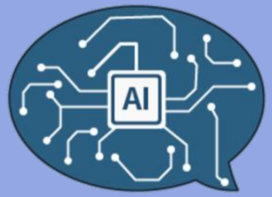
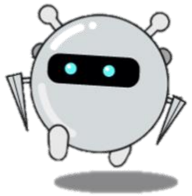


2장 - (4). 신경망



멘토 이현수

1. 신경망의 정의

<머신러닝 첫 단추 끼우기>
4.1.절

OUTTA

신경망의 정의

- Neural Network (NN)
- 입력(input) 데이터와 출력(output) 데이터 사이의 관계를 하나의 함수 관계로 연결해 표현해주는 인공지능 모델
  - 모델 : 관찰한 데이터로부터 예측과 결정을 얻어내는 수학적 함수

- 신경망을 사용할 경우, 충분한 관계가 표현 되는지 보장되는가?

→ Universal Approximation Thm. (교육자료 참고)

2. 신경망의 구조 (1)

<머신러닝 첫 단추 끼우기>

4.2.절

신경망의 구조

- 입력(input) 데이터와 출력(output) 데이터 사이의 관계를 하나의 함수 관계로 연결해 표현해주는 인공지능 모델
  - 모델 : 관찰한 데이터로부터 예측과 결정을 얻어내는 수학적 함수 $F(\mathbf{x})$

비선형 함수

$$F(\{v_i\},\{\mathbf{w}_i\},\{b_i\} : \mathbf{x}) = \sum_{i=1}^{N} v_i \varphi(\mathbf{x} \cdot \mathbf{w}_i + b_i)$$

m차원
실벡터

$$\boxed{(\underbrace{0, 1, 4, 5, \cdots 3}_{\text{m개}})}$$

신경망의 구조

- 입력(input) 데이터와 출력(output) 데이터 사이의 관계를 하나의 함수 관계로 연결해 표현해주는 인공지능 모델
  - 모델 : 관찰한 데이터로부터 예측과 결정을 얻어내는 수학적 함수 $F(\mathbf{x})$

$$\underbrace{F(\{v_i\},\{\mathbf{w}_i\},\{b_i\}}_{\text{N개의 원소로 이루어진 집합}} : \mathbf{x}) = \boxed{\sum_{i=1}^{N}} v_i \varphi(\mathbf{x}\cdot\mathbf{w}_i + b_i)$$

N개 스칼라의 합

N개의 원소로
이루어진 집합

$$\{v_1, v_2, \cdots v_N\}$$
$$\{\mathbf{w}_1, \mathbf{w}_2, \cdots \mathbf{w}_N\}$$

N개

$$\begin{aligned} = \quad & v_1\varphi(xw_1 + b_1) \\ & + v_2\varphi(xw_2 + b_2) \\ & + v_3\varphi(xw_3 + b_2) \\ & + \cdots \\ & + v_N\varphi(xw_N + b_5) \end{aligned}$$

신경망의 구조

- 입력(input) 데이터와 출력(output) 데이터 사이의 관계를 하나의 함수 관계로 연결해 표현해주는 인공지능 모델
  - 모델 : 관찰한 데이터로부터 예측과 결정을 얻어내는 수학적 함수

$$F(\{v_i\}, \{\mathbf{w}_i\}, \{b_i\} : \mathbf{x}) = \sum_{i=1}^{N} v_i \varphi(\mathbf{x} \cdot \mathbf{w}_i + b_i)$$

"Universal Approximation Theorem"

입력값(x)에 **적절한 상수와 벡터**,
그리고 **비선형 함수**를 이용해 연산을 진행하면
어떠한 **수학적 함수** F를 **함수** f**로 근사**시킬 수 있다.
(단, $f : R^m$의 닫힌 유계 부분집합 $\rightarrow R$ 이며 φ와 독립인 연속함수)

신경망의 구조

- 입력(input) 데이터와 출력(output) 데이터 사이의 관계를 하나의 함수 관계로 연결해 표현해주는 인공지능 모델
  - 모델 : 관찰한 데이터로부터 예측과 결정을 얻어내는 수학적 함수

"Universal Approximation Theorem"

Notation	Meaning
φ	Nonconstant, continuous, bounded, and monotonically increasing non-linear function
I_m	R^m 의 닫힌 유계 부분집합
$C(I_m)$	I_m 에서 $\mathbb{R}$ 로 가는 연속함수의 집합
f	$C(I_m)$ 에 속하면서 φ 와 독립인 임의의 함수

위 표의 조건을 만족하는 φ 에 대해,

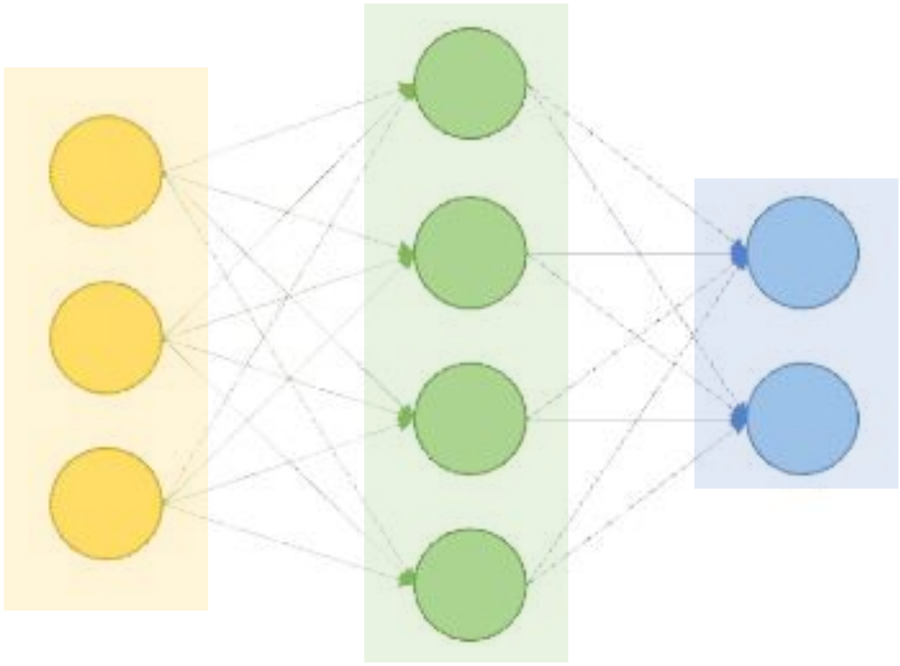
$$F(\{v_i\},\{\mathbf{w}_i\},\{b_i\};\mathbf{x}) = \sum_{i=1}^{N} v_i \varphi(\mathbf{x}\cdot\mathbf{w}_i + b_i)$$

와 같이 함수 F 를 정의하면, 역시 위 표의 조건을 만족하면서 정의된 임의의 함수 f 에 대해,

$$^{\forall}\epsilon > 0, ^{\exists}\{v_i\},\{\mathbf{w}_i\},\{b_i\}\ s.t. |F(\mathbf{x}) - f(\mathbf{x})| < \epsilon$$

신경망의 구조

- 입력(input) 데이터와 출력(output) 데이터 사이의 관계를 하나의 함수 관계로 연결해 표현해주는 인공지능 모델



신경망의 예시 1

- 일변수함수 F, 스칼라 x (1차원), $N = 5$

$$F(\{v_i\},\{w_i\},\{b_i\} : x) = \sum_{i=1}^{5} v_i \varphi(x \cdot w_i + b_i)$$

- 기존에 살펴본 수식과의 비교

5

$$F(\{v_i\},\{\mathbf{w}_i\},\{b_i\} : \mathbf{x}) = \sum_{i=1}^{N} v_i \varphi(\mathbf{x} \cdot \mathbf{w}_i + b_i)$$

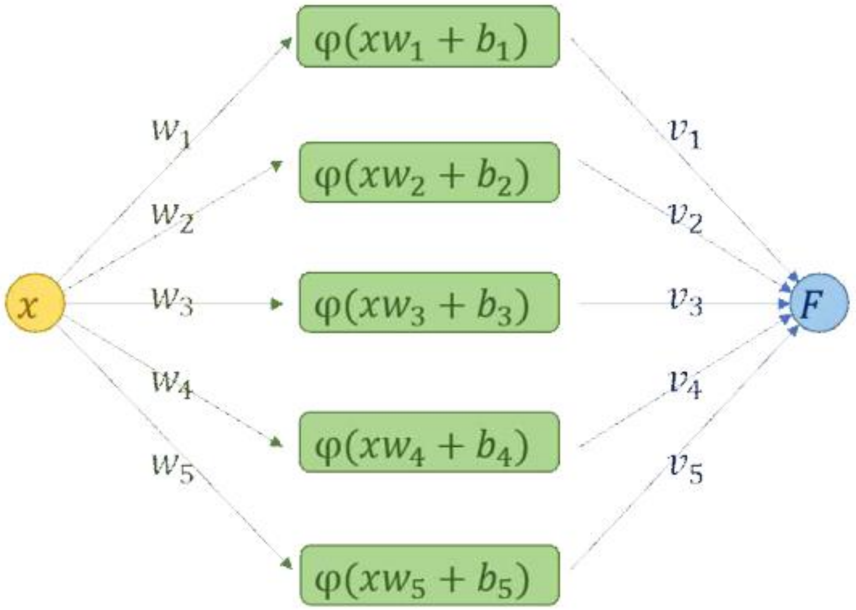
1차원
스칼라

신경망의 예시 1

- 일변수함수 F, 스칼라 x (1차원), $N = 5$

$$F(\{v_i\},\{w_i\},\{b_i\} : x) = \sum_{i=1}^{5} v_i\varphi(x \cdot w_i + b_i)$$

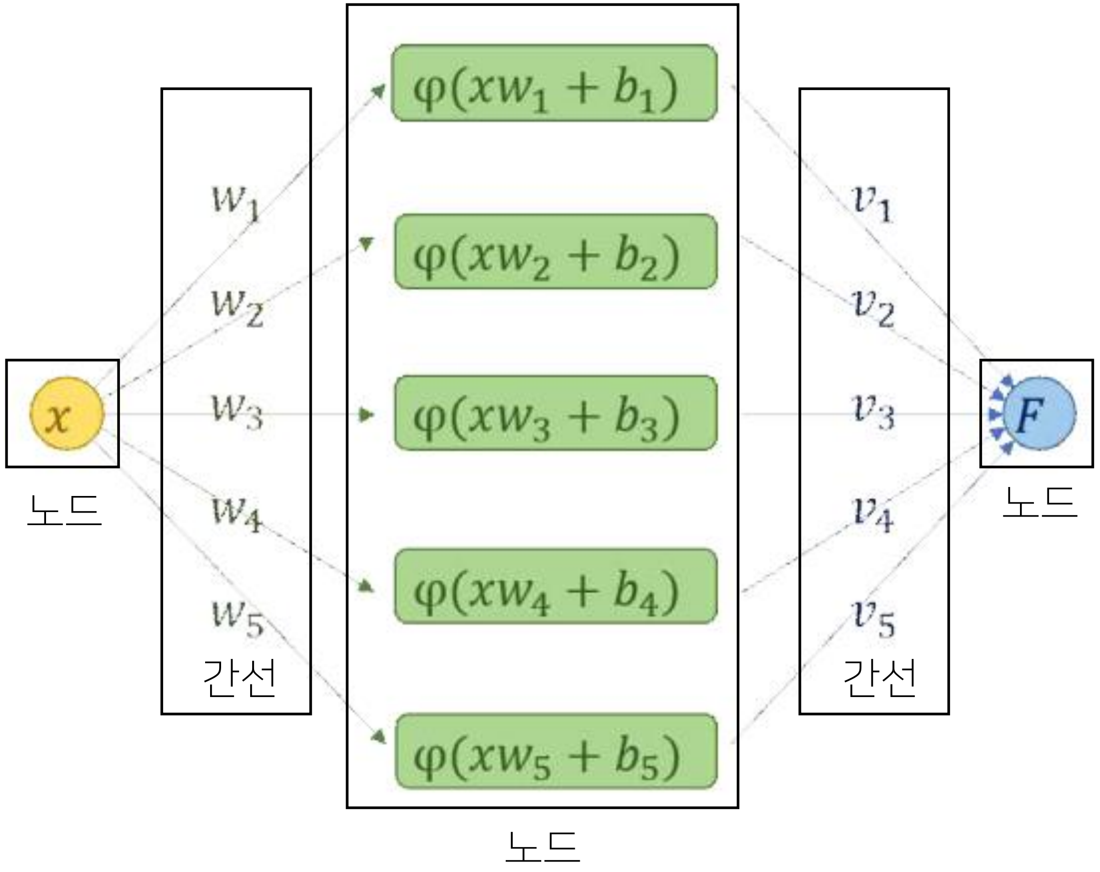
$$= v_1\varphi(xw_1 + b_1)$$
$$+ v_2\varphi(xw_2 + b_2)$$
$$+ \cdots$$
$$+ v_5\varphi(xw_5 + b_5)$$



"방향 그래프"

방향 그래프 : 각 간선에서 이동할 수 있는 방향이 정해져 있음

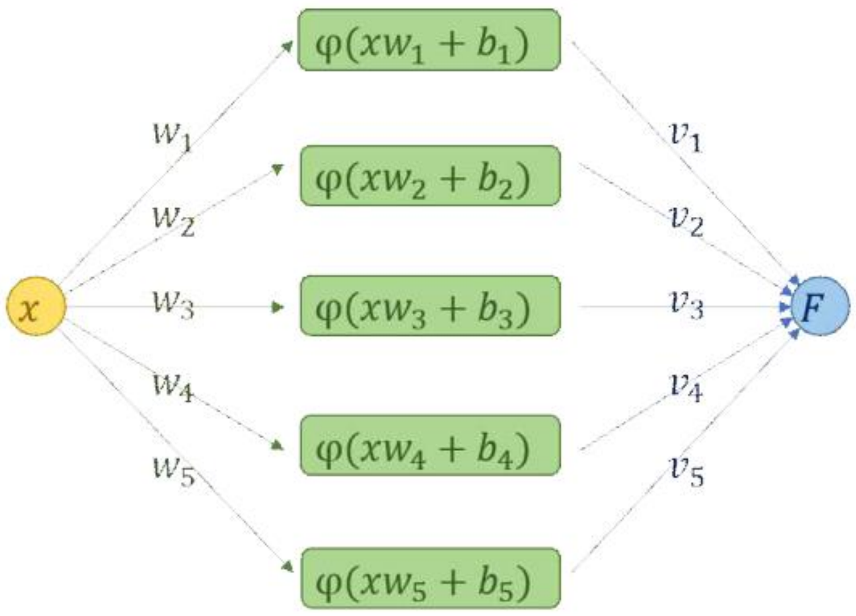
- 구성 요소
  1) 노드
  2) 간선



신경망의 예시 2

- 일변수함수 F, 스칼라 x (1차원), $N = 5$

$$F(\{v_i\},\{w_i\},\{b_i\} : x) = \sum_{i=1}^{5} v_i\varphi(x \cdot w_i + b_i)$$



Q) 그래프의 **가중치**와 **편향**은?

- 가중치 : w_i (입력값에 곱해주는 값)
- 편향 : b_i (함수 φ에서 더해주는 값)

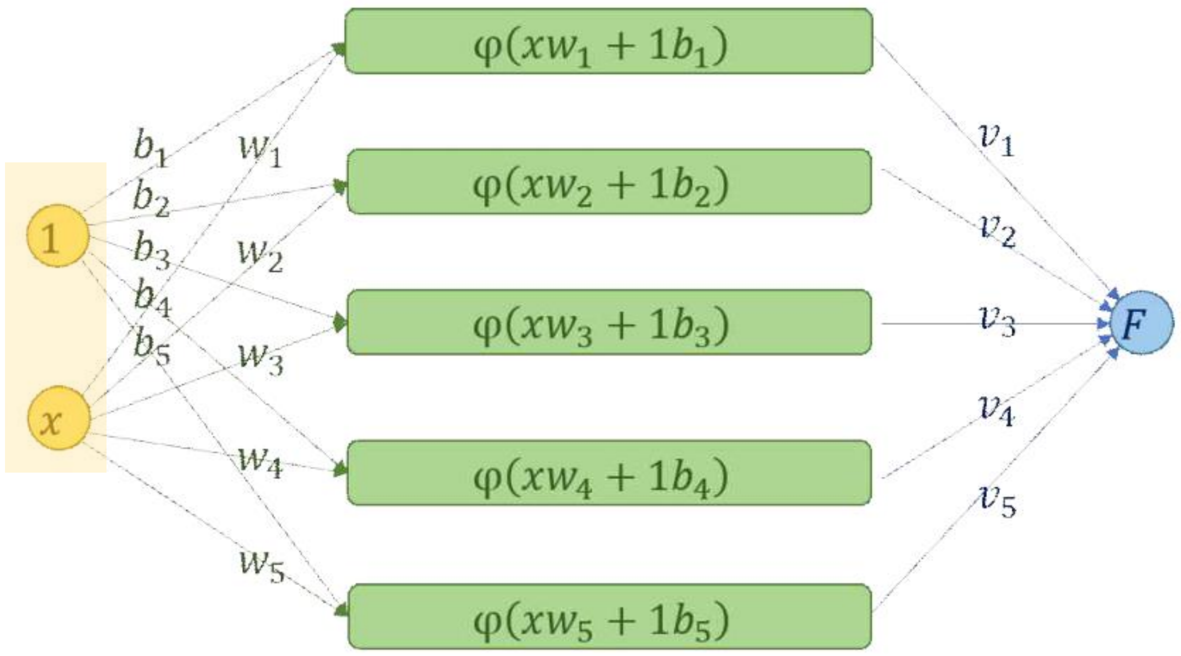
Q) 편향을 가중치 처럼 보는 방법?

신경망의 예시 2

- 일변수함수 F, 스칼라 x (1차원), $N = 5$

A) 편향을 가중치 처럼 보는 방법

$$F(\{v_i\},\{w_i\},\{b_i\} : x) = \sum_{i=1}^{5} v_i \varphi(x \cdot w_i + \mathbf{1} \cdot b_i)$$

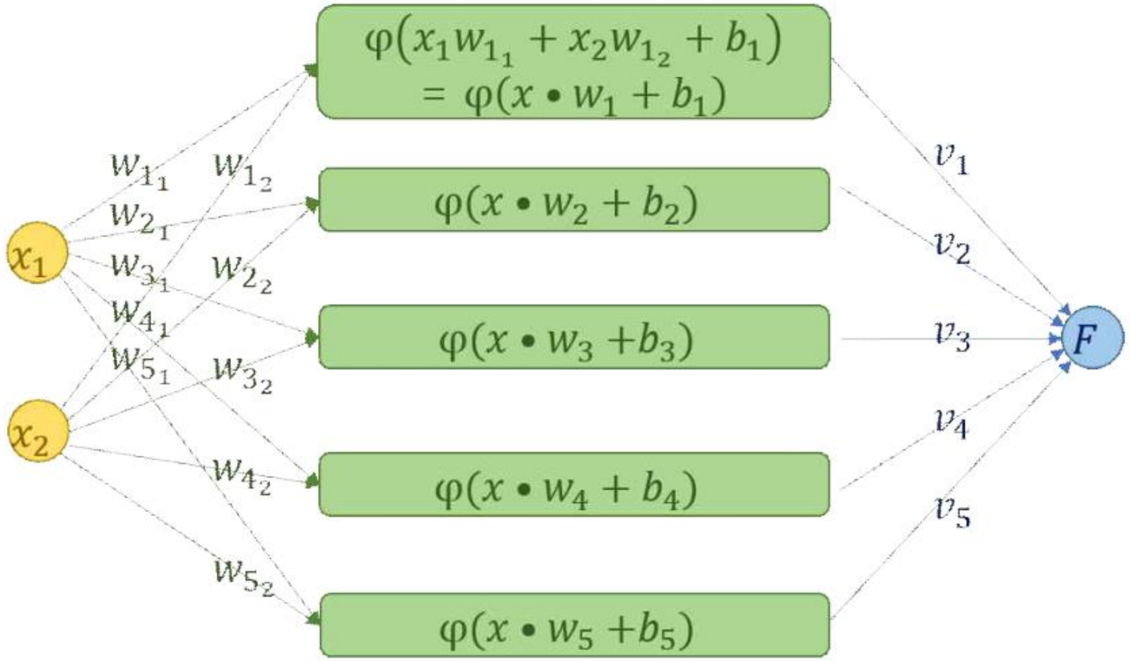


신경망의 예시 3

- 이변수함수 F, 벡터 $\mathbf{x}$ (2차원), $N = 5$

$$F(\{v_i\},\{\mathbf{w_i}\},\{b_i\} : \mathbf{x}) = \sum_{i=1}^{5} v_i \varphi(\mathbf{x} \cdot \mathbf{w}_i + b_i)$$

$$\mathbf{x} = (x_1, x_2)^T, \ \mathbf{w} = (w_1, w_2)^T$$

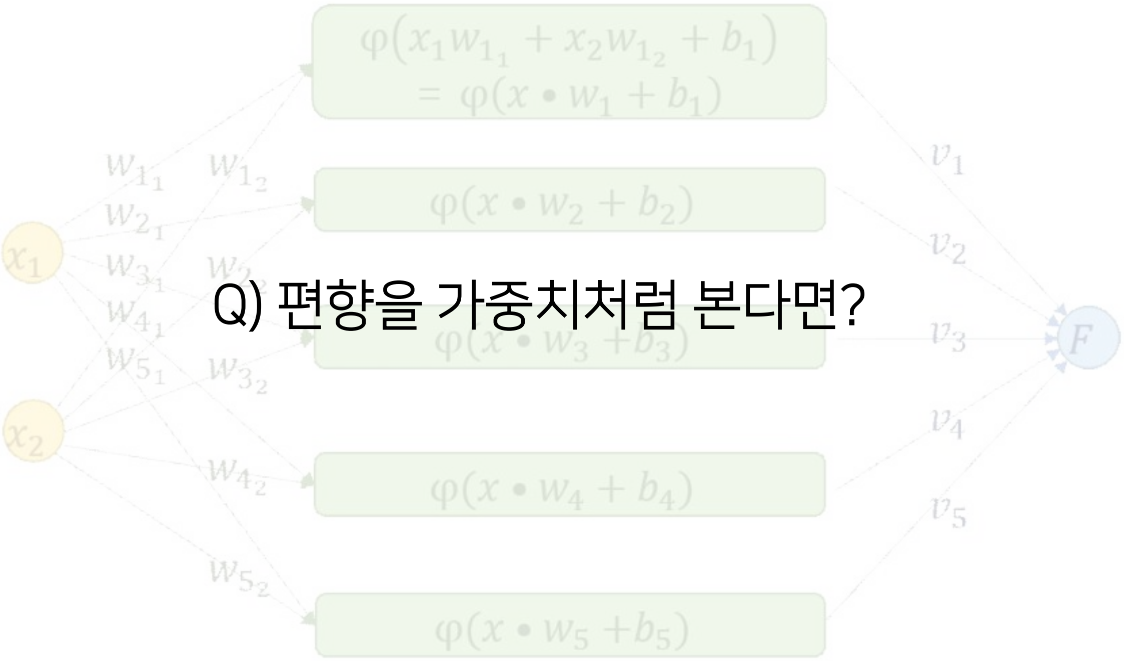


신경망의 예시 3

- 이변수함수 F, 벡터 $\mathbf{x}$ (2차원), $N = 5$

$$F(\{v_i\},\{\mathbf{w}_i\},\{b_i\} : \mathbf{x}) = \sum_{i=1}^{5} v_i \varphi(\mathbf{x} \cdot \mathbf{w}_i + b_i)$$

$$\mathbf{x} = (x_1, x_2)^T, \ \mathbf{w} = (w_1, w_2)^T$$



Q) 편향을 가중치처럼 본다면?

신경망의 정의

● **신경망의 구조**

신경망의 학습

순전파와 역전파 반복

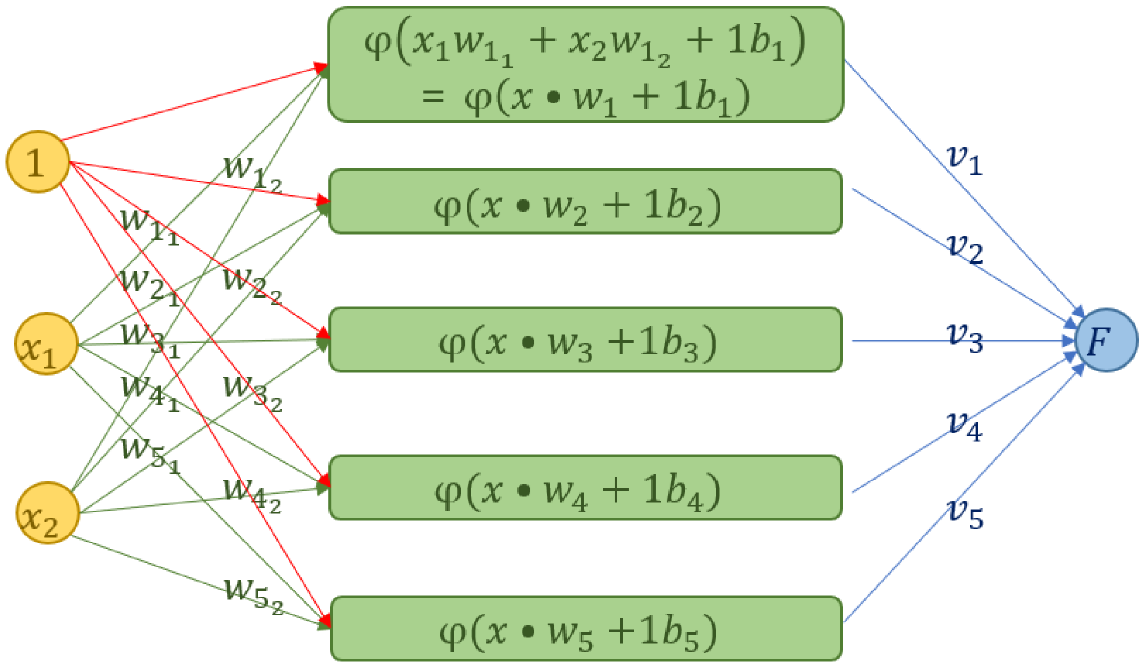
Hyper-parameter

신경망의 예시 4

- 이변수함수 F, 벡터 $\mathbf{x}$ (2차원), $N = 5$

$$F(\{v_i\},\{\mathbf{w}_i\},\{b_i\}:\mathbf{x}) = \sum_{i=1}^{5} v_i \varphi(\mathbf{x}\cdot\mathbf{w}_i + 1\cdot b_i)$$

$$\mathbf{x} = (x_1, x_2)^T, \ \mathbf{w} = (w_1, w_2)^T$$



OUTTA

3. 신경망의 구조 (2)

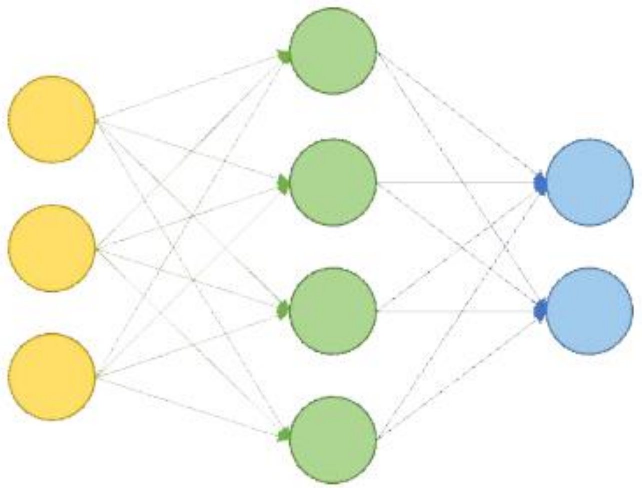
<머신러닝 첫 단추 끼우기>
4.2.절

신경망의 구조

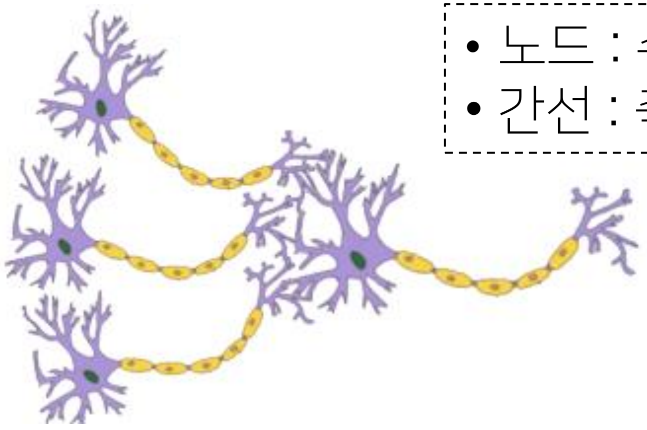
- 노드와 간선으로 이루어진, 방향 그래프

- 신경망 전체에 걸쳐 한 방향으로만 데이터가 전달됨
  : 뉴런의 신호전달 기작과 유사한 양상

- 신경망 vs. 뉴런



신경망 (Neural Network, NN)



• 노드 : 수상돌기
• 간선 : 축삭돌기

뉴런 (Neuron)

노드 = 뉴런

신경망의 정의

• **신경망의 구조**

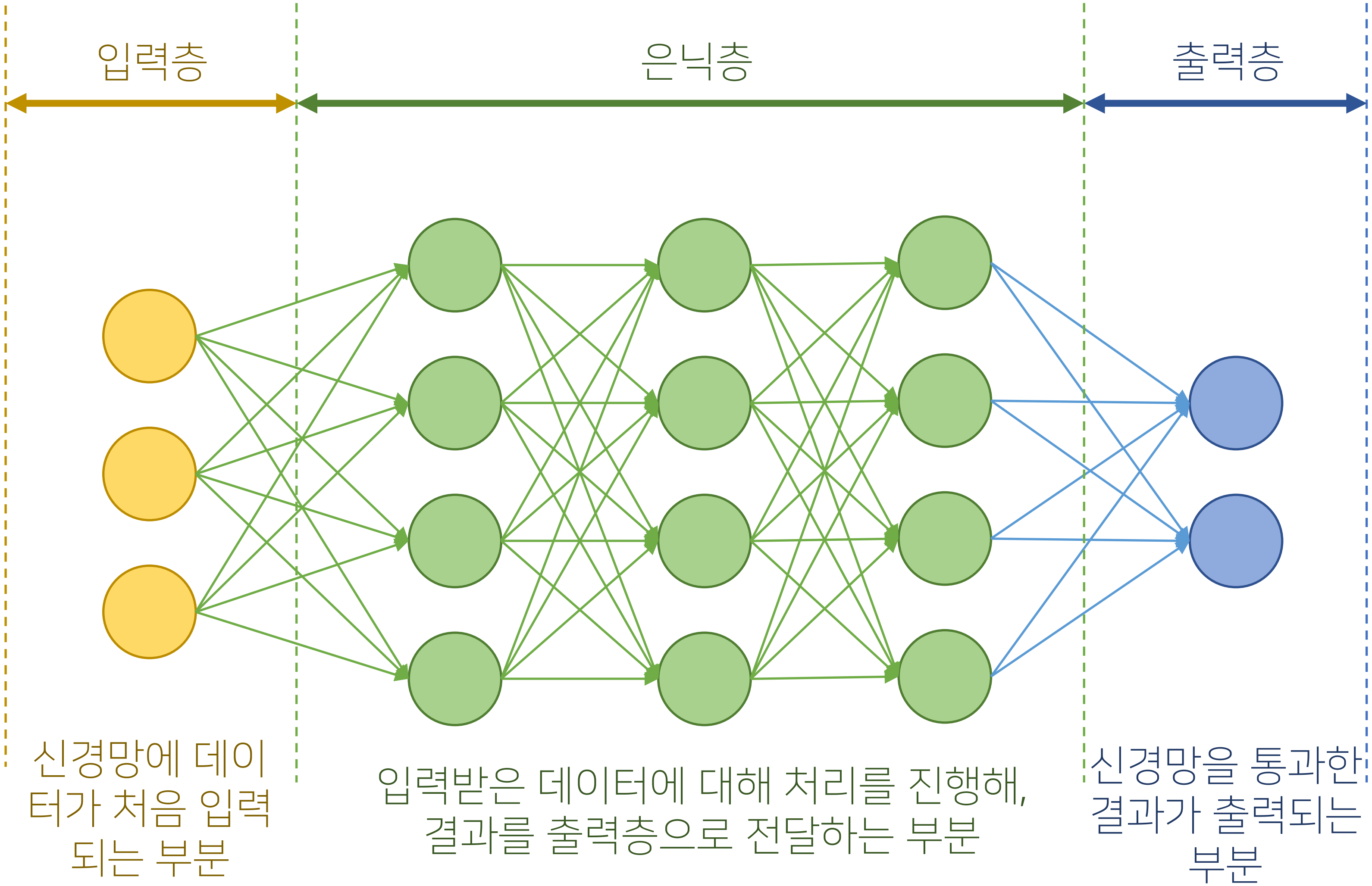
신경망의 학습

순전파와 역전파 반복

Hyper-parameter

계층 (Layer)

- 같은 선상에 있는 노드들의 집합



심층 신경망

- 은닉층의 개수가 2개 이상인 신경망

딥러닝 (Deep Learning, DL)

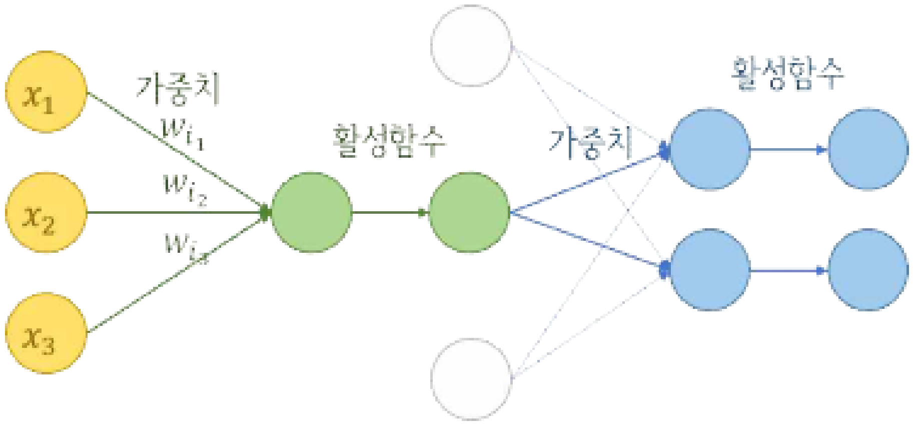
- 심층 신경망에 대한 머신러닝
- 딥러닝 ⊂ 머신러닝
- 딥러닝과 머신러닝의 구체적인 차이 → <머신러닝 첫 단추 끼우기> 참고

COUTA

각 뉴런의 역할 : 순전파

- 모든 뉴런 (= 노드) 이 공유하는 동일한 규칙
  1) 각 뉴런에 입력되는 값은 여러 개 가능, 그러나 **출력되는 값은 오직 하나**
  2) 각 뉴런에 입력되는 값에는 **가중치가 곱해짐**
  3) 출력되는 값은 **활성화 함수**를 통과함

- 순전파 : 데이터가 입력층, 은닉층, 출력층을 차례로 통과해 출력값이 나오는 과정



신경망의 정의

• **신경망의 구조**

신경망의 학습

순전파와 역전파 반복

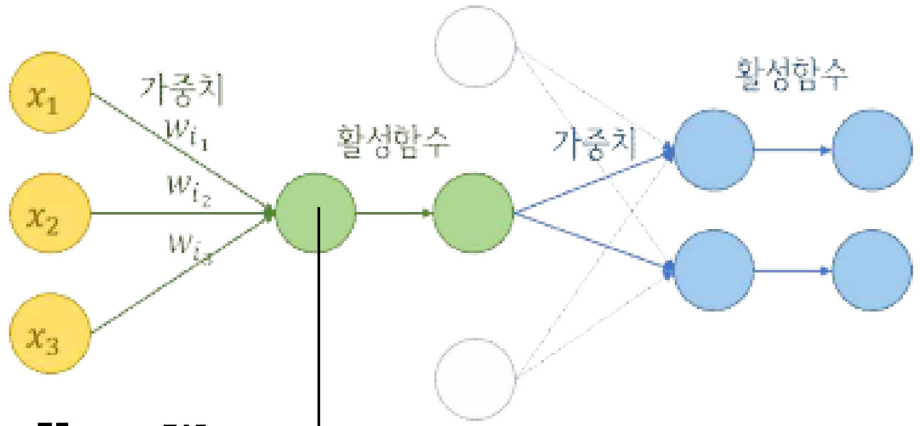
Hyper-parameter

가중치 곱

- 밀집층 (Dense Layer) or 완전연결 계층 (Fully-Connected Layer)

  : 모든 입력 데이터에 대해 가중치를 곱하는 뉴런들로 이루어진 계층

  : 각 층의 노드들끼리 완전하게 연결된 신경망의 Layer



$\mathbf{x}$ 1차원 벡터

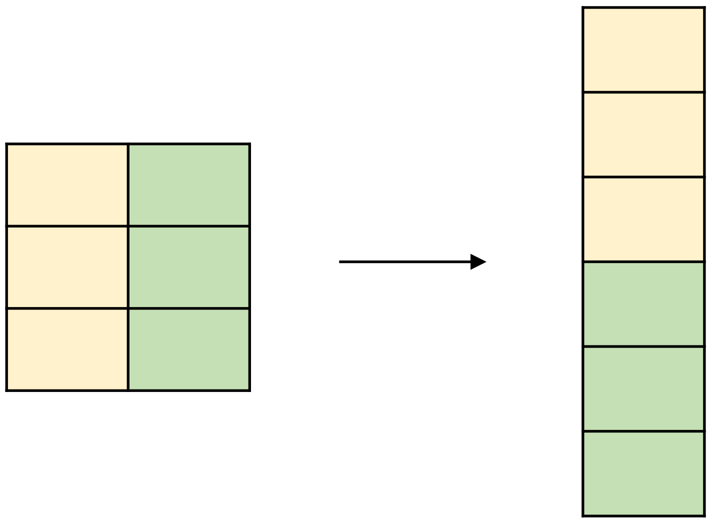
$\mathbf{w_i}$ 1차원 벡터

$\mathbf{x} \cdot \mathbf{w_i}$

가중치 곱

- 데이터가 2차원 이상인 경우?

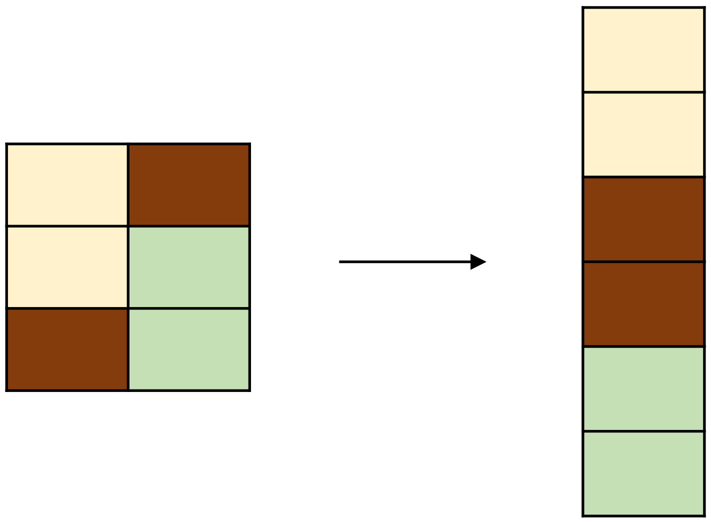
벡터화 (Flatten, Vectorization) : n차원 이상의 데이터를 1차원으로 만들어 주는 과정



가중치 곱

- 데이터가 2차원 이상인 경우?

벡터화 (Flatten, Vectorization) : n차원 이상의 데이터를 1차원으로 만들어 주는 과정



- 벡터화 과정의 문제점

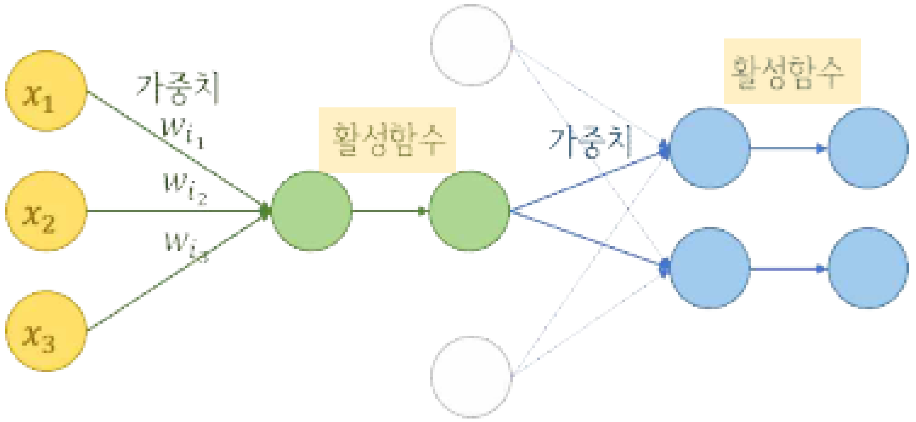
: 데이터의 **공간적인 정보가 무시**됨

: 해결 → 합성곱 신경망 (CNN, Convolution Neural Network)

활성화 함수 (Activation Function)

- 뉴런에서 최종적인 값을 내보내기 전에 통과 시켜주는 비선형 함수

$$F(\{v_i\},\{w_i\},\{b_i\} : x) = \sum_{i=1}^{5} v_i \varphi(x \cdot w_i + b_i)$$



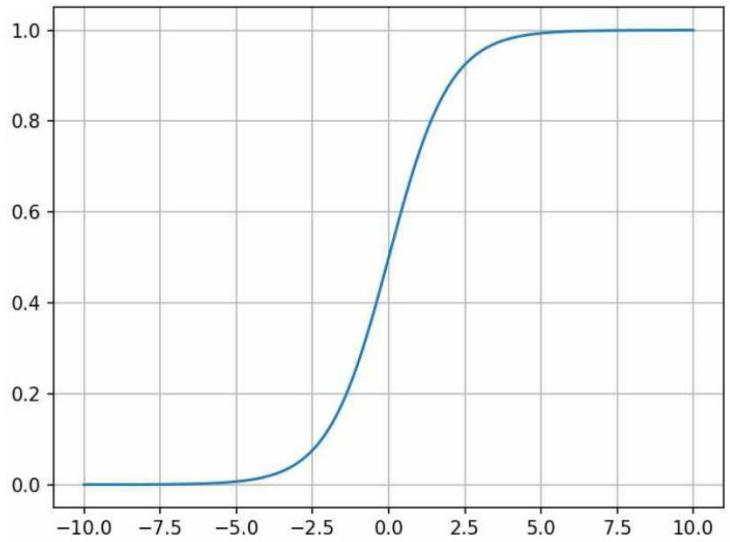
활성화 함수의 종류

1. 항등함수

- $\varphi(x) = x$
- 주로 출력층에서 사용
- **회귀 문제**에서 많이 사용

2. Sigmoid

- $\varphi(x) = \frac{1}{1+e^{-x}}$
- **분류 문제**에서 많이 사용



∵ 출력값이 0과 1사이의 값 → 해당 **Class에 속할 확률**로써 해석 가능!

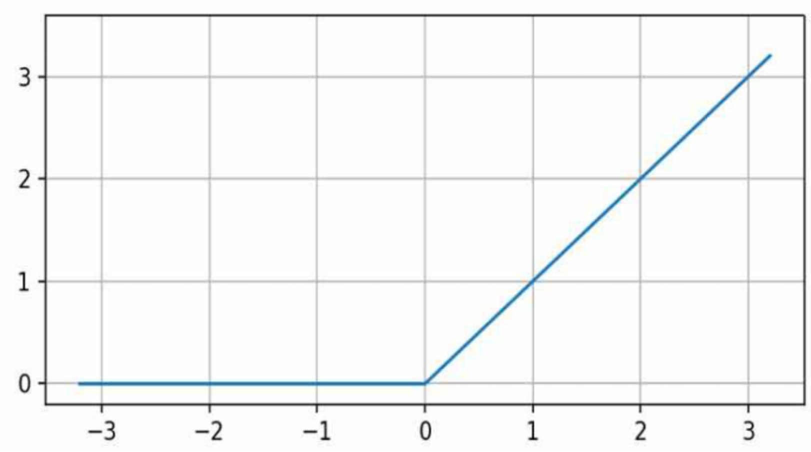
Ex) 1에 가까울 경우 class A로 분류, 0에 가까울 경우 class B로 분류

- DL에서는 잘 활용되지 않음 : Gradient Vanishing

활성화 함수의 종류

3. ReLU

- $\varphi(x) = \max(x, 0)$
- Gradient Vanishing 문제 **해결 가능**
- 여전히 문제점을 가짐 → Leaky ReLU



활성화 함수의 종류

4. Softmax

- 입력값이 $x_1, x_2, \cdots, x_n$, 출력값이 $y_1, y_2, \cdots, y_n$일 때

$$y_k = \text{softmax}(x_k) = \frac{e^{x_k}}{\sum_{i=1}^{n} e^{x_i}}$$

- $0 < \varphi(x_k) < 1$
- $\sum_{i=1}^{n} \varphi(x_k) = 1$
- 증가함수

→ 분류 문제에서 **각 class 에 속할 확률**로 해석 가능

다중 클래스 분류 문제에 많이 사용됨

활성화 함수의 종류

4. Softmax

- 입력값이 $x_1, x_2, \cdots, x_n$, 출력값이 $y_1, y_2, \cdots, y_n$일 때

$$y_k = \text{softmax}(x_k) = \frac{e^{x_k}}{\sum_{i=1}^{n} e^{x_i}}$$

x

1	2	1.5	0.6

↓

y

$\frac{e^{1}}{e^{1}+e^{2}+e^{1.5}+e^{0.6}}$	$\frac{e^{2}}{e^{1}+e^{2}+e^{1.5}+e^{0.6}}$	$\frac{e^{1.5}}{e^{1}+e^{2}+e^{1.5}+e^{0.6}}$	$\frac{e^{0.6}}{e^{1}+e^{2}+e^{1.5}+e^{0.6}}$

활성화 함수의 종류

4. Softmax

- 입력값이 $x_1, x_2, \cdots, x_n$, 출력값이 $y_1, y_2, \cdots, y_n$일 때

$$y_k = \text{softmax}(x_k) = \frac{e^{x_k}}{\sum_{i=1}^{n} e^{x_i}}$$

x

1	2	1.5	0.6

↓

y

$\frac{2.72}{16.41}$	$\frac{7.39}{16.41}$	$\frac{4.48}{16.41}$	$\frac{1.82}{16.41}$

활성화 함수의 종류

4. Softmax

- 입력값이 $x_1, x_2, \cdots, x_n$, 출력값이 $y_1, y_2, \cdots, y_n$일 때

$$y_k = \mathrm{softmax}(x_k) = \frac{e^{x_k}}{\sum_{i=1}^{n} e^{x_i}}$$

x

1	2	1.5	0.6

↓

y

0.166	**0.450**	**0.273**	**0.111**

활성화 함수의 종류

4. Softmax

- 입력값이 $x_1, x_2, \cdots, x_n$, 출력값이 $y_1, y_2, \cdots, y_n$ 일 때

$$y_k = \mathrm{softmax}(x_k) = \frac{e^{x_k}}{\sum_{i=1}^{n} e^{x_i}}$$

x

1	2	1.5	0.6

↓

y

0.166	**0.450**	**0.273**	**0.111**

$0.166 + 0.450 + 0.273 + 0.111 = 1$

소프트맥스 함수의 값은 '**확률**' 로써 해석 가능!

활성화 함수의 필요성

1) 활성화 함수를 곱해주지 않는다면, **선형 모델밖에 표현하지 못함.**

- 가중치를 곱해주는 과정은 선형 결합!

$$\mathbf{x} \cdot \mathbf{w_i} \longrightarrow (\mathbf{x} \cdot \mathbf{w_i}) \cdot \mathbf{w_j} \longrightarrow ((\mathbf{x} \cdot \mathbf{w_i}) \cdot \mathbf{w_j}) \cdot \mathbf{w_k}$$

- 비선형 함수를 추가해, **보다 일반적인 수학 모델** 표현가능
- **정교한 규칙성 모델링** 가능

2) 활성화 함수가 이용해야 **신경망의 층 수를 증가시키는 것이 의미있음**

- 선형 결합의 합성 = **선형 결합 → 하나의 층을 사용하는 것과 같은 효과**
- 활성화 함수를 사용할 경우, 층을 추가함에 따라 **비선형성이 같이 추가됨**

$$\mathbf{x} \cdot \mathbf{w_i} \longrightarrow f(\mathbf{x} \cdot \mathbf{w_i}) \cdot \mathbf{w_j} \longrightarrow f(f(\mathbf{x} \cdot \mathbf{w_i}) \cdot \mathbf{w_j}) \cdot \mathbf{w_k}$$

4. 신경망의 학습

<머신러닝 첫 단추 끼우기>
4.3.절

OUTTA

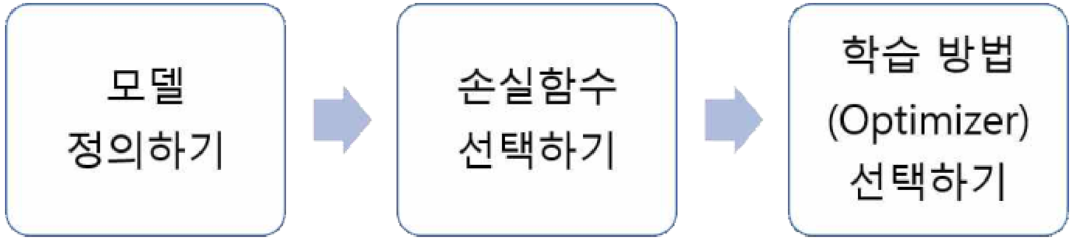
0. 데이터 전처리

1. 신경망 모델 구성

2. 손실함수 정의 및 계산

3. 손실함수 최적화 (학습)

- Backpropagation

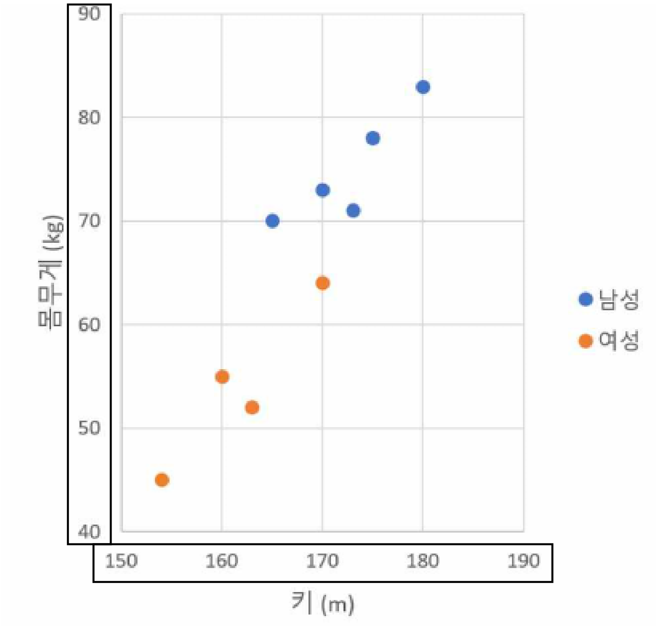


0. 데이터 전처리

1) 특성 스케일링 (Feature Scaling) or 정규화 (Normalization)

- Sample Data : Height and Weight

키 (m)	몸무게 (kg)	성별
180	83	남
170	73	남
175	78	남
170	64	여
163	52	여
154	45	여
165	70	남
160	55	여
173	71	남



키와 몸무게의 스케일이 매우 다름.

발생할 수 있는 문제점은?

신경망의 정의

신경망의 구조

- **신경망의 학습**

순전파와 역전파 반복

Hyper-parameter

0. 데이터 전처리

1) 특성 스케일링 (Feature Scaling) or 정규화 (Normalization)

- ML/DL 모델은 ‘**단위를 제외한 숫자**’ 만을 이용해 학습

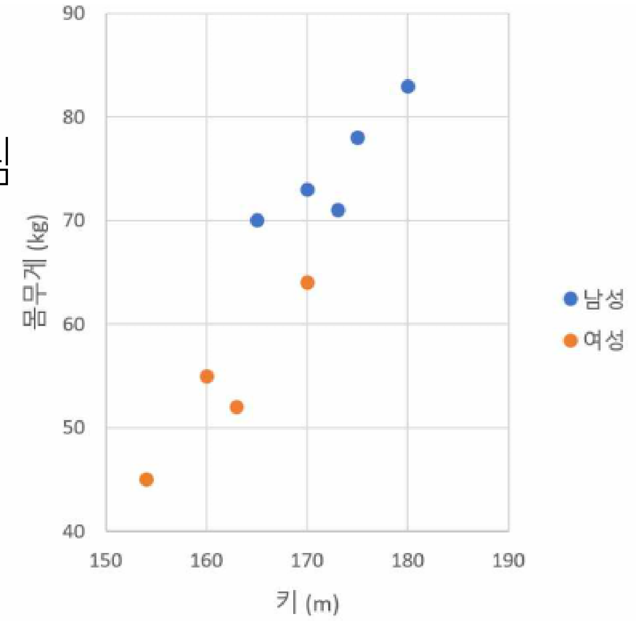
→ 키 값의 중요도 증가

→ ‘**몸무게**’ 에 대한 가중치 업데이트가 느리게 진행됨

(실습 4에서 확인해보실 수 있습니다)

→ $\mathcal{L} = \left(\hat{y} - (wx + b)\right)^2$

0.1배 10배



- 데이터의 **스케일을 맞춰 주는 과정**이 필요

OUTTA

0. 데이터 전처리

2-1) 해결방안 1 : 최소 – 최대 정규화

- 데이터의 최솟값 = 0 / 최댓값 = 1

$$x' = \frac{x - \min(Data)}{\max(Data) - \min(Data)}$$

- ex)

0	0	1	1	1	2	2	3	100

↓

0	0	0.01	0.01	0.01	0.02	0.02	0.03	1

0. 데이터 전처리

2-2) 해결방안 2 : Z-점수 정규화

- 표준화 (Standardization)

$$Z = \frac{X - \mu}{\sigma}$$

- 데이터 값의 **평균을 0**으로 만들어, 대칭적인 분포로 변환
- 최대-최소 정규화에 비해 **Outlier 의 영향이 적음**

(실습 4에서 확인해보실 수 있습니다)

1. 신경망 모델 구성

- 선형 회귀의 경우 : 수학적 함수 F(x)

$$\boldsymbol{F(m, b; x) = mx + b}$$

- 학습을 진행시킬 모델 : 신경망 모델

Layer 수 / 각 Layer의 종류 / Activation Function

신경망이 F(x)의 값에 가까운 값을 도출하도록 학습시킬 예정

2. 손실함수 정의 및 계산

- 주어진 문제에 적합한 오차함수를 찾기 전에, 먼저 여러 오차함수에 대해 알아보자.

1) SSE (Sum of Squares for Error, 오차제곱합)

$$\mathcal{L}_{SSE} = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$

Q) 모델의 출력값이 [0.1, 0.7, 0.2]이고, 실제 값이 [0, 1, 0]일 때 SSE는?

$$(0 - 0.1)^2 + (1 - 0.7)^2 + (0 - 0.2)^2 = 0.01 + 0.49 + 0.04 = 0.54$$

2) MSE (Mean Squared Error, 평균 제곱 오차)

$$\mathcal{L}_{MSE} = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 = \frac{\mathcal{L}_{SSE}}{n}$$

2. 손실함수 정의 및 계산

- 주어진 문제에 적합한 오차함수를 찾기 전에, 먼저 여러 오차함수에 대해 알아보자.

3) |Residuals| (오차 절댓값의 합)

- N개의 데이터셋
- True data points : $(x_i, \hat{y}_i),\ 0 \le i \le N-1$
- Expected data points : (x_i, y_i)
- 잔차(Residual) : $d_i = \hat{y}_i - y_i$

$$\mathcal{L}_{abs} = \sum_{i=1}^{n} |d_i|$$

4) CEE (Cross-Entropy Error, 교차 엔트로피 함수)

$$\mathcal{L}_{CEE} = -\sum_{i=1}^{n} y_i \log \hat{y}_i$$

2. 손실함수 정의 및 계산

- ex) 선형 회귀 문제에 적합한 오차함수는?

SSE (Sum of Squares for Error, 오차제곱합)

$$\mathcal{L}_{SSE} = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$

3. 손실함수 최적화 (학습)

- 손실함수를 최적화 시키는 방법? Optimizer

Optimizer

- 경사하강법 (GD)를 기반으로 손실함수를 최소화하는 모델 파라미터를 찾는 알고리즘
- 기존 GD의 단점 보완 (Local minimum에 수렴하는 등의 문제)
- Adam, SGD

Gradient Descent (Recap)

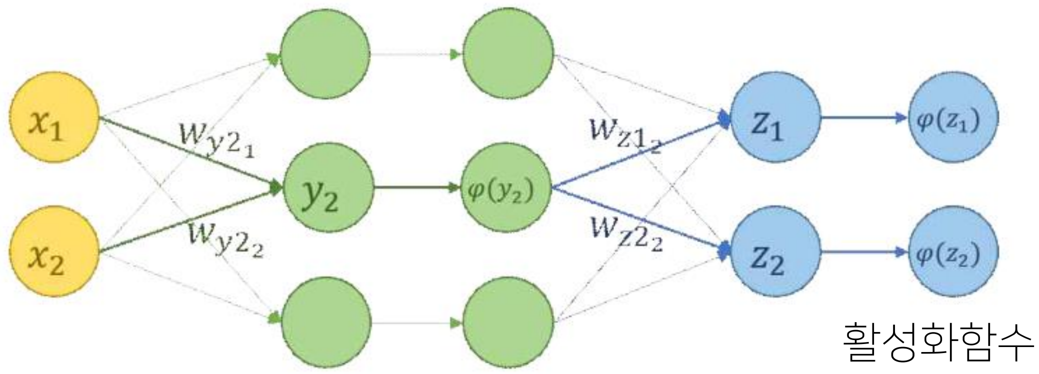
$$\boldsymbol{w}_{new} = \boldsymbol{w}_{old} - \delta \cdot \boldsymbol{\nabla}\mathcal{L}(\boldsymbol{w})\Big|_{\boldsymbol{w}=\boldsymbol{w}_{old}}$$

- 편미분 계수 계산 : 오차역전파법 (Backpropagation)

Backpropagation

순전파

입력층의 값으로부터
출력층의 값 계산



활성화함수 φ

역전파

손실함수의 값으로부터
가중치 (간선) 값 업데이트

How to Calculate $\nabla\mathcal{L}(\boldsymbol{w})\Big|_{\boldsymbol{w}=\boldsymbol{w}_{old}}$? Use CHAIN RULE

COUTA

Recap : Quiz 01

1. 다음 행렬곱 결과의 전치행렬을 구하여라.

$$\begin{pmatrix} 2 & 3 & 1 \\ -1 & 2 & 3 \end{pmatrix}\begin{pmatrix} 5 \\ -2 \\ 3 \end{pmatrix}$$

- 결과 행렬의 Size : 2행 1열

$$\begin{pmatrix} 2 \times 5 + 3 \times (-2) + 1 \times 3 \\ (-1) \times 5 + 2 \times (-2) + 3 \times 3 \end{pmatrix}$$

◉ (7,0)

○ (3,6)

○ (0,7)

Recap : Quiz 01

2. $f(x,y,z)=xy+yz$일 때, ∇f 는?

- ◉ $(y, x+z, y)$
- ○ $(y, x+z, xy)$
- ○ $(0, x, y)$

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right)$$

$$= (y, x+z, y)$$

Recap : Quiz 01

3. $L(x, y) = xy, f(a, b) = (ab, a + b) = (x, y)$ 일 때, $\frac{\partial L}{\partial a}$ 를 구하시오.

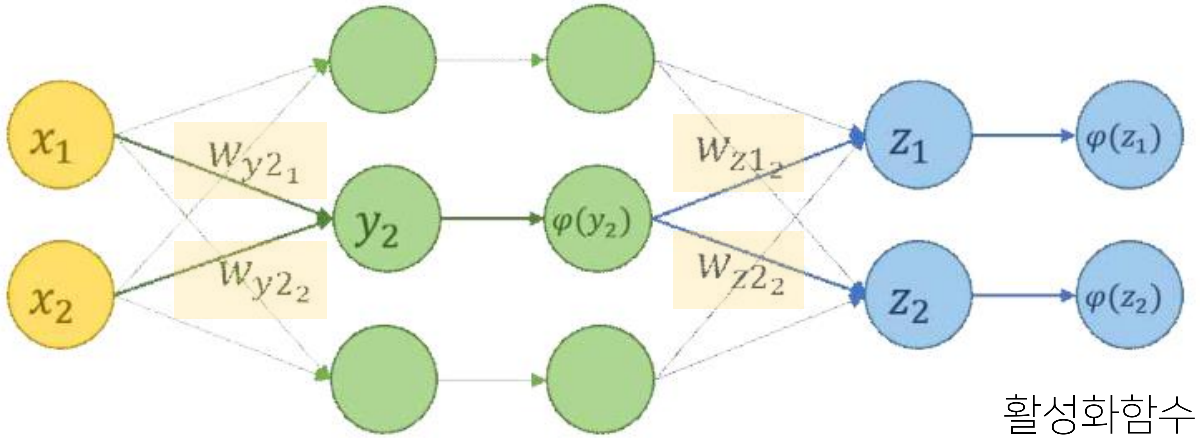
Use CHAIN RULE

$$\frac{\partial L}{\partial a} = \frac{\partial L}{\partial x} \cdot \frac{\partial x}{\partial a} + \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial a}$$

$$= y \cdot b + x \cdot 1$$

$$= yb + x$$

Backpropagation



활성화함수 φ

$$\boldsymbol{w}_{new} = \boldsymbol{w}_{old} - \delta \cdot \nabla \mathcal{L}(\boldsymbol{w}) \Big|_{\boldsymbol{w}=\boldsymbol{w}_{old}}$$

Loss Function $\mathcal{L}(\boldsymbol{w})$

신경망의 정의

신경망의 구조

● **신경망의 학습**

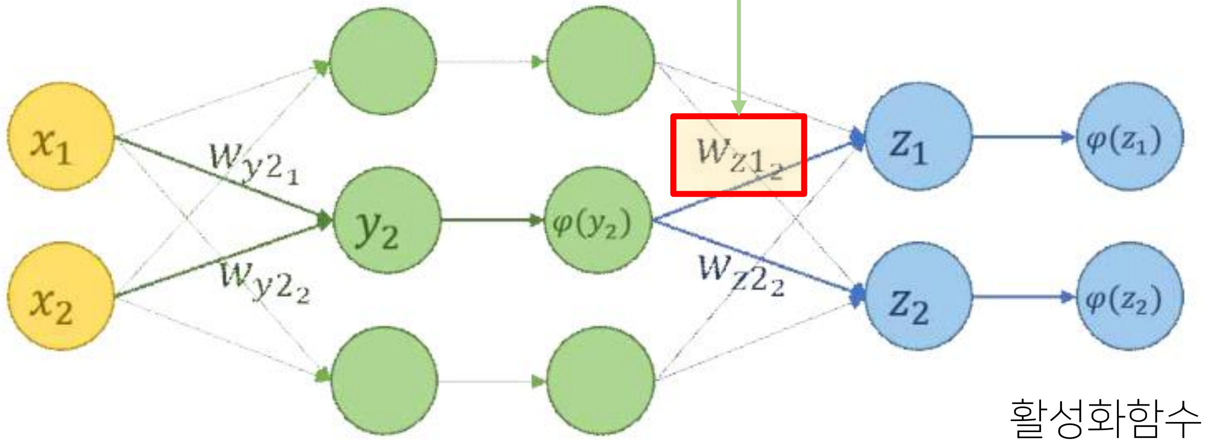
순전파와 역전파 반복

Hyper-parameter

Backpropagation

① $\boldsymbol{w_{z1_2}}$: $w_{z1_2\,new} = w_{z1_2\,old} - \delta \cdot \nabla\mathcal{L}(w)|_{w=w_{z1_2\,old}}$

$$\frac{\partial \mathcal{L}}{\partial w_{z1_2}} = \frac{\partial \mathcal{L}}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_{z1_2}}$$



활성화함수 φ

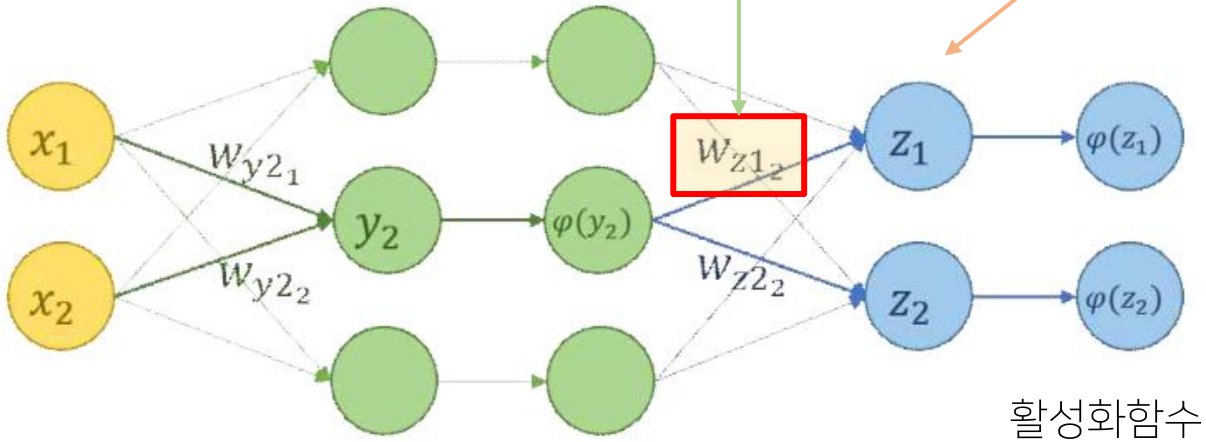
OUTTA

Backpropagation

① w_{z1_2} : $w_{z1_2\,new} = w_{z1_2\,old} - \delta \cdot \nabla\mathcal{L}(w)|_{w=w_{z1_2\,old}}$

$$\frac{\partial \mathcal{L}}{\partial w_{z1_2}} = \frac{\partial \mathcal{L}}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_{z1_2}}$$

$$\frac{\partial \mathcal{L}}{\partial z_1} = \frac{\partial \mathcal{L}}{\partial \varphi(z_1)} \cdot \frac{\partial \varphi(z_1)}{\partial z_1}$$



활성화함수 φ

신경망의 정의

신경망의 구조

- **신경망의 학습**

순전파와 역전파 반복

Hyper-parameter

Backpropagation

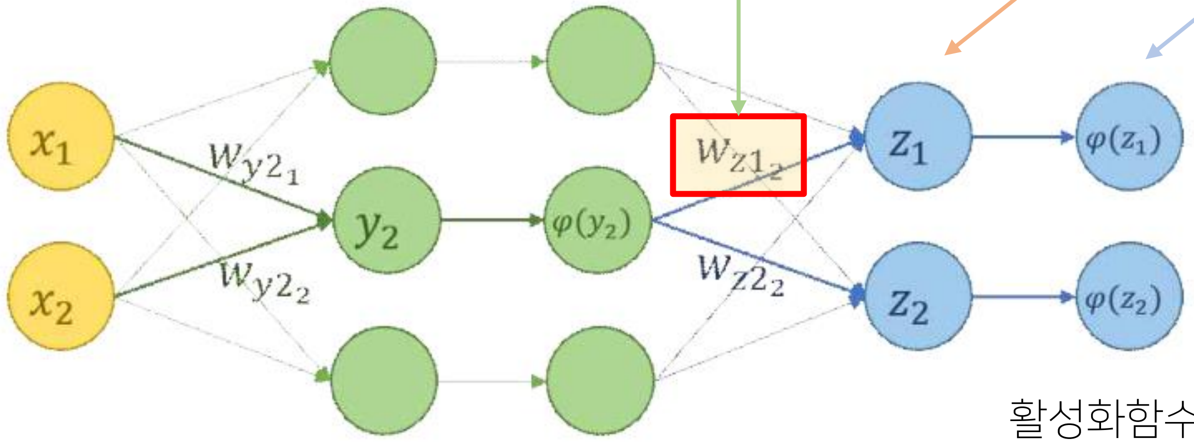
① w_{z1_2} : $w_{z1_2\,new} = w_{z1_2\,old} - \delta \cdot \nabla\mathcal{L}(w)|_{w=w_{z1_2\,old}}$

$$\frac{\partial \mathcal{L}}{\partial w_{z1_2}} = \frac{\partial \mathcal{L}}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_{z1_2}}$$

$$\frac{\partial \mathcal{L}}{\partial z_1} = \frac{\partial \mathcal{L}}{\partial \varphi(z_1)} \cdot \frac{\partial \varphi(z_1)}{\partial z_1}$$

$$\frac{\partial \mathcal{L}}{\partial \varphi(z_1)}$$

($\mathcal{L}$ 은 $\varphi(\cdot)$의 함수)



활성화함수 φ

Backpropagation

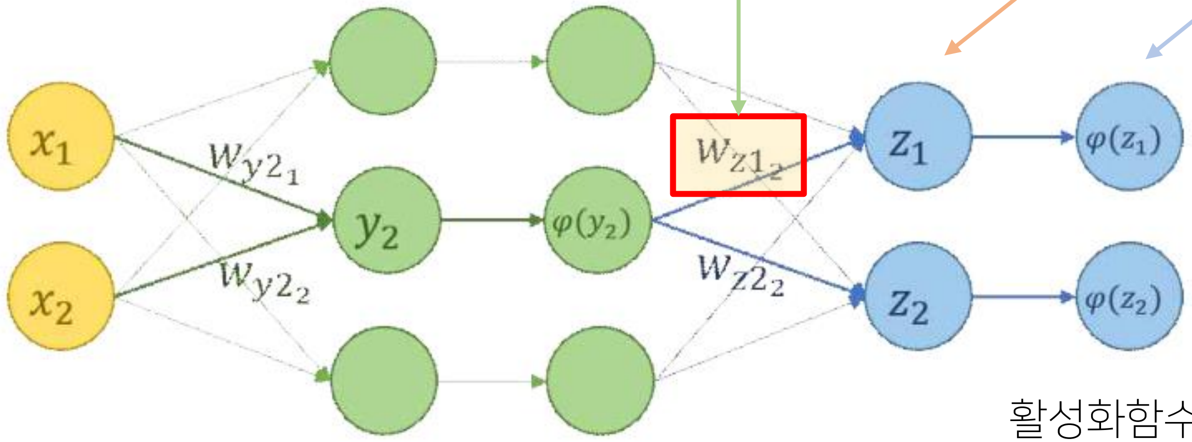
① w_{z1_2} : $w_{z1_2\,new} = w_{z1_2\,old} - \delta \cdot \nabla\mathcal{L}(w)|_{w=w_{z1_2\,old}}$

$$\frac{\partial\mathcal{L}}{\partial w_{z1_2}} = \frac{\partial\mathcal{L}}{\partial z_1}\cdot\varphi(y_2)$$

$$\frac{\partial\mathcal{L}}{\partial z_1} = \frac{\partial\mathcal{L}}{\partial\varphi(z_1)}\cdot\frac{\partial\varphi(z_1)}{\partial z_1}$$

$$z_1 = \varphi(y_1)w_{z1_1} + \varphi(y_2)w_{z1_2} + \varphi(y_3)w_{z1_3}$$

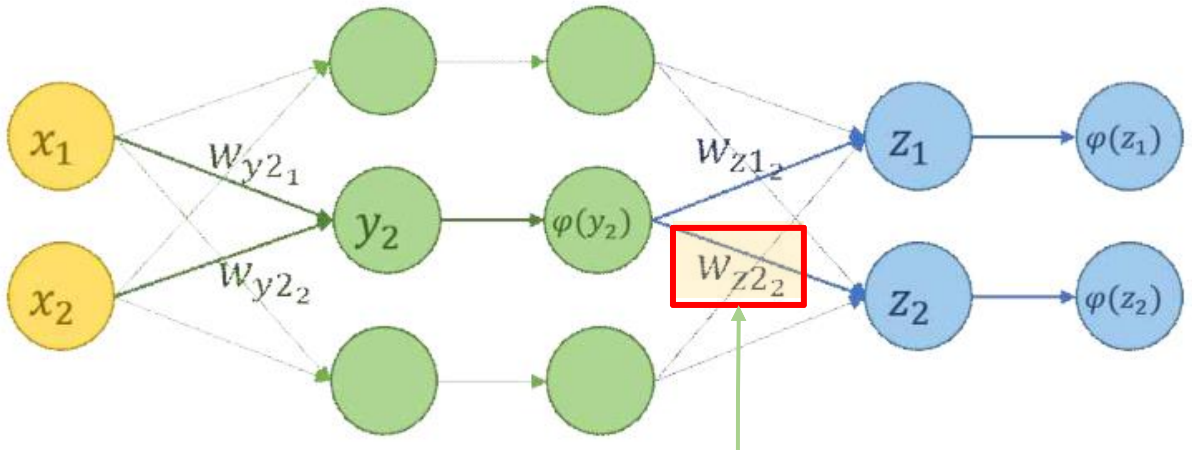
$$\frac{\partial\mathcal{L}}{\partial\varphi(z_1)}$$ ($\mathcal{L}$ 은 $\varphi(\cdot)$의 함수)



활성화함수 φ

Backpropagation

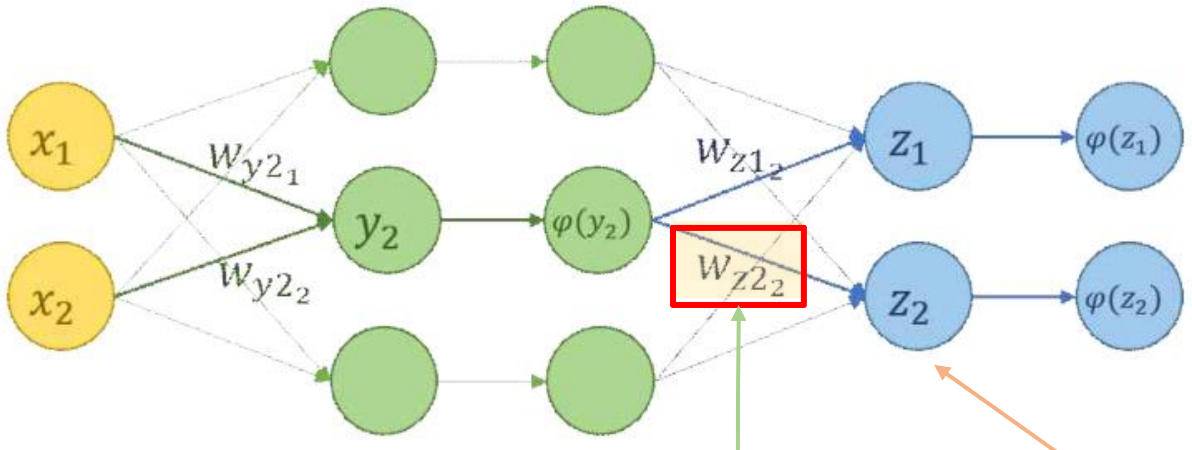
② $\boldsymbol{w_{z2_2}}$: $w_{z2_2\,new} = w_{z2_2\,old} - \delta \cdot \nabla\mathcal{L}(w)|_{w=w_{z2_2\,old}}$



$$\frac{\partial\mathcal{L}}{\partial w_{z2_2}} = \frac{\partial\mathcal{L}}{\partial z_2} \cdot \frac{\partial z_2}{\partial w_{z2_2}}$$

Backpropagation

② w_{z2_2} : $w_{z2_2\,new} = w_{z2_2\,old} - \delta \cdot \nabla\mathcal{L}(w)|_{w=w_{z2_2\,old}}$

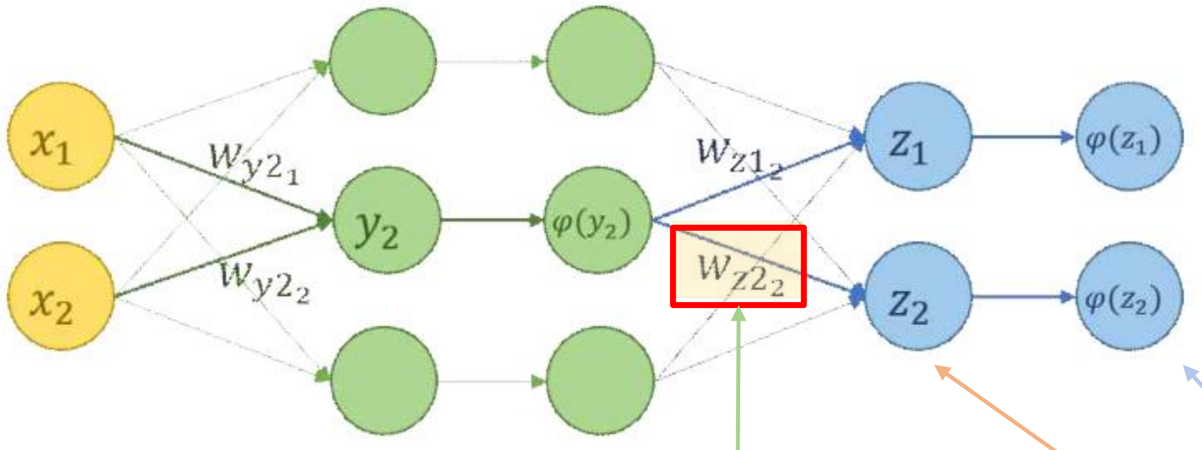


$$\frac{\partial\mathcal{L}}{\partial w_{z2_2}} = \frac{\partial\mathcal{L}}{\partial z_2} \cdot \frac{\partial z_2}{\partial w_{z2_2}}$$

$$\frac{\partial\mathcal{L}}{\partial z_2} = \frac{\partial\mathcal{L}}{\partial \varphi(z_2)} \cdot \frac{\partial \varphi(z_2)}{\partial z_2}$$

Backpropagation

② w_{z2_2} : $w_{z2_2\,new} = w_{z2_2\,old} - \delta \cdot \nabla\mathcal{L}(w)|_{w=w_{z2_2\,old}}$



$$\frac{\partial\mathcal{L}}{\partial\varphi(z_2)}$$ ($\mathcal{L}$ 은 $\varphi(\cdot)$의 함수)

$$\frac{\partial\mathcal{L}}{\partial w_{z2_2}} = \frac{\partial\mathcal{L}}{\partial z_2} \cdot \frac{\partial z_2}{\partial w_{z2_2}}$$

$$\frac{\partial\mathcal{L}}{\partial z_2} = \frac{\partial\mathcal{L}}{\partial\varphi(z_2)} \cdot \frac{\partial\varphi(z_2)}{\partial z_2}$$

신경망의 정의

신경망의 구조

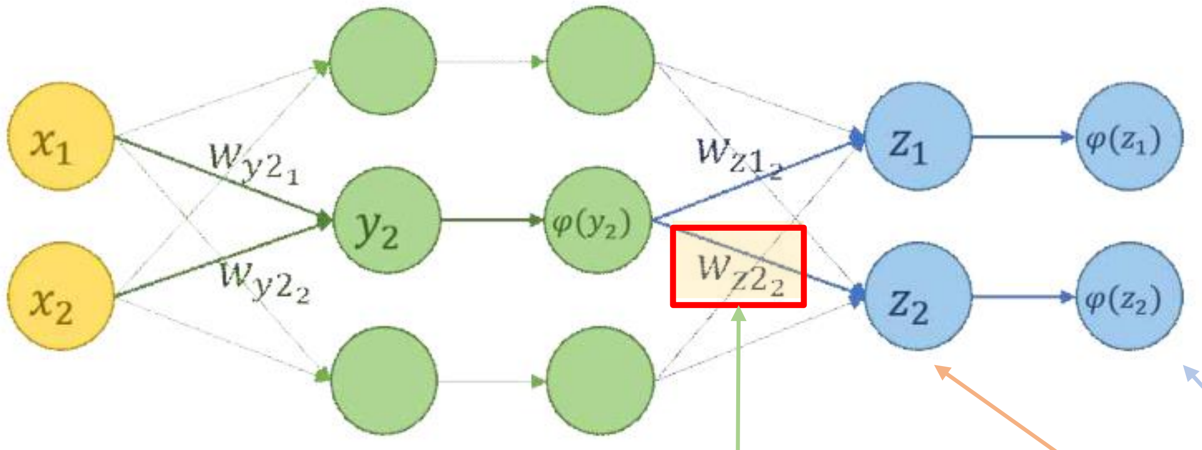
• **신경망의 학습**

순전파와 역전파 반복

Hyper-parameter

Backpropagation

② w_{z2_2} : $w_{z2_2\,new} = w_{z2_2\,old} - \delta \cdot \nabla\mathcal{L}(w)|_{w=w_{z2_2\,old}}$



$\dfrac{\partial\mathcal{L}}{\partial\varphi(z_2)}$ ($\mathcal{L}$ 은 $\varphi(\cdot)$의 함수)

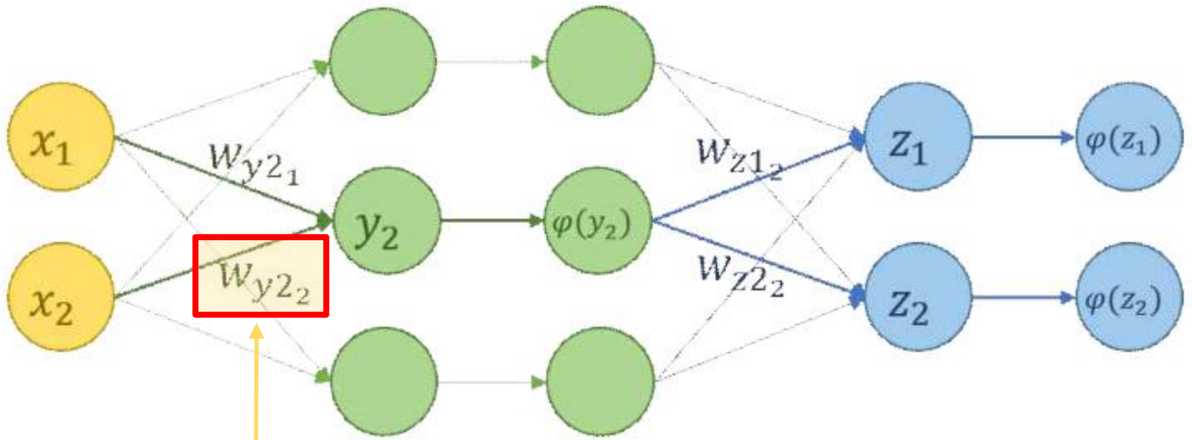
$$\frac{\partial\mathcal{L}}{\partial z_2} = \frac{\partial\mathcal{L}}{\partial\varphi(z_2)} \cdot \frac{\partial\varphi(z_2)}{\partial z_2}$$

$$\frac{\partial\mathcal{L}}{\partial w_{z2_2}} = \frac{\partial\mathcal{L}}{\partial z_2} \cdot \varphi(y_2)$$

COUTA

Backpropagation

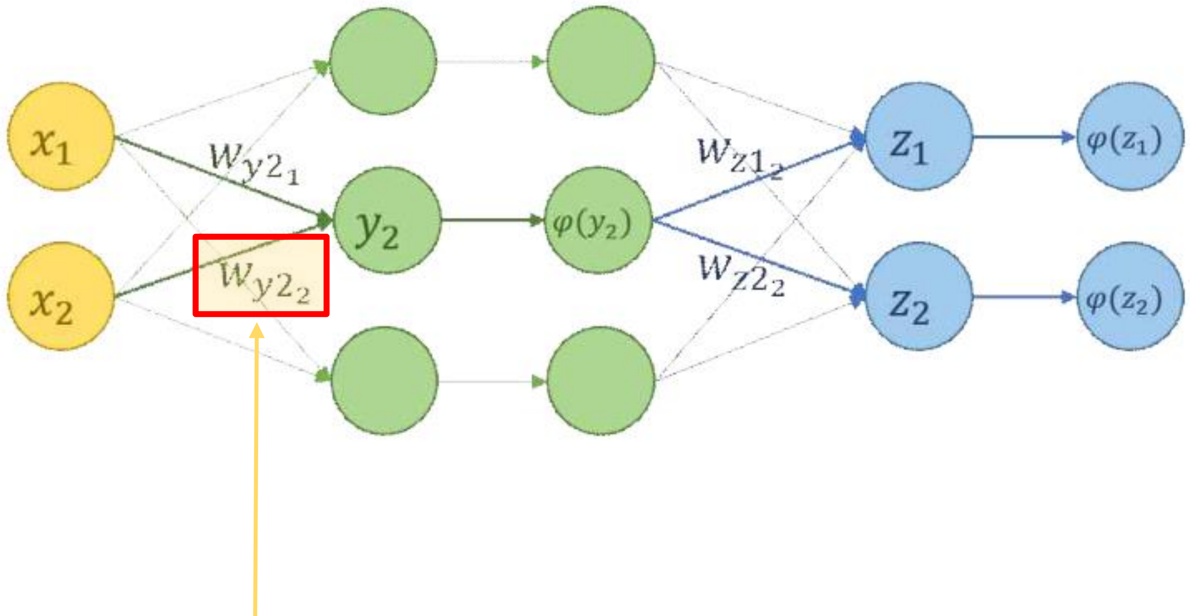
③ $\boldsymbol{w_{y2_2}}$: $w_{y2_2\,new} = w_{y2_2\,old} - \delta \cdot \nabla\mathcal{L}(w)|_{w=w_{y2_2\,old}}$



$$\frac{\partial\mathcal{L}}{\partial w_{y2_2}} = \frac{\partial\mathcal{L}}{\partial y_2} \cdot \frac{\partial y_2}{\partial w_{y2_2}}$$

Backpropagation

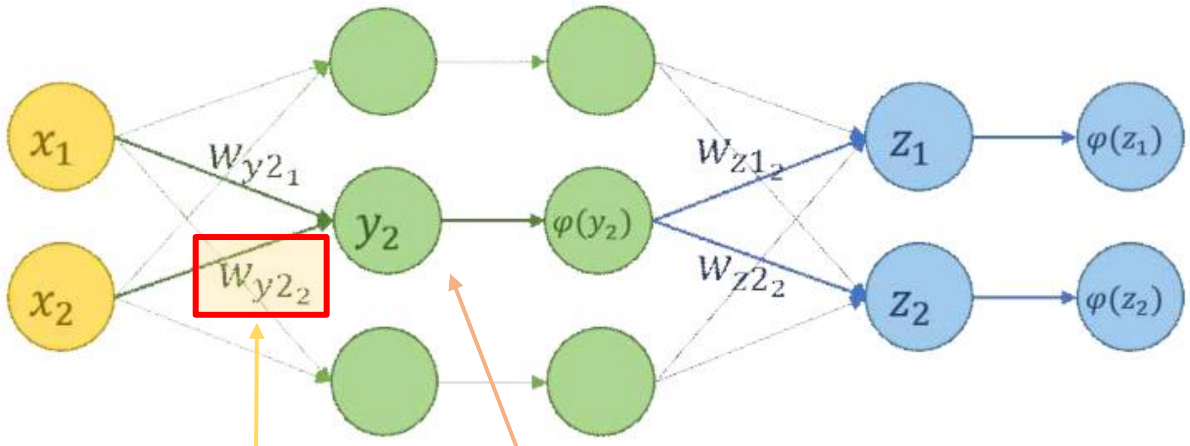
③ w_{y2_2} : $w_{y2_2\,new} = w_{y2_2\,old} - \delta \cdot \nabla\mathcal{L}(w)|_{w=w_{y2_2\,old}}$



$$\frac{\partial\mathcal{L}}{\partial w_{y2_2}} = \frac{\partial\mathcal{L}}{\partial y_2} \cdot \quad x_2$$

Backpropagation

③ w_{y2_2} : $w_{y2_2\,new} = w_{y2_2\,old} - \delta \cdot \nabla\mathcal{L}(w)|_{w=w_{y2_2\,old}}$

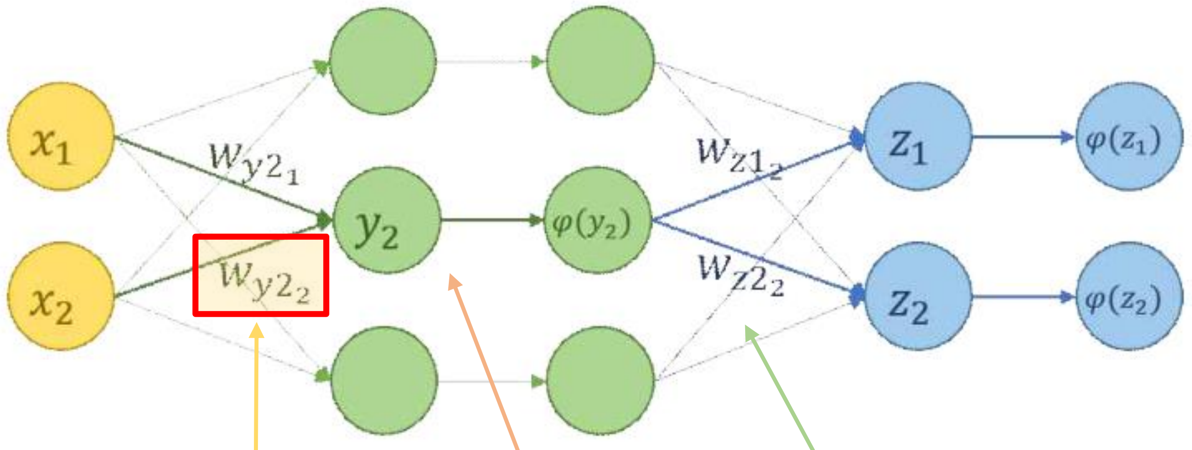


$$\frac{\partial\mathcal{L}}{\partial w_{y2_2}} = \frac{\partial\mathcal{L}}{\partial y_2} \cdot \quad x_2$$

$$\frac{\partial\mathcal{L}}{\partial y_2} = \frac{\partial\mathcal{L}}{\partial\varphi(y_2)} \cdot \frac{\partial\varphi(y_2)}{\partial y_2}$$

Backpropagation

③ w_{y2_2} : $w_{y2_2\,new} = w_{y2_2\,old} - \delta \cdot \nabla\mathcal{L}(w)|_{w=w_{y2_2\,old}}$



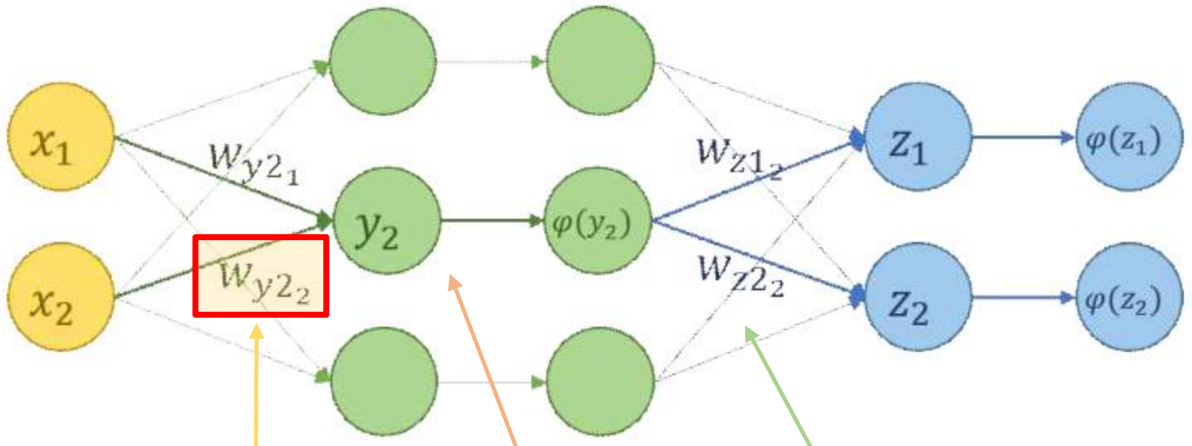
$$\frac{\partial\mathcal{L}}{\partial w_{y2_2}} = \frac{\partial\mathcal{L}}{\partial y_2} \cdot x_2$$

$$\frac{\partial\mathcal{L}}{\partial y_2} = \frac{\partial\mathcal{L}}{\partial\varphi(y_2)} \cdot \frac{\partial\varphi(y_2)}{\partial y_2}$$

$$\frac{\partial\mathcal{L}}{\partial\varphi(y_2)} = \frac{\partial\mathcal{L}}{\partial z_1} \cdot \frac{\partial z_1}{\partial\varphi(y_2)} + \frac{\partial\mathcal{L}}{\partial z_2} \cdot \frac{\partial z_2}{\partial\varphi(y_2)}$$

Backpropagation

③ w_{y2_2} : $w_{y2_2\,new} = w_{y2_2\,old} - \delta \cdot \nabla\mathcal{L}(w)|_{w=w_{y2_2\,old}}$



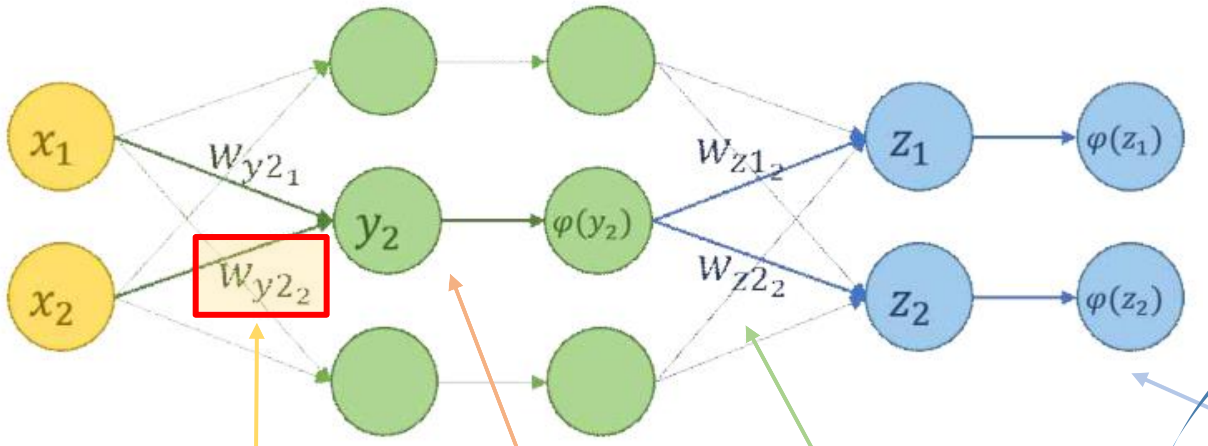
$$\frac{\partial\mathcal{L}}{\partial w_{y2_2}} = \frac{\partial\mathcal{L}}{\partial y_2} \cdot x_2$$

$$\frac{\partial\mathcal{L}}{\partial\varphi(y_2)} = \frac{\partial\mathcal{L}}{\partial z_1} \cdot w_{z1_2} + \frac{\partial\mathcal{L}}{\partial z_2} \cdot w_{z2_2}$$

$$\frac{\partial\mathcal{L}}{\partial y_2} = \frac{\partial\mathcal{L}}{\partial\varphi(y_2)} \cdot \frac{\partial\varphi(y_2)}{\partial y_2}$$

Backpropagation

③ w_{y2_2} : $w_{y2_2\,new} = w_{y2_2\,old} - \delta \cdot \nabla \mathcal{L}(w)|_{w=w_{y2_2\,old}}$



$$\frac{\partial \mathcal{L}}{\partial z_1} = \frac{\partial \mathcal{L}}{\partial \varphi(z_1)} \cdot \frac{\partial \varphi(z_1)}{\partial z_1}$$

($\mathcal{L}$ 은 $\varphi(\cdot)$의 함수)

$$\frac{\partial \mathcal{L}}{\partial z_2} = \frac{\partial \mathcal{L}}{\partial \varphi(z_2)} \cdot \frac{\partial \varphi(z_2)}{\partial z_2}$$

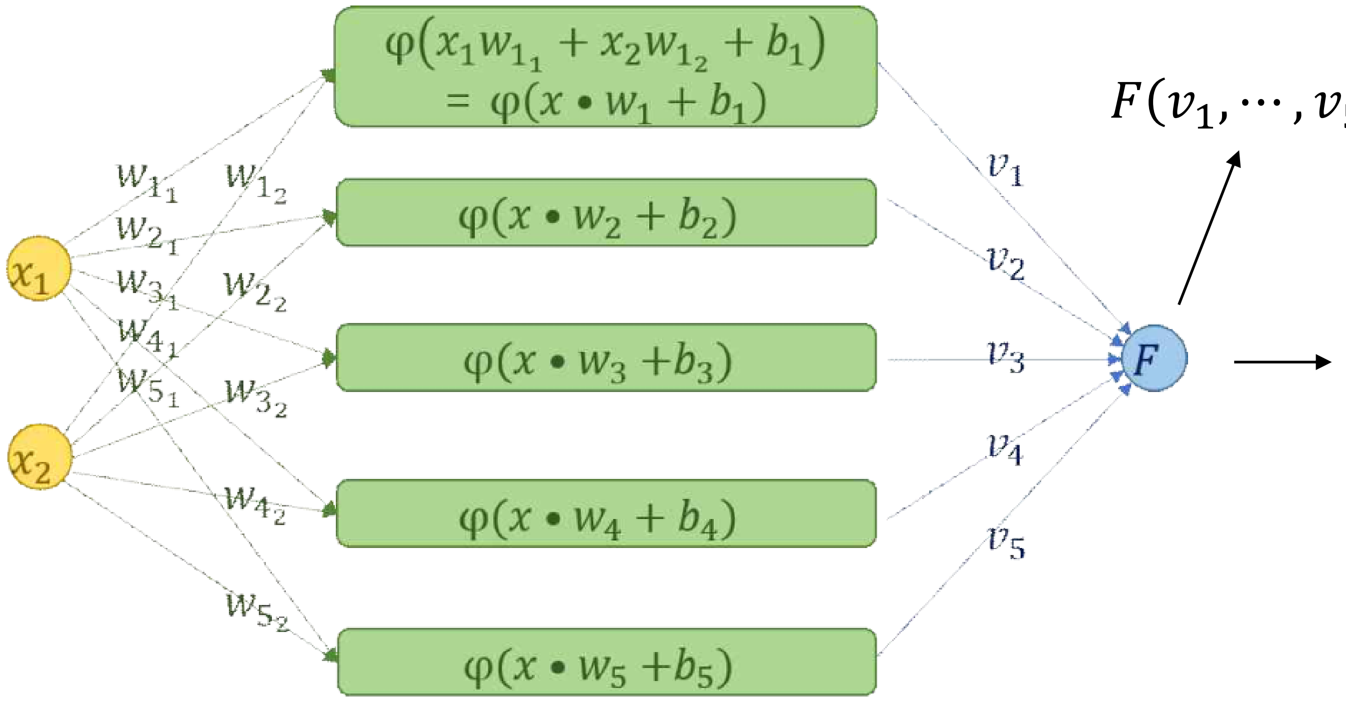
$$\frac{\partial \mathcal{L}}{\partial \varphi(y_2)} = \frac{\partial \mathcal{L}}{\partial z_1} \cdot w_{z1_2} + \frac{\partial \mathcal{L}}{\partial z_2} \cdot w_{z2_2}$$

$$\frac{\partial \mathcal{L}}{\partial w_{y2_2}} = \frac{\partial \mathcal{L}}{\partial y_2} \cdot x_2$$

$$\frac{\partial \mathcal{L}}{\partial y_2} = \frac{\partial \mathcal{L}}{\partial \varphi(y_2)} \cdot \frac{\partial \varphi(y_2)}{\partial y_2}$$

Backpropagation

Another Example



$$F(v_1, \cdots, v_5) = v_1 \cdot v_2 + \frac{v_3}{v_4} - v_5$$

활성화 함수 2

$$\varphi_2(x) = \frac{1}{1 + e^{-x}}$$

오차함수

$$\mathcal{L}(x) = (x - 2)^2$$

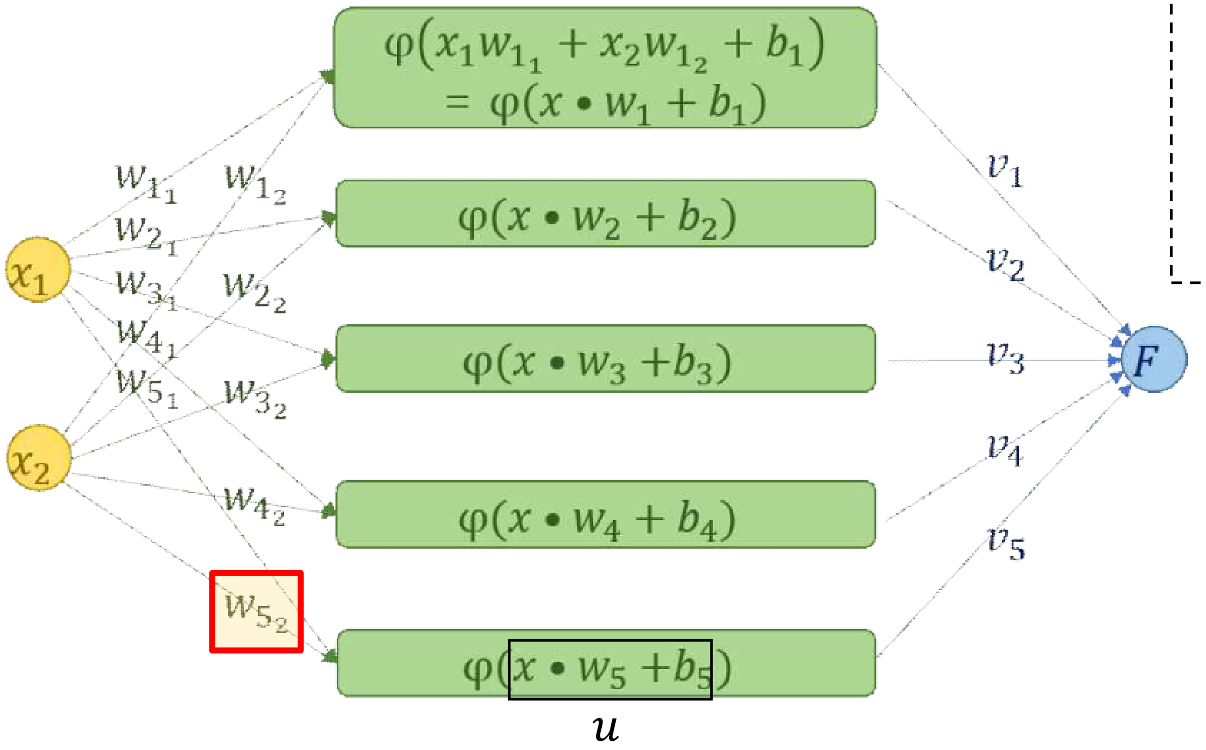
활성화 함수 1

$$\varphi(x) = \frac{1}{x}$$

OUTA

Backpropagation

Another Example



$$F(v_1, \cdots, v_5) = v_1 \cdot v_2 + \frac{v_3}{v_4} - v_5$$

활성화 함수 1 : $\varphi(x) = \frac{1}{x}$

활성화 함수 2 : $\varphi_2(x) = \frac{1}{1+e^{-x}}$

오차함수 : $\mathcal{L}(x) = (x-2)^2$

What is $\frac{\partial \mathcal{L}}{\partial w_{5_2}}$?

신경망의 정의

신경망의 구조

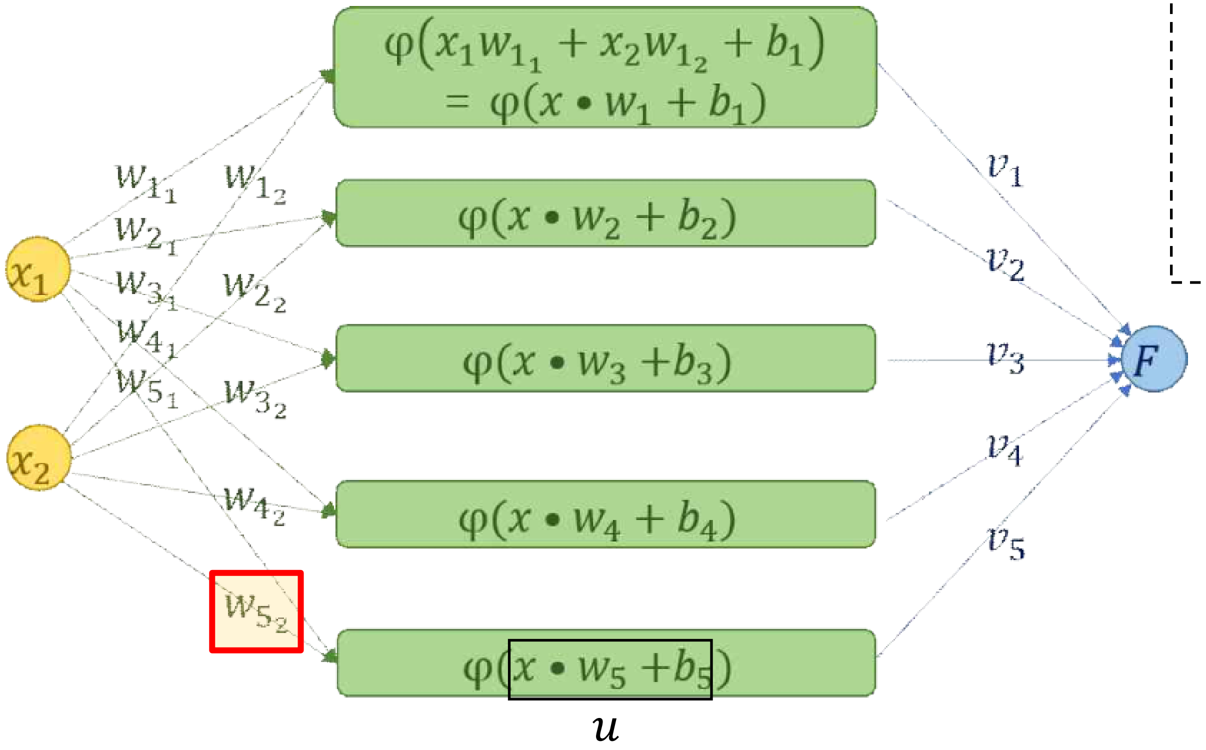
• **신경망의 학습**

순전파와 역전파 반복

Hyper-parameter

Backpropagation

Another Example



$$F(v_1, \cdots, v_5) = v_1 \cdot v_2 + \frac{v_3}{v_4} - v_5$$

활성화 함수 1 : $\varphi(x) = \frac{1}{x}$

활성화 함수 2 : $\varphi_2(x) = \frac{1}{1+e^{-x}}$

오차함수 : $\mathcal{L}(x) = (x-2)^2$

① $$\frac{\partial \mathcal{L}}{\partial w_{5_2}} = \frac{\partial \mathcal{L}}{\partial u} \cdot \frac{\partial u}{\partial w_{5_2}}$$

신경망의 정의

신경망의 구조

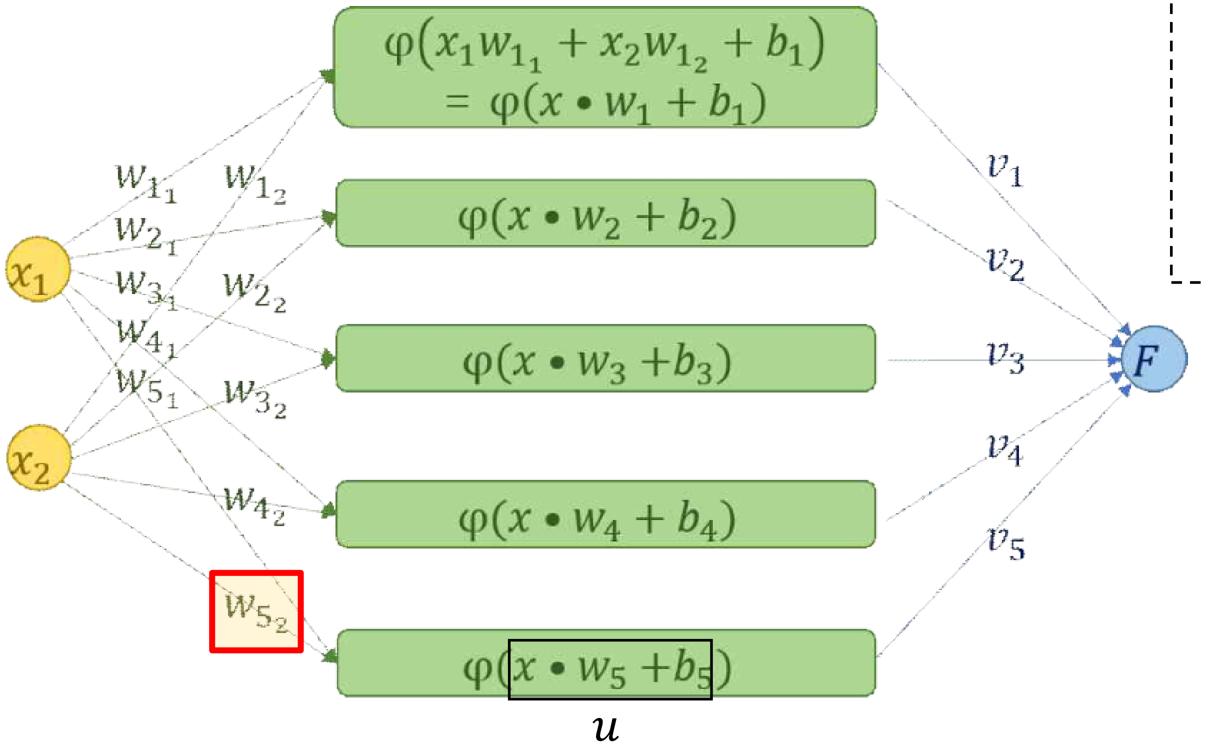
● **신경망의 학습**

순전파와 역전파 반복

Hyper-parameter

Backpropagation

Another Example



$$F(v_1, \cdots, v_5) = v_1 \cdot v_2 + \frac{v_3}{v_4} - v_5$$

활성화 함수 1 : $\varphi(x) = \frac{1}{x}$

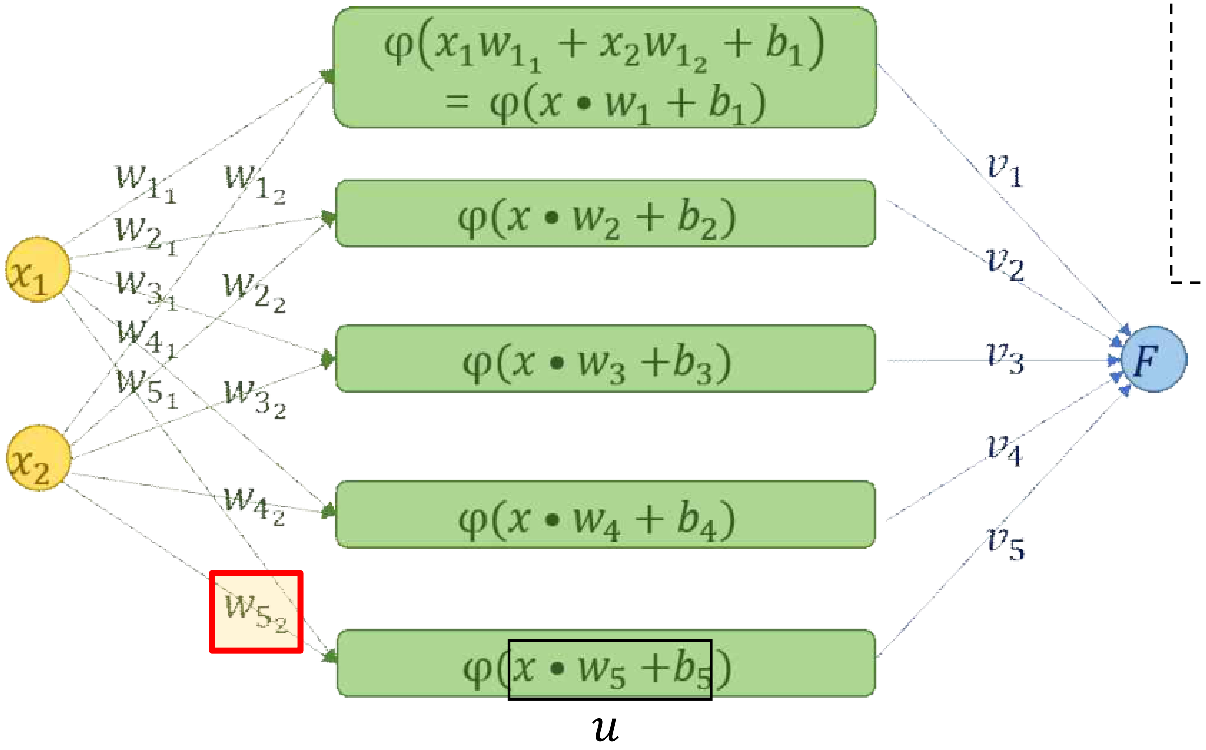
활성화 함수 2 : $\varphi_2(x) = \frac{1}{1+e^{-x}}$

오차함수 : $\mathcal{L}(x) = (x-2)^2$

① $$\frac{\partial \mathcal{L}}{\partial w_{5_2}} = \frac{\partial \mathcal{L}}{\partial u} \cdot x_2$$

Backpropagation

Another Example



$$F(v_1, \cdots, v_5) = v_1 \cdot v_2 + \frac{v_3}{v_4} - v_5$$

활성화 함수 1 : $\varphi(x) = \frac{1}{x}$

활성화 함수 2 : $\varphi_2(x) = \frac{1}{1+e^{-x}}$

오차함수 : $\mathcal{L}(x) = (x-2)^2$

① $\frac{\partial \mathcal{L}}{\partial w_{5_2}} = \frac{\partial \mathcal{L}}{\partial u} \cdot x_2$

② $\frac{\partial \mathcal{L}}{\partial u} = \frac{\partial \mathcal{L}}{\partial \varphi(u)} \cdot \frac{\partial \varphi(u)}{\partial u}$

Backpropagation

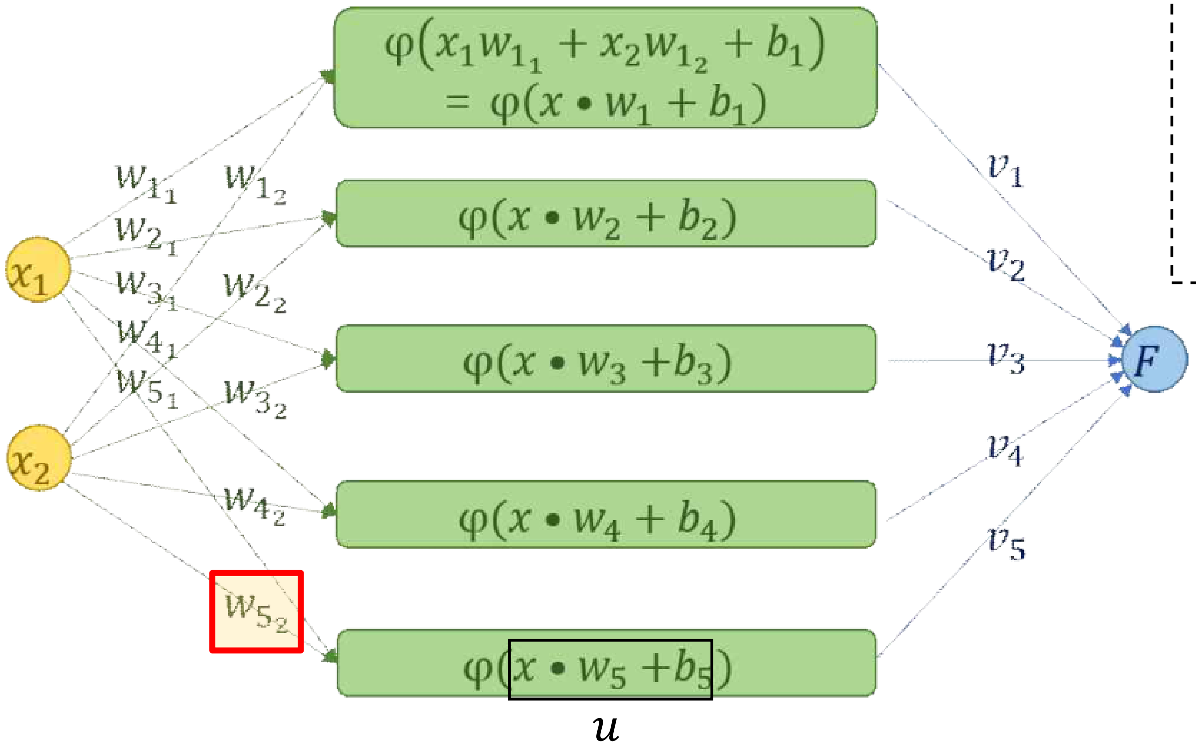
Another Example

$$F(v_1, \cdots, v_5) = v_1 \cdot v_2 + \frac{v_3}{v_4} - v_5$$

활성화 함수 1 : $\varphi(x) = \frac{1}{x}$

활성화 함수 2 : $\varphi_2(x) = \frac{1}{1+e^{-x}}$

오차함수 : $\mathcal{L}(x) = (x-2)^2$



① $\frac{\partial \mathcal{L}}{\partial w_{5_2}} = \frac{\partial \mathcal{L}}{\partial u} \cdot x_2$

② $\frac{\partial \mathcal{L}}{\partial u} = \frac{\partial \mathcal{L}}{\partial \varphi(u)} \cdot \left(-\frac{1}{u^2}\right)$

Backpropagation

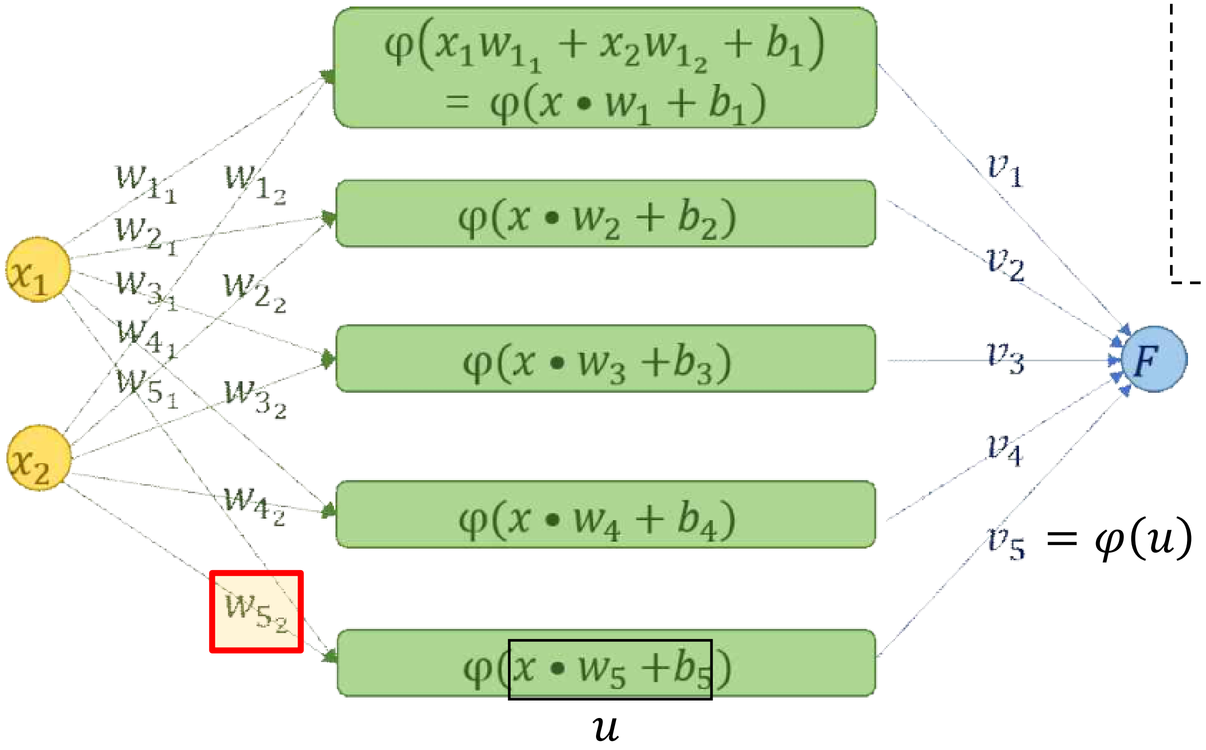
Another Example

$$F(v_1, \cdots, v_5) = v_1 \cdot v_2 + \frac{v_3}{v_4} - v_5$$

활성화 함수 1 : $\varphi(x) = \frac{1}{x}$

활성화 함수 2 : $\varphi_2(x) = \frac{1}{1+e^{-x}}$

오차함수 : $\mathcal{L}(x) = (x-2)^2$



$v_5 = \varphi(u)$

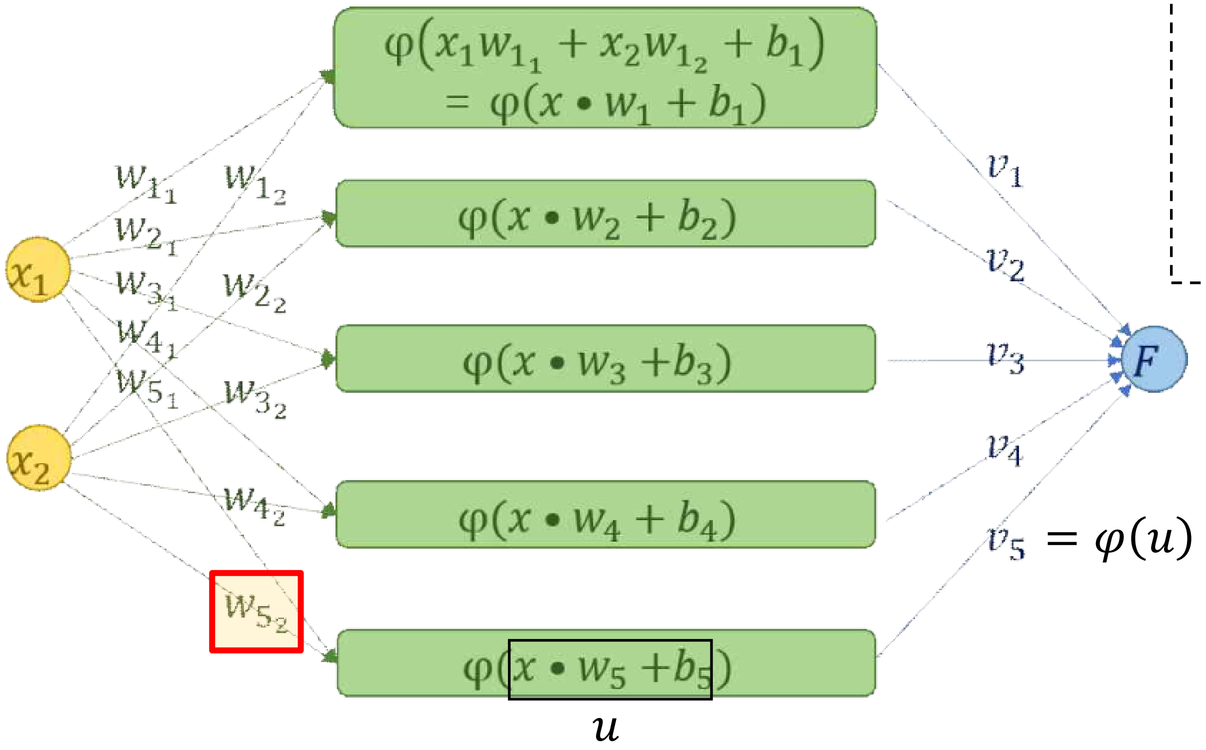
① $\frac{\partial \mathcal{L}}{\partial w_{5_2}} = \frac{\partial \mathcal{L}}{\partial u} \cdot x_2$

② $\frac{\partial \mathcal{L}}{\partial u} = \frac{\partial \mathcal{L}}{\partial \varphi(u)} \cdot \left(-\frac{1}{u^2}\right)$

③ $\frac{\partial \mathcal{L}}{\partial \varphi(u)} = \frac{\partial \mathcal{L}}{\partial F} \cdot \frac{\partial F}{\partial \varphi(u)}$

Backpropagation

Another Example



$$F(v_1, \cdots, v_5) = v_1 \cdot v_2 + \frac{v_3}{v_4} - v_5$$

활성화 함수 1 : $\varphi(x) = \frac{1}{x}$

활성화 함수 2 : $\varphi_2(x) = \frac{1}{1+e^{-x}}$

오차함수 : $\mathcal{L}(x) = (x-2)^2$

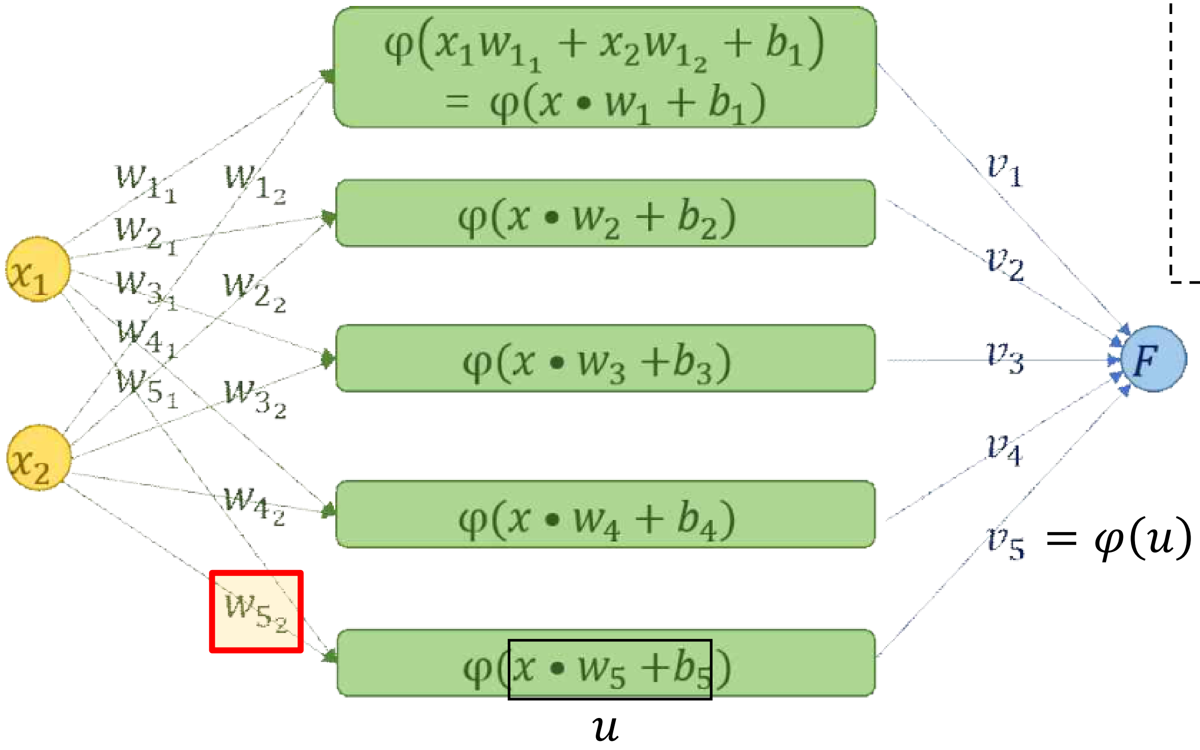
① $\frac{\partial \mathcal{L}}{\partial w_{5_2}} = \frac{\partial \mathcal{L}}{\partial u} \cdot x_2$

② $\frac{\partial \mathcal{L}}{\partial u} = \frac{\partial \mathcal{L}}{\partial \varphi(u)} \cdot \left(-\frac{1}{u^2}\right)$

③ $\frac{\partial \mathcal{L}}{\partial \varphi(u)} = \frac{\partial \mathcal{L}}{\partial F} \cdot (-1)$

Backpropagation

Another Example



$v_5 = \varphi(u)$

$$F(v_1, \cdots, v_5) = v_1 \cdot v_2 + \frac{v_3}{v_4} - v_5$$

활성화 함수 1 : $\varphi(x) = \frac{1}{x}$

활성화 함수 2 : $\varphi_2(x) = \frac{1}{1+e^{-x}}$

오차함수 : $\mathcal{L}(x) = (x-2)^2$

① $\frac{\partial \mathcal{L}}{\partial w_{5_2}} = \frac{\partial \mathcal{L}}{\partial u} \cdot x_2$

② $\frac{\partial \mathcal{L}}{\partial u} = \frac{\partial \mathcal{L}}{\partial \varphi(u)} \cdot \left(-\frac{1}{u^2}\right)$

③ $\frac{\partial \mathcal{L}}{\partial \varphi(u)} = \frac{\partial \mathcal{L}}{\partial F} \cdot (-1)$

④ $\frac{\partial \mathcal{L}}{\partial F} = \frac{\partial \mathcal{L}}{\partial \varphi_2(F)} \cdot \frac{\partial \varphi_2(F)}{\partial F}$

OUTTA

Backpropagation

Sigmoid 함수 미분

$$\varphi_2(x) = \frac{1}{1+e^{-x}}$$

$$\frac{\partial \varphi_2(x)}{\partial x} = \frac{-e^{-x}}{(1+e^{-x})^2} = \left(1 - \varphi_2(x)\right) \cdot \varphi_2(x)$$

신경망의 정의

신경망의 구조

- **신경망의 학습**

순전파와 역전파 반복

Hyper-parameter

Backpropagation

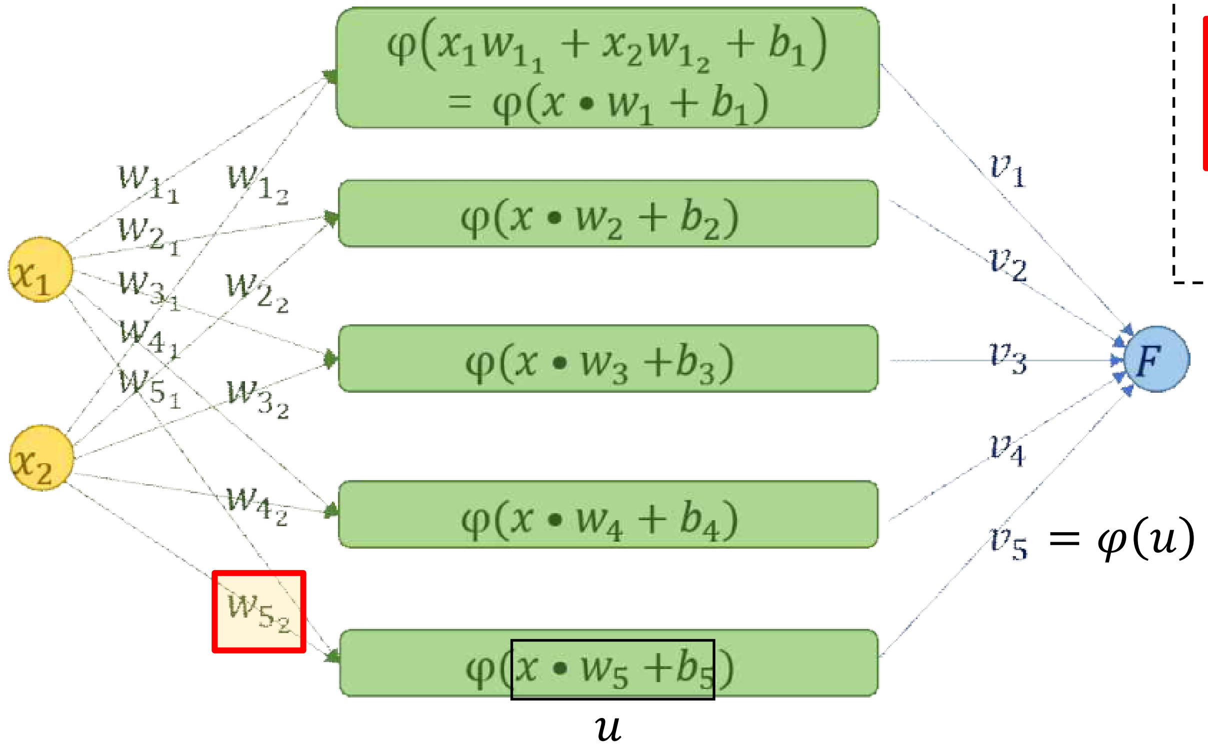
Another Example

$$F(v_1,\cdots,v_5) = v_1 \cdot v_2 + \frac{v_3}{v_4} - v_5$$

활성화 함수 1 : $\varphi(x) = \frac{1}{x}$

활성화 함수 2 : $\varphi_2(x) = \frac{1}{1+e^{-x}}$

오차함수 : $\mathcal{L}(x) = (x-2)^2$



① $\frac{\partial \mathcal{L}}{\partial w_{5_2}} = \frac{\partial \mathcal{L}}{\partial u} \cdot x_2$

② $\frac{\partial \mathcal{L}}{\partial u} = \frac{\partial \mathcal{L}}{\partial \varphi(u)} \cdot \left(-\frac{1}{u^2}\right)$

③ $\frac{\partial \mathcal{L}}{\partial \varphi(u)} = \frac{\partial \mathcal{L}}{\partial F} \cdot (-1)$

④ $\frac{\partial \mathcal{L}}{\partial F} = \frac{\partial \mathcal{L}}{\partial \varphi_2(F)} \cdot (1-\varphi_2(F)) \cdot \varphi_2(F)$

신경망의 정의

신경망의 구조

● **신경망의 학습**

순전파와 역전파 반복

Hyper-parameter

Backpropagation

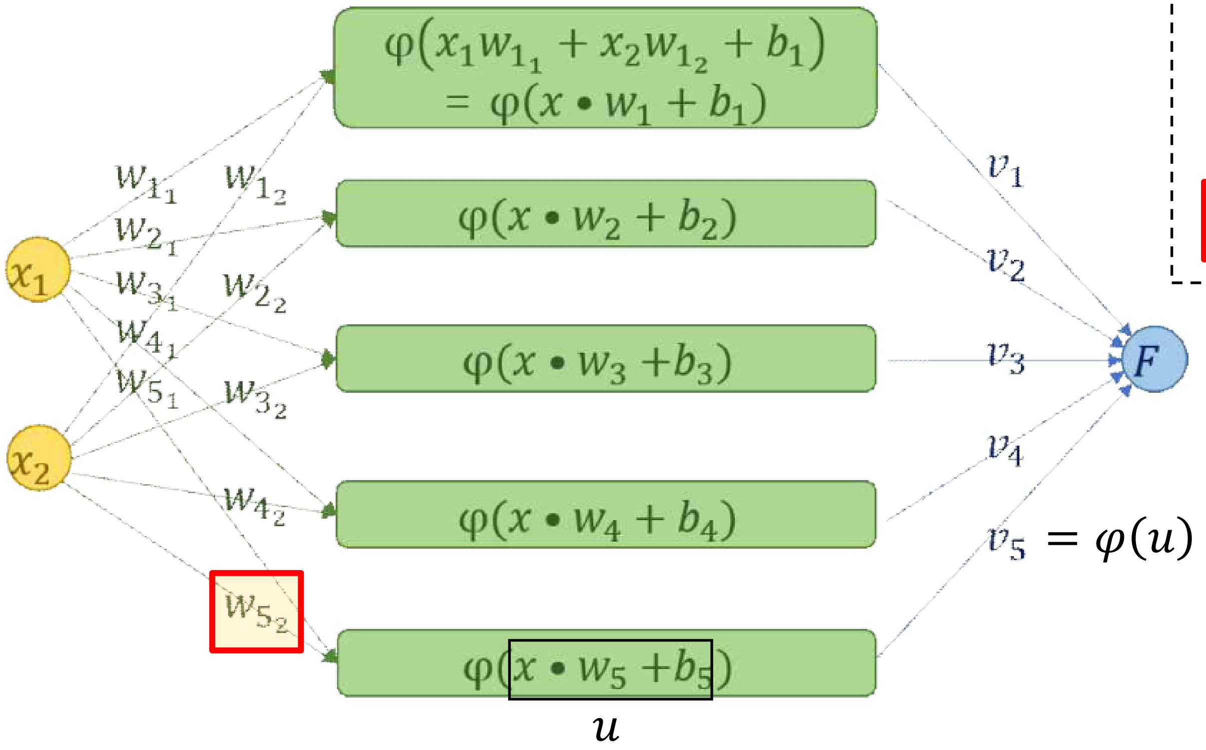
Another Example

$$F(v_1, \cdots, v_5) = v_1 \cdot v_2 + \frac{v_3}{v_4} - v_5$$

활성화 함수 1 : $\varphi(x) = \frac{1}{x}$

활성화 함수 2 : $\varphi_2(x) = \frac{1}{1+e^{-x}}$

오차함수 : $\mathcal{L}(x) = (x-2)^2$



① $\frac{\partial \mathcal{L}}{\partial w_{5_2}} = \frac{\partial \mathcal{L}}{\partial u} \cdot x_2$

② $\frac{\partial \mathcal{L}}{\partial u} = \frac{\partial \mathcal{L}}{\partial \varphi(u)} \cdot \left(-\frac{1}{u^2}\right)$

③ $\frac{\partial \mathcal{L}}{\partial \varphi(u)} = \frac{\partial \mathcal{L}}{\partial F} \cdot (-1)$

④ $\frac{\partial \mathcal{L}}{\partial F} = \frac{\partial \mathcal{L}}{\partial \varphi_2(F)} \cdot (1 - \varphi_2(F)) \cdot \varphi_2(F)$

⑤ $\frac{\partial \mathcal{L}}{\partial \varphi_2(F)} = 2(\varphi_2(F) - 2)$

COUTA

Backpropagation

Another Example

① $$\frac{\partial \mathcal{L}}{\partial w_{5_2}} = \frac{\partial \mathcal{L}}{\partial u} \cdot x_2$$

② $$\frac{\partial \mathcal{L}}{\partial u} = \frac{\partial \mathcal{L}}{\partial \varphi(u)} \cdot \left(-\frac{1}{u^2}\right)$$

③ $$\frac{\partial \mathcal{L}}{\partial \varphi(u)} = \frac{\partial \mathcal{L}}{\partial F} \cdot (-1)$$

④ $$\frac{\partial \mathcal{L}}{\partial F} = \frac{\partial \mathcal{L}}{\partial \varphi_2(F)} \cdot \big(1 - \varphi_2(F)\big) \cdot \varphi_2(F)$$

⑤ $$\frac{\partial \mathcal{L}}{\partial \varphi_2(F)} = 2(\varphi_2(F) - 2)$$

⑥ $$F(v_1, \cdots, v_5) = v_1 \cdot v_2 + \frac{v_3}{v_4} - v_5$$

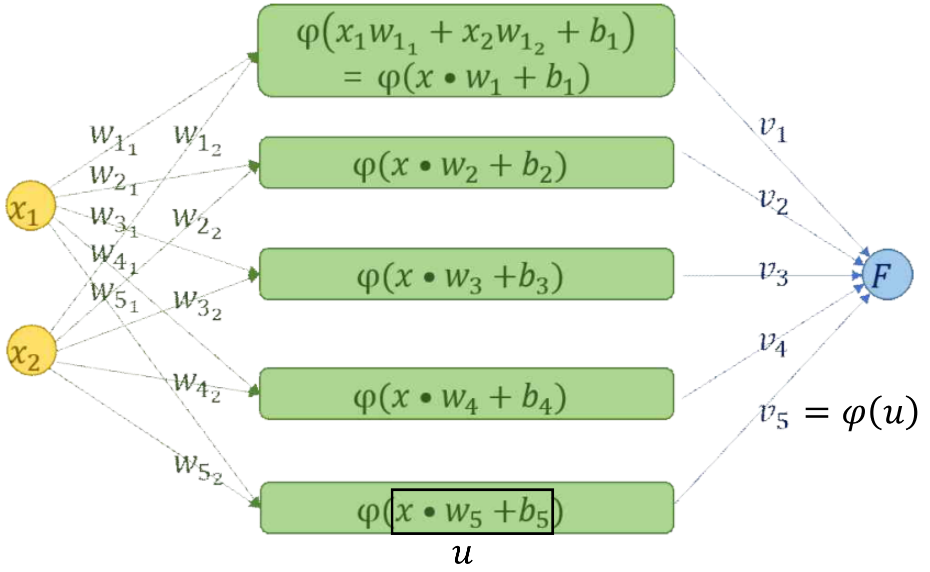
⑦ $$\varphi(x) = \frac{1}{x}$$

⑧ $$\varphi_2(x) = \frac{1}{1 + e^{-x}}$$

⑨ $$\mathcal{L}(x) = (x - 2)^2$$

⑩

w_{1_1}	0.1	w_{2_1}	1.1
w_{1_2}	0.5	w_{2_2}	0.0
w_{1_3}	0.3	w_{2_3}	0.4
w_{1_4}	2.1	w_{2_4}	-0.1
w_{1_5}	-0.2	w_{2_5}	1.3
x_1	1.7	x_2	3.5



신경망의 정의

신경망의 구조

- **신경망의 학습**

순전파와 역전파 반복

Hyper-parameter

Backpropagation

Another Example

① $\boxed{\frac{\partial \mathcal{L}}{\partial w_{5_2}}} = \frac{\partial \mathcal{L}}{\partial u} \cdot x_2$??

② $\frac{\partial \mathcal{L}}{\partial u} = \frac{\partial \mathcal{L}}{\partial \varphi(u)} \cdot \left(-\frac{1}{u^2}\right)$

③ $\frac{\partial \mathcal{L}}{\partial \varphi(u)} = \frac{\partial \mathcal{L}}{\partial F} \cdot (-1)$

④ $\frac{\partial \mathcal{L}}{\partial F} = \frac{\partial \mathcal{L}}{\partial \varphi_2(F)} \cdot (1-\varphi_2(F)) \cdot \varphi_2(F)$

⑤ $\frac{\partial \mathcal{L}}{\partial \varphi_2(F)} = 2(\varphi_2(F) - 2)$

⑥ $F(v_1, \cdots, v_5) = v_1 \cdot v_2 + \frac{v_3}{v_4} - v_5$

⑦ $\varphi(x) = \frac{1}{x}$

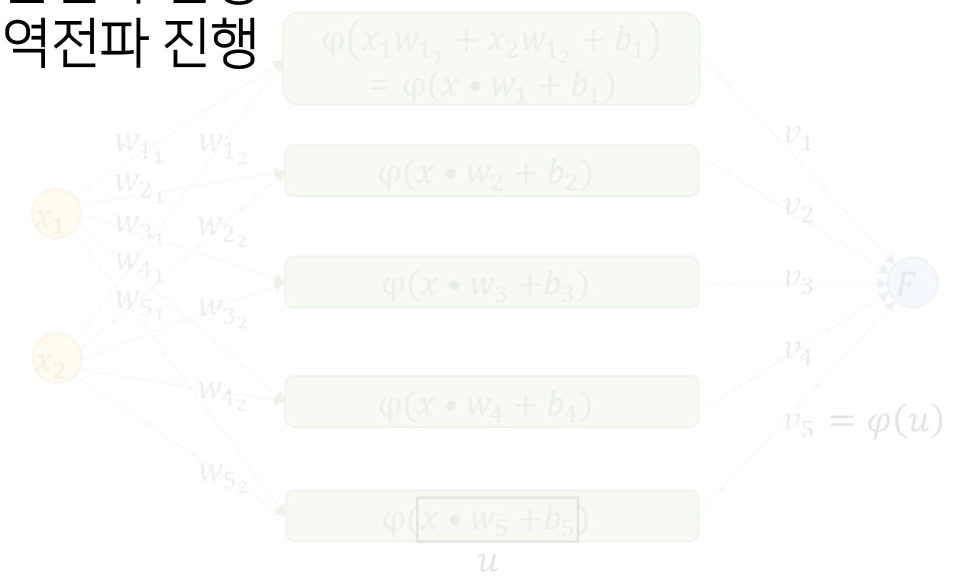
⑧ $\varphi_2(x) = \frac{1}{1+e^{-x}}$

⑨ $\mathcal{L}(x) = (x-2)^2$

⑩

w_{1_1}	0.1	w_{2_1}	1.1
w_{1_2}	0.5	w_{2_2}	0.0
w_{1_3}	0.3	w_{2_3}	0.4
w_{1_4}	2.1	w_{2_4}	-0.1
w_{1_5}	-0.2	w_{2_5}	1.3
x_1	1.7	x_2	3.5

연산 순서 :
(1) 순전파 진행
(2) 역전파 진행



Backpropagation

Another Example

(1) 순전파 진행

① $\boxed{\frac{\partial \mathcal{L}}{\partial w_{5_2}}} = \frac{\partial \mathcal{L}}{\partial u} \cdot x_2$??

② $\frac{\partial \mathcal{L}}{\partial u} = \frac{\partial \mathcal{L}}{\partial \varphi(u)} \cdot \left(-\frac{1}{u^2}\right)$

③ $\frac{\partial \mathcal{L}}{\partial \varphi(u)} = \frac{\partial \mathcal{L}}{\partial F} \cdot (-1)$

④ $\frac{\partial \mathcal{L}}{\partial F} = \frac{\partial \mathcal{L}}{\partial \varphi_2(F)} \cdot \left(1 - \varphi_2(F)\right) \cdot \varphi_2(F)$

⑤ $\frac{\partial \mathcal{L}}{\partial \varphi_2(F)} = 2(\varphi_2(F) - 2)$

⑥ $F(v_1, \cdots, v_5) = v_1 \cdot v_2 + \frac{v_3}{v_4} - v_5$

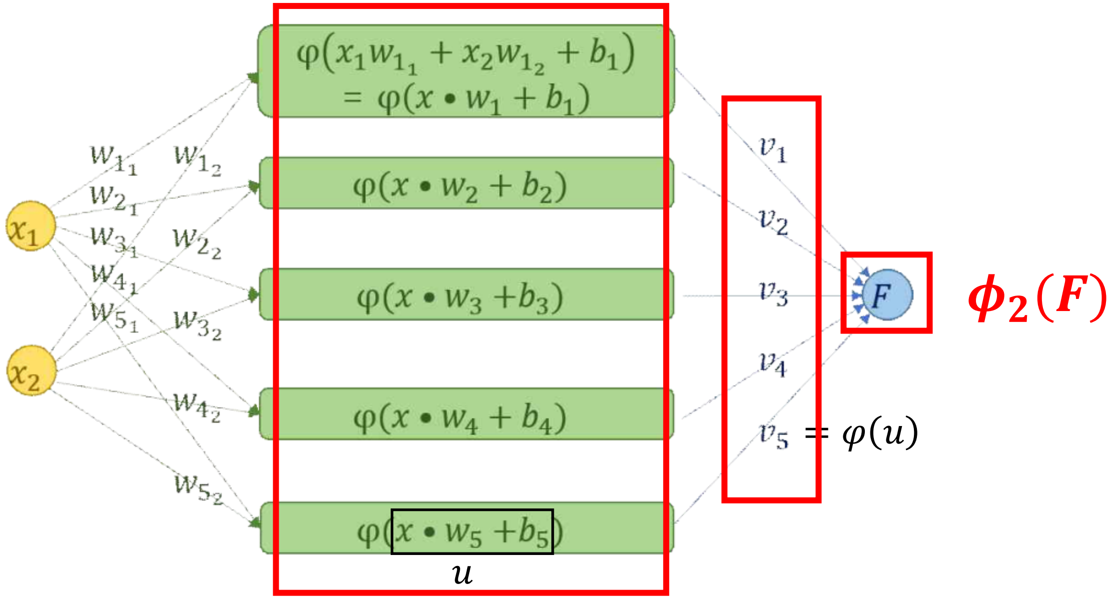
⑦ $\varphi(x) = \frac{1}{x}$

⑧ $\varphi_2(x) = \frac{1}{1 + e^{-x}}$

⑨ $\mathcal{L}(x) = (x - 2)^2$

⑩

w_{1_1}	0.1	w_{2_1}	1.1
w_{1_2}	0.5	w_{2_2}	0.0
w_{1_3}	0.3	w_{2_3}	0.4
w_{1_4}	2.1	w_{2_4}	-0.1
w_{1_5}	-0.2	w_{2_5}	1.3
x_1	1.7	x_2	3.5



- 신경망의 정의
- 신경망의 구조
- • 신경망의 학습
- 순전파와 역전파 반복
- Hyper-parameter

Backpropagation

Another Example

(2) 역전파 진행

① $\boxed{\dfrac{\partial \mathcal{L}}{\partial w_{5_2}}} = \dfrac{\partial \mathcal{L}}{\partial u} \cdot x_2$

② $\dfrac{\partial \mathcal{L}}{\partial u} = \dfrac{\partial \mathcal{L}}{\partial \varphi(u)} \cdot \left(-\dfrac{1}{u^2}\right)$

③ $\dfrac{\partial \mathcal{L}}{\partial \varphi(u)} = \dfrac{\partial \mathcal{L}}{\partial F} \cdot (-1)$

④ $\dfrac{\partial \mathcal{L}}{\partial F} = \dfrac{\partial \mathcal{L}}{\partial \varphi_2(F)} \cdot \left(1 - \varphi_2(F)\right) \cdot \varphi_2(F)$

⑤ $\dfrac{\partial \mathcal{L}}{\partial \varphi_2(F)} = 2(\varphi_2(F) - 2)$

⑥ $F(v_1, \cdots, v_5) = v_1 \cdot v_2 + \dfrac{v_3}{v_4} - v_5$

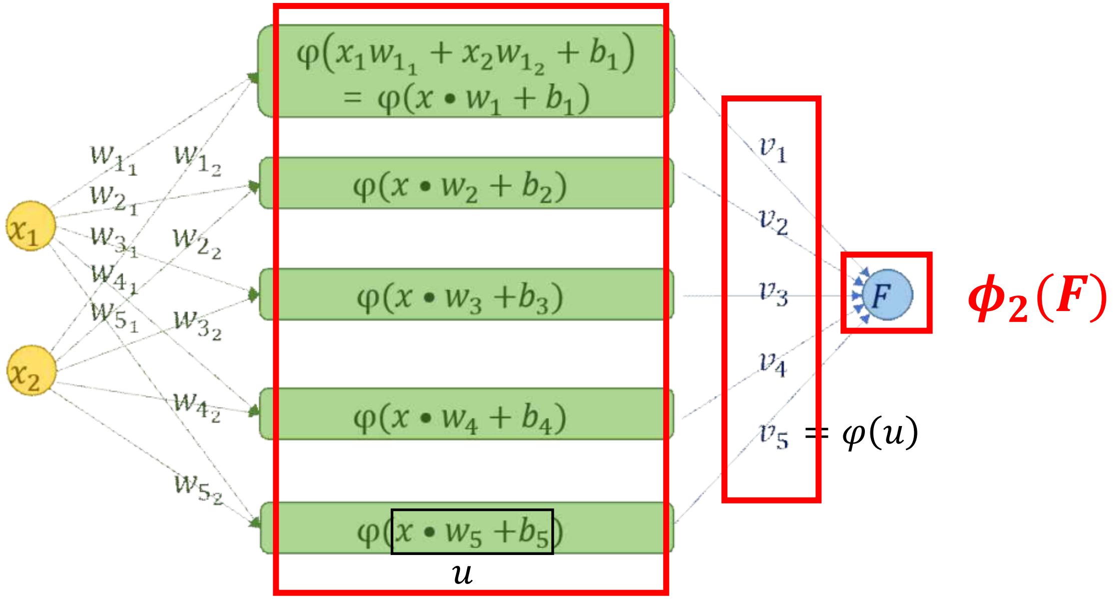
⑦ $\varphi(x) = \dfrac{1}{x}$

⑧ $\varphi_2(x) = \dfrac{1}{1 + e^{-x}}$

⑨ $\mathcal{L}(x) = (x - 2)^2$

⑩

w_{1_1}	0.1	w_{2_1}	1.1
w_{1_2}	0.5	w_{2_2}	0.0
w_{1_3}	0.3	w_{2_3}	0.4
w_{1_4}	2.1	w_{2_4}	-0.1
w_{1_5}	-0.2	w_{2_5}	1.3
x_1	1.7	x_2	3.5



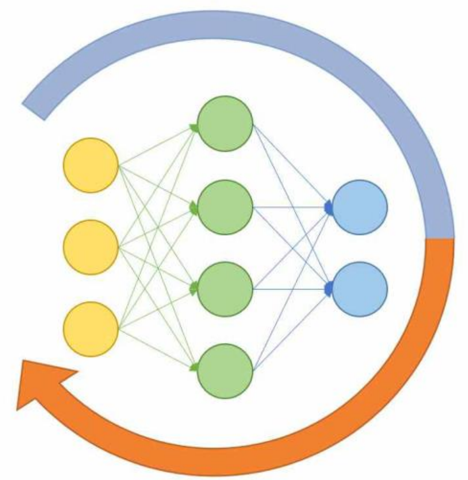
5. 순전파와 역전파의 반복

<머신러닝 첫 단추 끼우기>
4.4.절

COUTA

1 Iteration 의 정의

- 데이터 집합에 대해 순전파, 역전파를 통해 가중치 업데이트가 **1회** 일어나는 것



- 신경망에 1개의 Data씩 입력해줄 때 (Data 총 개수 : N개)

  → 전체 데이터에 대해 가중치 업데이트를 하기 위해 N iteration 필요

- 신경망에 모든 Data를 한 번에 입력할 때

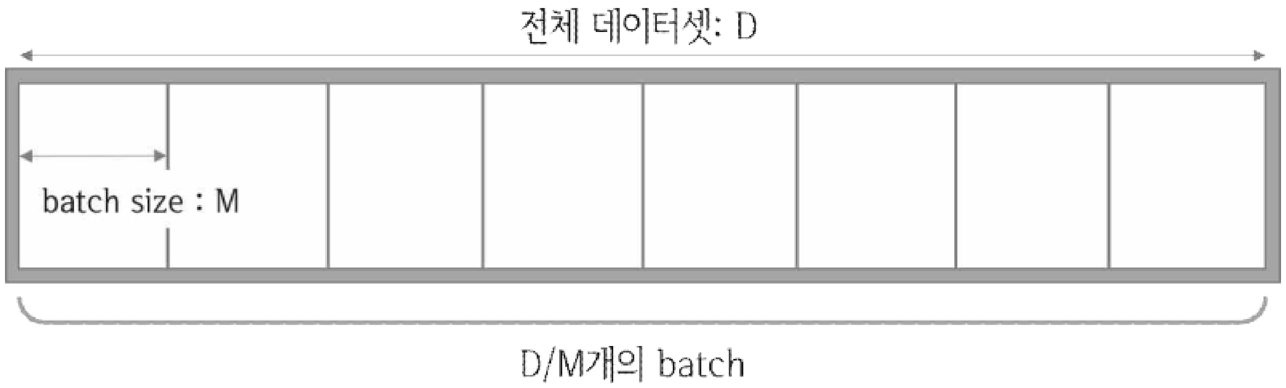
  → 전체 데이터에 대해 가중치 업데이트를 하기 위해 1 iteration 필요

실제 신경망에서는?

- N개의 전체 데이터를 **M개씩 묶어** 신경망에 입력

Batch (배치) / Batch Size

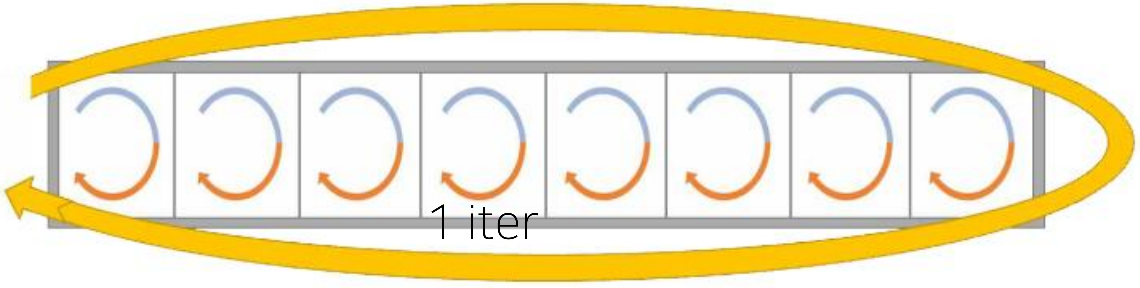
- Batch : **한 번에 신경망에 입력하는 데이터 묶음**
- Batch Size : **1개의 Batch 내의 데이터 수 (= M)**
- N개의 Data를 랜덤하게 섞고, M개씩 골라 Batch 생성



Batch에서의 순전파와 역전파

- 순전파 : 각 batch를 이루는 데이터에 대해 독립적으로 진행
- 역전파 : 각 batch에서 계산된 손실함수의 평균을 이용해 진행
- More Details on ML LAB 5! (Next Seminar)

1 epoch 의 정의



- **전체 데이터**에 대해 순전파, 역전파를 통해 가중치 업데이트가 **1회** 일어나는 것

Q) 전체 데이터가 D개, Batch Size가 M일 때

- Batch 의 개수는? **D/M**
- 1 epoch 는 몇 iteration으로 이루어져 있는가? **D/M**

Batch 를 사용하는 이유?

In Next Seminar!

COUTA

6. Hyperparameter

<머신러닝 첫 단추 끼우기>

4.5.절

What is Hyperparameter?

- 딥러닝 모델의 성능에 영향을 주는 변수
- 사용자가 직접 설정해 입력하며, 최적의 성능을 위해서는 Tuning 과정이 필요함
- "Parameter (가중치)를 효과적으로 업데이트 해 주기 위한 요소"

- Examples
  - 가중치의 초깃값
  - Learning rate
  - 학습의 종료조건
  - Layer Depth
  - Numbers of nodes in 1 Layer
  - Activation Function
  - Loss Function
  - Optimizer
  - batch size

7. Summary

Summary

- **수학적 모델을 신경망으로 근사**할 수 있음

- 신경망 간선에 들어가는 **가중치 및 편향의 적절한 값**을 찾는 것이 해야할 일 (학습)

- 우선 파라미터 값을 적당히 초기화한 후, **순전파** 결과를 도출하고 **손실함수**를 통해
  그 결과가 얼마나 좋은지 평가

- “손실함수 값을 작게 하려면 파라미터 값을 어떻게 변화시킬까?” 에 대한 답을 주는 것이 **경사하강법**

- 가중치 및 편향에 대한 미분값을 구하는 **역전파** 과정에서 연쇄법칙을 사용