



부산대학교

PUSAN NATIONAL UNIVERSITY

Multilayer Perceptrons

Stacking the Layers

Hidden Units and Nonlinear Activation

BX3600087-140 / DS3600087-140 / AB3600087-140

기계학습개론

2026학년도 2학기

오용경

데이터사이언스전공

yongkyung.oh@pusan.ac.kr

Section 1 • Multilayer Perceptrons

Adding a Hidden Layer

입력과 출력 사이에서 값을 변환하는 은닉층을 설명한다

층을 하나 더 두면

입력과 출력 사이에 계산 단계를 하나 끼워 넣는다

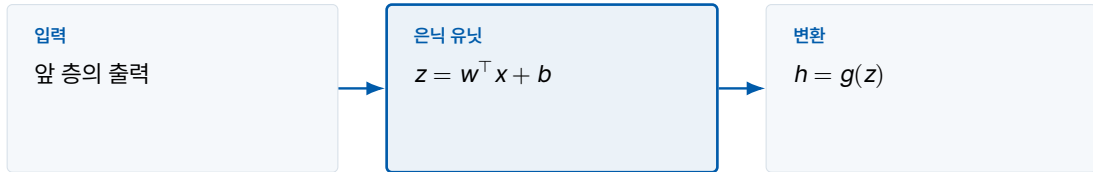
층(layer)	담는 것	하는 일
입력층	관측의 특성	입력값을 전달
은닉층	변환된 중간값	가중합과 활성화함수 계산
출력층	예측값	은닉층의 출력을 조합

은닉층과 은닉 유닛

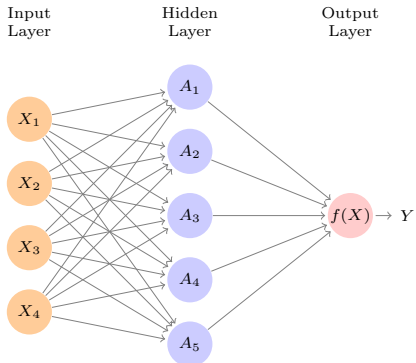
DEFINITION

은닉층(hidden layer)

입력층과 출력층 사이에서 값을 변환하는 층
각 출력값을 계산하는 단위를 은닉 유닛(hidden unit)이라 한다



은닉층이 입력과 출력을 연결한다



왼쪽 입력이 가운데 은닉층을 거쳐 오른쪽 출력으로 이어진다.

Source: James et al. (2023), Figure 10.1.

은닉층은 가중합에 활성화함수를 적용한다

완전 연결 은닉층에서는 모든 입력이 모든 은닉 유닛에 연결된다

$$h = g(W_1 x + b_1), \quad s = w_2^\top h + b_2$$

W_1 의 행 수는 은닉 유닛 수이며, g 는 각 성분에 적용하는 비선형 활성화함수다

은닉층 계산의 모양을 확인한다

기호	모양	담고 있는 것
x	$p \times 1$	한 관측의 특성
W_1	$K \times p$	은닉 유닛마다 한 행
h	$K \times 1$	새로 만든 특성

W_1 의 각 행에는 은닉 유닛 하나의 가중치 벡터가 들어 있다

은닉 유닛 수는 검증 성능으로 선택한다

구분	설명
정하는 시점	학습 전에 정한다
값의 성격	하이퍼파라미터
선택 기준	검증 성능 비교

층 수와 유닛 수를 정한 뒤, 그 구조에서 가중치와 절편을 학습한다

MLP는 순방향으로 값을 계산하는 신경망이다

명칭	범위와 특징
MLP	은닉층을 포함하는 다층 퍼셉트론
순방향 신경망	이전 층으로 되돌아가는 연결 없이 값을 계산
신경망(neural network)	MLP를 포함하는 더 넓은 모델 범주

MLP는 다층 퍼셉트론(multilayer perceptron)의 약자이며, 순방향 신경망(feedforward network)의 한 종류다

Checkpoint

Adding a Hidden Layer

- 은닉층은 무엇을 하는 층인가?
- 은닉 유닛 하나는 어떤 단층 계산과 같은 모양인가?
- W_1 의 행과 열은 각각 무엇의 수인가?

Why a Nonlinear Activation

비선형 활성화함수가 필요한 이유를 설명한다

활성함수가 없으면 여러 층도 하나의 선형식이 된다

상수항을 포함한 두 선형식을 합성해도 하나의 선형식으로 나타낼 수 있다

$$s = w_2^\top (W_1 x + b_1) + b_2 = \underbrace{(w_2^\top W_1)}_{\text{새 } w^\top} x + \underbrace{(w_2^\top b_1 + b_2)}_{\text{새 } b}$$

비선형 활성화함수가 없다면 층을 여러 번 합성해도 $w^\top x + b$ 로 정리된다

선형식 하나로는 XOR을 분류할 수 없다

확인 항목	결과
층 수 증가	비선형 함수가 없으면 표현력은 같음
합성된 식	단층 선형 모델
XOR 분류 결과	네 점을 모두 맞힐 수 없음

Caution 활성화함수 없이 층을 늘려도 XOR의 결정 경계는 직선 하나로 남는다

DEFINITION

활성함수(activation function)

각 유닛의 가중합 z 를 출력 $g(z)$ 로 바꾸는 함수

$$x \mapsto z = w^T x + b \mapsto h = g(z)$$

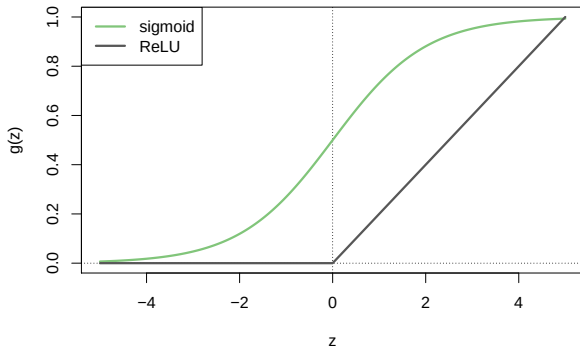
비선형 활성함수를 넣으면 선형식의 합성으로 표현할 수 없는 관계도 표현할 수 있다

sigmoid와 ReLU

이름	식	모양
sigmoid	$1/(1 + e^{-z})$	0과 1 사이의 S자
ReLU(rectified linear unit)	$\max(0, z)$	음수는 0, 양수는 그대로

sigmoid의 출력은 1보다 작지만 ReLU의 출력에는 상한이 없다

sigmoid와 ReLU의 모양



두 활성화함수의 모양이다. 그림의 ReLU 곡선은 sigmoid와 비교하기 쉽도록 $1/5$ 배 한 $\max(0, z)/5$ 다.

Source: James et al. (2023), Figure 10.2.

은닉층은 학습으로 새 특성을 만든다

은닉 유닛의 값은 입력에서 계산한 새로운 특성 표현이다

비교 항목	사람이 설계한 특성	은닉층의 특성
생성 방식	사람이 변환식을 정한다	가중치 학습으로 정해진다
예	곱 특성 $x_1 x_2$	$h = g(W_1 x + b_1)$ 의 각 성분

은닉층은 가중치 학습을 통해 입력을 어떤 중간값으로 변환할지 정한다

sigmoid와 ReLU의 미분 특성

비교 항목	sigmoid	ReLU
계산	지수 함수	크기 비교 하나
출력 범위	$(0, 1)$	$[0, \infty)$
큰 $ z $ 에서	도함수가 0에 가깝다	양수 쪽은 1로 일정

sigmoid의 도함수는 양끝에서 0에 가까워지고, ReLU의 도함수는 양수 구간에서 1, 음수 구간에서 0이다

은닉층과 출력층의 함수는 역할이 다르다

자리	목적	보기
은닉층	비선형 관계를 표현	ReLU 등
출력층	예측값의 형태를 결정	회귀는 선형 출력, 이진 분류는 sigmoid

로지스틱 회귀는 출력에 sigmoid를 적용하여 이진 범주의 확률을 구한다

Checkpoint

Why a Nonlinear Activation

- 활성화함수가 없으면 여러 층은 무엇과 같아지는가?
- g 에 요구되는 조건은 무엇인가?
- 은닉 유닛의 값을 새 특성이라고 부르는 이유는 무엇인가?

Section 3 • Multilayer Perceptrons

Forward Pass by Hand

2-2-1 구조로 XOR을 직접 계산한다

XOR을 푸는 2-2-1 신경망

입력 2개, 은닉 유닛 2개, 출력 1개인 구조에 숫자를 넣어 본다

유닛	식
은닉 유닛 1	$h_1 = \text{ReLU}(x_1 + x_2)$
은닉 유닛 2	$h_2 = \text{ReLU}(x_1 + x_2 - 1)$
출력	$s = h_1 - 4h_2 - 0.5$

주어진 가중치로 계산하는 예제이며, 학습 실험의 결과는 아니다. 출력은 점수 s 의 부호로 범주를 정하며, sigmoid를 붙여도 $\sigma(s) > 0.5$ 와 $s > 0$ 은 같은 판정이다

네 입력의 순전파 결과

x	h_1	h_2	s	예측
(0, 0)	0	0	-0.5	-1
(0, 1)	1	0	+0.5	+1
(1, 0)	1	0	+0.5	+1
(1, 1)	2	1	-2.5	-1

네 점의 예측이 모두 XOR의 정답 범주와 맞는다

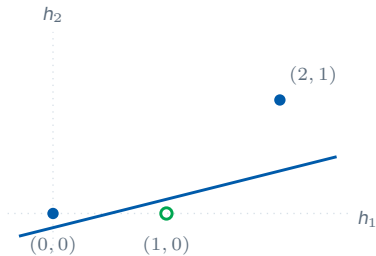
은닉층이 XOR을 분리 가능한 좌표로 바꾼다

은닉층을 지난 뒤의 좌표 (h_1, h_2) 에서 네 점의 배치를 비교한다

원래 x	은닉 좌표 (h_1, h_2)	정답 범주
$(0, 0)$	$(0, 0)$	-1
$(0, 1)$	$(1, 0)$	+1
$(1, 0)$	$(1, 0)$	+1
$(1, 1)$	$(2, 1)$	-1

은닉 좌표에서는 $(1, 0)$ 에 모인 +1 관측과 나머지 -1 관측을 직선으로 분리할 수 있다

은닉공간에서는 두 범주를 직선으로 나눌 수 있다



채운 점은 정답 범주 -1 , 빈 원은 $+1$ 이다. 출력식의 경계 $h_2 = 0.25h_1 - 0.125$ 가 두 정답 범주를 가른다

활성함수를 빼면 $(0, 0)$ 을 잘못 분류한다

x	s	예측	정답 범주
$(0, 0)$	$+3.5$	$+1$	-1
$(0, 1)$	$+0.5$	$+1$	$+1$
$(1, 0)$	$+0.5$	$+1$	$+1$
$(1, 1)$	-2.5	-1	-1

Caution ReLU를 빼면 식이 $-3(x_1 + x_2) + 3.5$ 로 정리되어 $(0, 0)$ 을 틀린다

ReLU가 음수인 중간값을 0으로 바꾼다

확인 항목	결과
(0, 0)에서 달라진 값	h_2 가 -1 이 아니라 0 이 된다
변화의 원인	ReLU가 음수를 0 으로 바꾼다
출력의 변화	$(0, 0)$ 의 점수가 3.5 에서 -0.5 로 바뀐다

같은 XOR 분류를 만드는 가중치는 여러 개다

$$s = h_1 - 4h_2 - 0.5, \quad s' = h_1 - 2h_2 - 0.5$$

입력	s	s'
(0, 0)	-0.5	-0.5
(0, 1), (1, 0)	0.5	0.5
(1, 1)	-2.5	-0.5

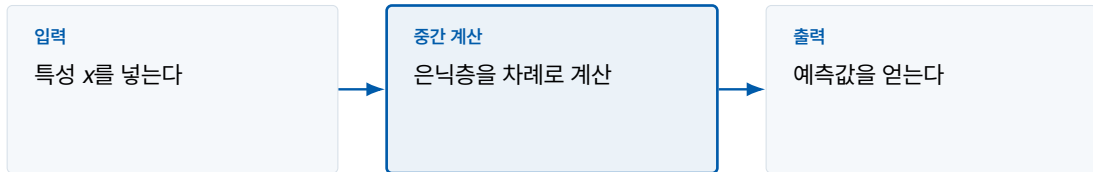
점수는 달라도 네 관측에서 부호가 같으면 같은 범주로 분류한다

순전파 계산

DEFINITION

순전파(forward pass)

입력부터 각 층의 값을 차례로 계산해 최종 예측값을 얻는 과정



학습에서는 순전파로 손실을 구한 뒤, 그 손실의 그래디언트로 가중치를 갱신한다

Checkpoint

Forward Pass by Hand

- $(1, 1)$ 에서 h_1 과 h_2 는 각각 얼마인가?
- 은닉공간에서 네 점은 어떻게 놓이는가?
- ReLU를 빼면 어느 점을 틀리고 왜 그런가?

Deeper and Wider

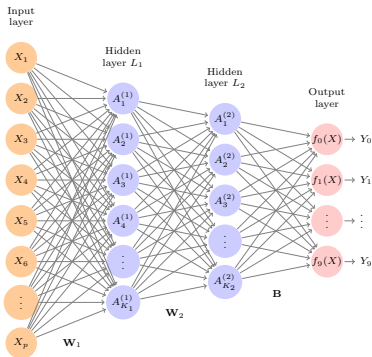
신경망의 깊이와 너비를 구분하고 매개변수 수를 센다

신경망은 넓게 또는 깊게 키운다

유닛을 늘리는 것과 층을 늘리는 것은 다른 일이다

키우는 방향	늘리는 것	뜻
넓게	한 층의 유닛 수	한 단계에서 더 많은 조합
깊게	층의 수	조합의 조합

은닉층이 둘인 신경망



첫 은닉층의 출력이 둘째 층의 입력이 된다. 이 그림의 출력은 손글씨 숫자 0부터 9까지를 구별한다.

Source: James et al. (2023), Figure 10.4.

매개변수를 세어 보면

완전 연결 층의 매개변수 수는 가중치 개수와 절편 개수의 합이다

연결	가중치 개수	절편 개수
입력층 → 은닉층	$K \times p$	K
은닉층 → 출력층	$1 \times K$	1

$$pK + K + K + 1 \quad 2 \times 2 + 2 + 2 + 1 = 9$$

층의 크기에 따라 매개변수 수가 늘어난다

구조	매개변수 수	계산 구성
$2 \rightarrow 2 \rightarrow 1$	9	가중치 6개, 절편 3개
$784 \rightarrow 256 \rightarrow 128 \rightarrow 1$	233,985	층별 가중치와 절편의 합(출력 1개)

앞 그림처럼 출력이 10개인 $784 \rightarrow 256 \rightarrow 128 \rightarrow 10$ 이면 235,146개다

매개변수가 많으면 각 편미분을 따로 계산하기보다 계산 과정을 재사용하는 방법이 필요하다

큰 신경망은 계산량과 과적합 가능성이 커진다

고려 항목	영향
더 큰 모델	더 복잡한 함수를 표현할 수 있음
학습 결과	최적화가 어렵거나 과적합할 수 있음
모델 선택	검증 결과와 계산 비용을 함께 비교

Caution 모델이 커져도 학습 손실이 반드시 줄거나 일반화 성능이 반드시 높아지지는 않는다

역전파로 구한 그래디언트를 최적화에 사용한다

계산	방법
가중치 갱신	SGD 등의 최적화 방법을 사용한다
그래디언트 계산	역전파를 사용한다

역전파는 그래디언트를 계산하고, 최적화 방법은 그 값으로 매개변수를 갱신한다

깊이와 너비의 효과는 함수에 따라 다르다

구성	표현과 학습에서 고려할 점
넓은 은닉층 하나	적절한 조건에서 복잡한 함수도 근사 가능
여러 은닉층	일부 함수는 더 적은 유닛으로 표현 가능
실제 학습	표현할 수 있다는 사실이 학습 성공을 보장하지 않음

Source: Zhang et al. (2023), Multilayer Perceptrons.

Checkpoint

Deeper and Wider

- 넓게 만드는 것과 깊게 만드는 것은 어떻게 다른가?
- $p \rightarrow K \rightarrow 1$ 구조의 매개변수 수는 얼마인가?
- 매개변수가 많아지면 무엇을 걱정해야 하는가?

How Do We Train All Those Weights?

Next Lecture: Backpropagation

- 계산 그래프와 연쇄법칙
- 출력에서 은닉층으로 미분값 전달
- 손계산 예제의 모든 매개변수에 대한 그래디언트