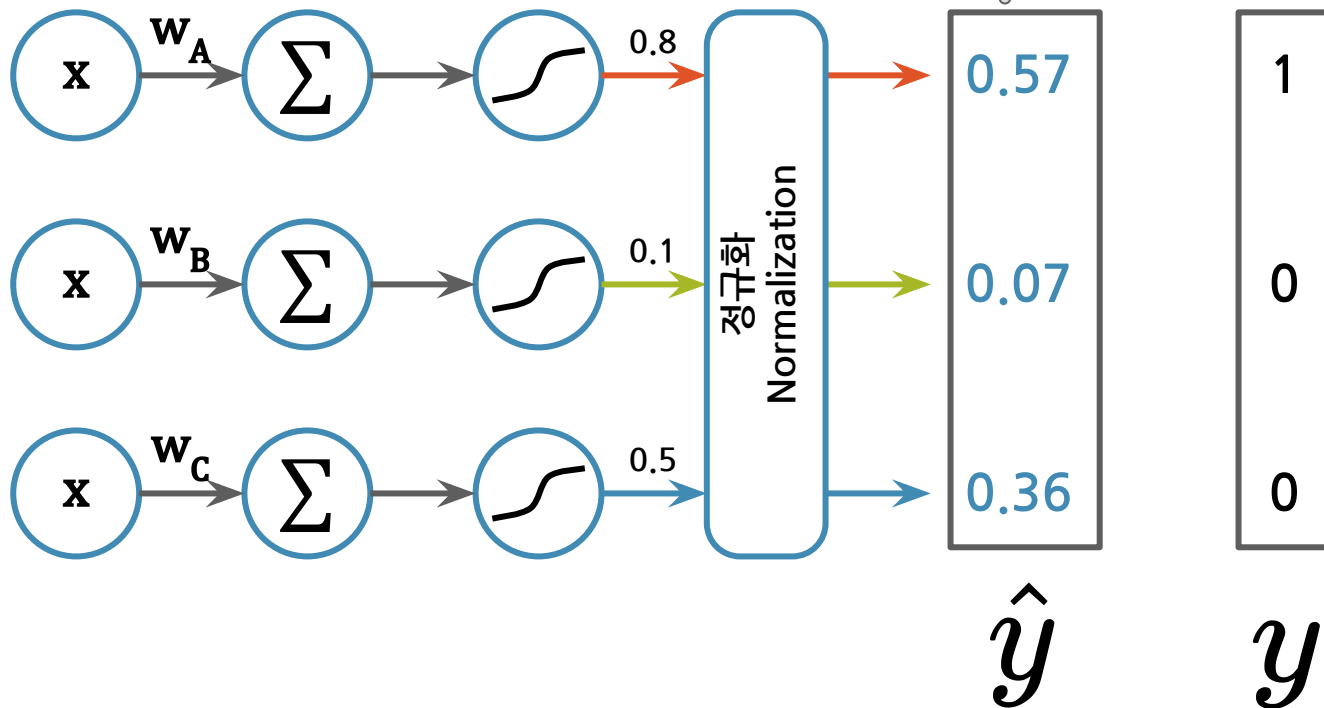


y

이진 분류 → 다중 분류

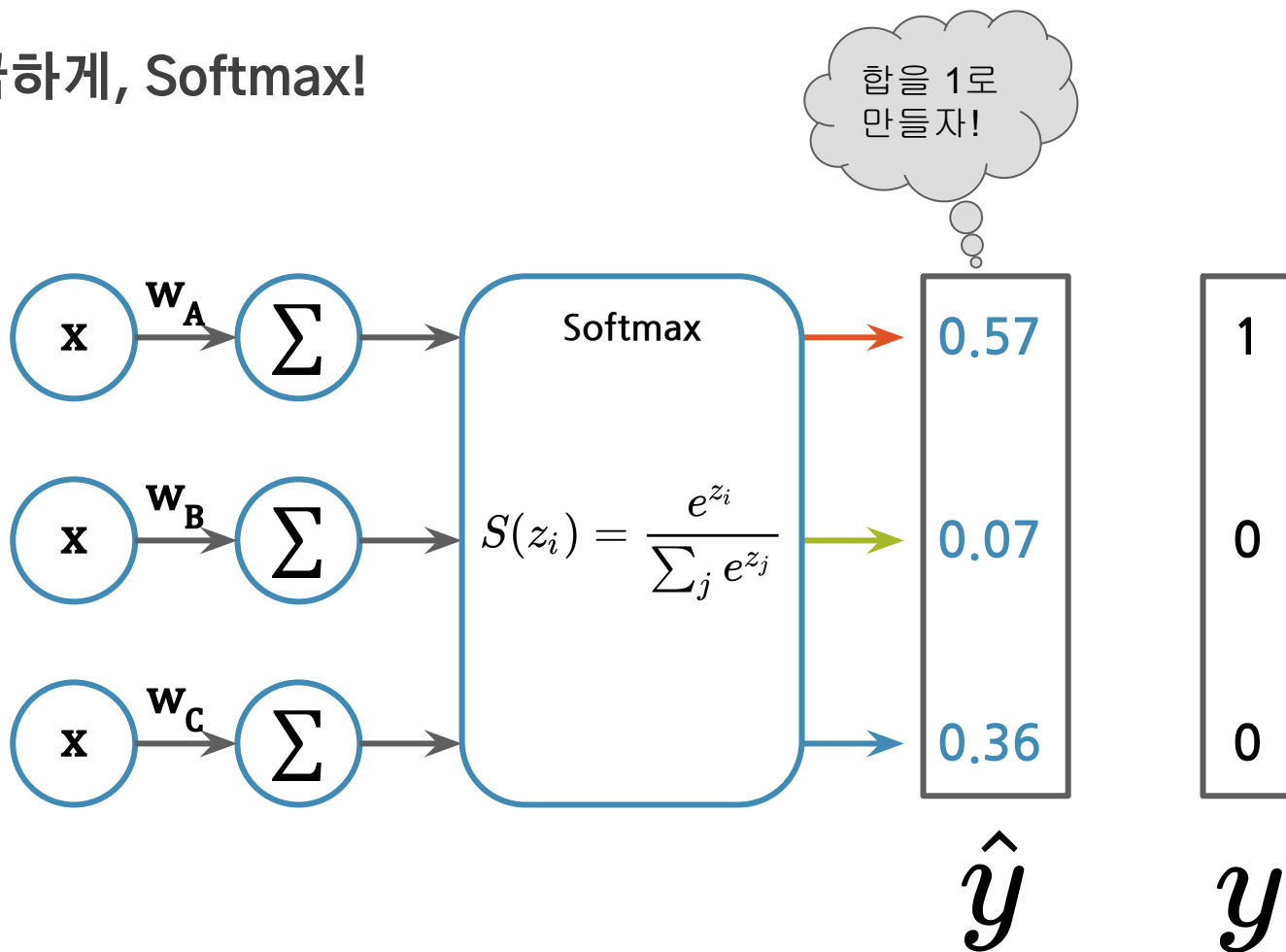
출력을 어떻게 해석해야 할까? → 확률로 해석

합을 1로
만들자!



이진 분류 → 다중 분류

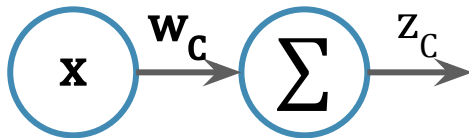
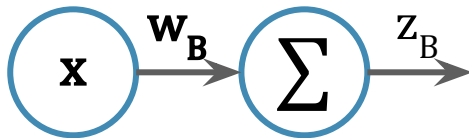
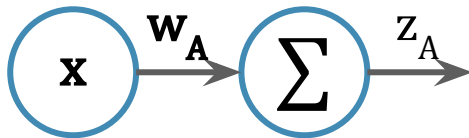
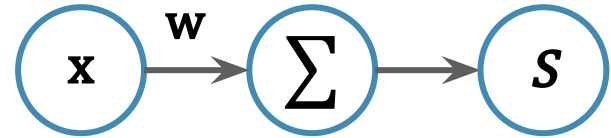
좀 더 깔끔하게, Softmax!



다중 분류 단순화

그래프 표현 → 행렬 표현

$$H(\mathbf{x}) = S(\mathbf{x}^T \mathbf{w} + b)$$



$$\mathbf{x} \times \mathbf{w}_A = z_A$$

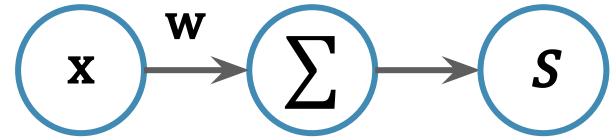
$$\mathbf{x} \times \mathbf{w}_B = z_B$$

$$\mathbf{x} \times \mathbf{w}_C = z_C$$

다중 분류 단순화

행렬로 간단하게

$$H(\mathbf{x}) = S(\mathbf{x}^T \mathbf{w} + b)$$



$$\mathbf{x} \times \mathbf{w}_A = z_A$$

$$\mathbf{x} \times \mathbf{w}_B = z_B$$

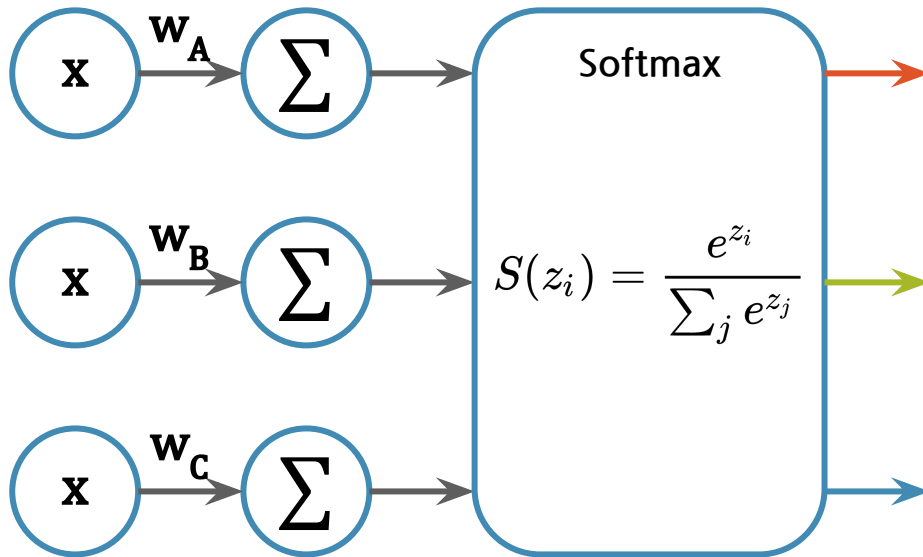
$$\mathbf{x} \times \mathbf{w}_C = z_C$$

$$\mathbf{x} \times [\mathbf{w}_A \quad \mathbf{w}_B \quad \mathbf{w}_C] = z_A, z_B, z_C$$

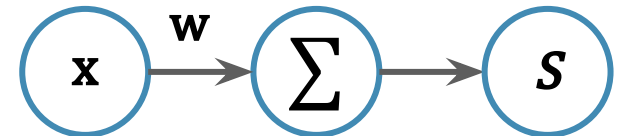
$$\mathbf{x} \times \mathbf{w} = \mathbf{z}$$

다중 분류 단순화

그래프 표현도 간단하게



$$H(\mathbf{x}) = S(\mathbf{x}^T \mathbf{w} + b)$$



비용: Cross Entropy

비용 함수: 현재의 가설이 얼마나 잘못되었는가

- 값이 작을수록 가설이 정확, 값이 클수록 가설이 잘못 됨

Training Data

[1.2, 3.8]	→	[1, 0, 0]
[3.2, -1.2]	→	[0, 1, 0]
[2.8, -1.4]	→	[0, 0, 1]
[1.2, 3.4]	→	[0, 1, 0]
[4.2, 2.1]	→	[1, 0, 0]
...		
[3.2, 2.2]	→	[1, 0, 0]

↑
입력 \mathbf{x}

↑
출력 (실제값) \mathbf{y}

$$H(\mathbf{x}) = S(\mathbf{x}^T \mathbf{w} + b)$$

[0.5, 0.2, 0.3]
[0.1, 0.8, 0.1]
[0.9, 0.1, 0.0]
[0.3, 0.4, 0.3]
[0.7, 0.2, 0.1]
...
[1.0, 0.0, 0.0]

↑
출력 (예측값) $H(\mathbf{x})$

$$C(y_1, H(x_1))$$

$$C(y_2, H(x_2))$$

$$C(y_3, H(x_3))$$

$$C(y_4, H(x_4))$$

$$C(y_5, H(x_5))$$

...

$$C(y_m, H(x_m))$$

$$Cost(\mathbf{w}, b) = \frac{1}{n} \sum_{i=1}^m C(y_i, H(x_i))$$

비용: Cross Entropy

$$C(y, \hat{y}) = - \sum_{j=1}^d y_j \log \hat{y}_j$$

$$H(\mathbf{x}) = S(\mathbf{x}^T \mathbf{w} + b)$$

→ [1, 0, 0]
→ [0, 1, 0]
→ [0, 0, 1]
→ [0, 1, 0]
→ [1, 0, 0]
...
→ [1, 0, 0]

[0.5, 0.2, 0.3]
[0.1, 0.8, 0.1]
[0.9, 0.0, 0.1]
[0.3, 0.4, 0.3]
[0.7, 0.2, 0.1]
...
[0.9, 0.1, 0.0]

$$C(y_1, H(x_1)) = -\log(0.5)$$

$$C(y_2, H(x_2)) = -\log(0.8)$$

$$C(y_3, H(x_3)) = -\log(0.1)$$

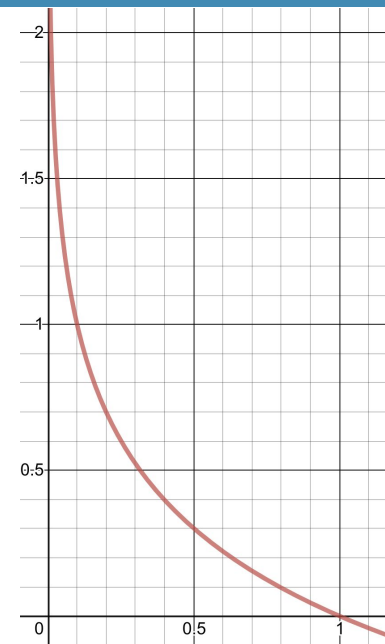
$$C(y_4, H(x_4)) = -\log(0.4)$$

$$C(y_5, H(x_5)) = -\log(0.7)$$

...

$$C(y_m, H(x_m)) = -\log(0.9)$$

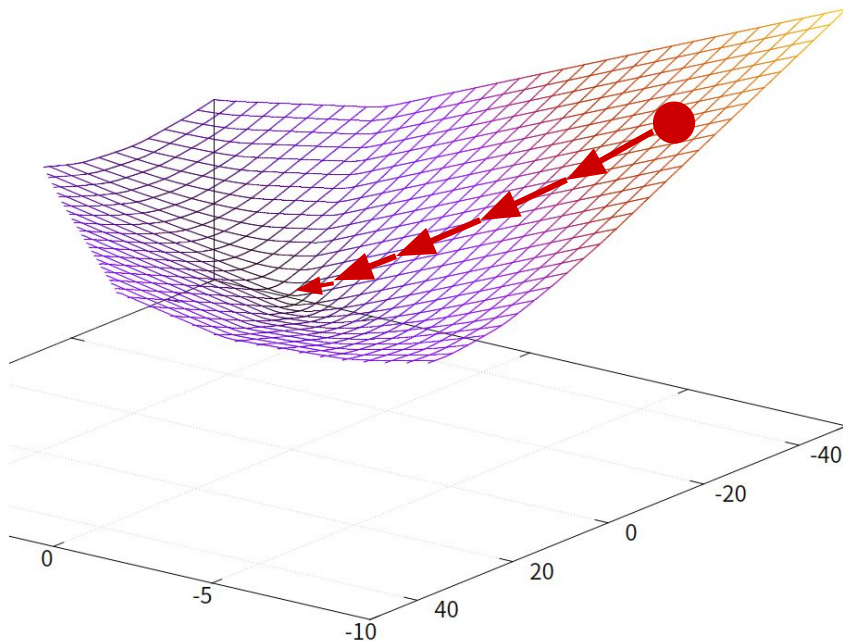
$$Cost(\mathbf{w}, b) = \sum_{i=1}^m C(y_i, H(x_i))$$



Gradient Descent

목표: 비용을 최소화 하자

$$\arg \min_{\mathbf{w}, b} cost(\mathbf{w}, b)$$



업데이트:

$$\mathbf{w} = \mathbf{w} - \alpha \frac{\partial cost(\mathbf{w}, b)}{\partial \mathbf{w}}$$

$$b = b - \alpha \frac{\partial cost(\mathbf{w}, b)}{\partial b}$$

Softmax Regression (2)

가설함수: $H(\mathbf{x}) = S(\mathbf{x}^T \mathbf{w} + b)$

비용: $Cost(\mathbf{w}, b) = \frac{1}{n} \sum_{i=1}^m C(y_i, H(x_i))$

$$C(y, \hat{y}) = - \sum_{j=1}^d y_j \log \hat{y}_j$$

업데이트:

$$\mathbf{w} = \mathbf{w} - \alpha \frac{\partial cost(\mathbf{w}, b)}{\partial \mathbf{w}}$$
$$b = b - \alpha \frac{\partial cost(\mathbf{w}, b)}{\partial b}$$

Questions?