

2022학년도 1학기 컴퓨터언어학

제8강 신경망 언어 모형 (2)

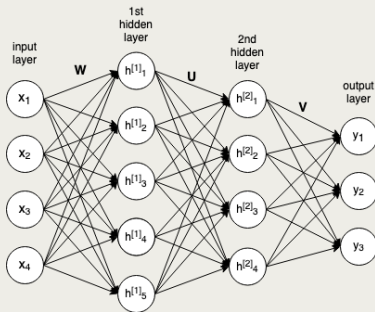
박수지

서울대학교 인문대학 언어학과

2022년 3월 30일 수요일

오늘의 목표

- 1 순방향신경망의 개념을 설명할 수 있다.
- 2 다층 신경망에서 은닉층의 활성화함수가 비선형함수라야 하는 이유를 설명할 수 있다.
- 3 아래의 그림이 무엇을 나타내며 각 부분이 무엇을 의미하는지 설명할 수 있다.



인공신경망(ANN: Artificial Neural Network)

순방향신경망(FFNN/FNN: Feed-forward Neural Network)

1 (다층)신경망의 핵심: 은닉층의 존재

- 하나의 은닉층은 (대개 여러 개의) 신경 단위로 이루어져 있다.
- 신경 단위(뉴런):

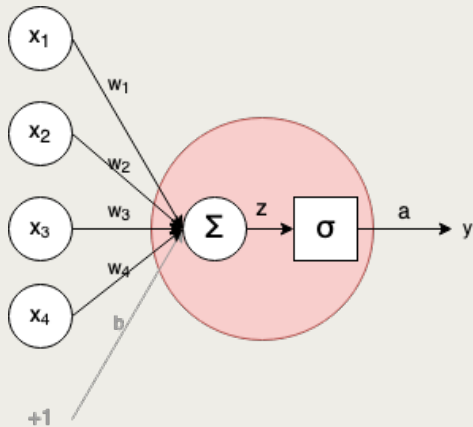
$$\text{입력} \cdots \vec{x} \xrightarrow[\text{linear function}]{\text{선형함수}} z = \vec{w} \cdot \vec{x} + b \xrightarrow[\text{activation function}]{\text{활성화함수}} a = f(z) \cdots \text{활성화값}$$

2 순방향신경망

- 입력층 → 출력층 사이에 은닉층을 한 방향으로 쌓아 나간다.

단층(1층) 신경망

로지스틱 회귀분석



예시: 입력의 특성값이 4개인 경우

입력 x_1, x_2, x_3, x_4

가중치 w_1, w_2, w_3, w_4

편향 b

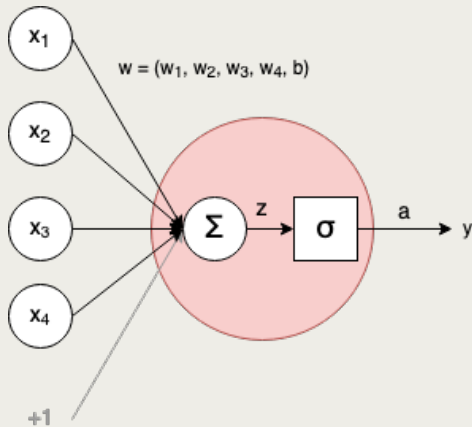
가중합 $z = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b$

활성화값 $a = \sigma(z)$

출력 $y = a$

단층(1층) 신경망

로지스틱 회귀분석

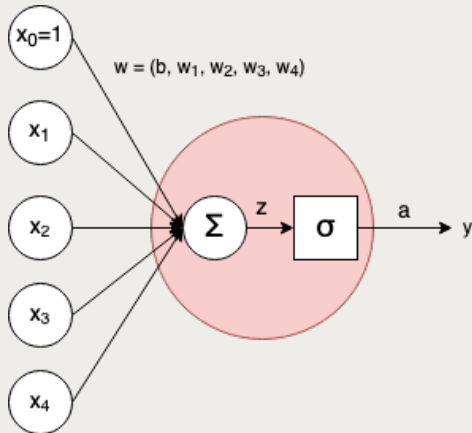


사실

한 신경 단위(뉴런)의 매개변수는 가중치 벡터 \vec{w} 와 편향 b 로 나타낼 수 있다.

단층(1층) 신경망

로지스틱 회귀분석



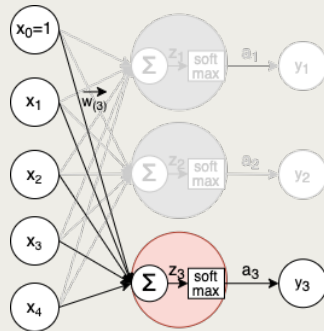
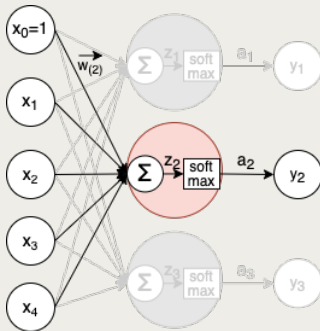
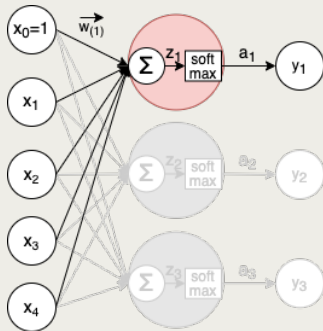
사실

한 단위의 매개변수는 가중치 벡터 \vec{w} 로 나타낼 수 있다. (7강 참조)

단층(1층) 신경망

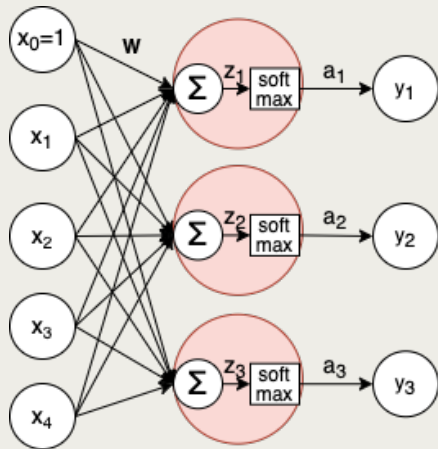
소프트맥스 회귀분석

- 출력층 뉴런의 개수는 데이터를 분류할 부류(class)의 개수와 같다.
- 뉴런마다 가중치 벡터를 가진다.



단층(1층) 신경망

소프트맥스 회귀분석



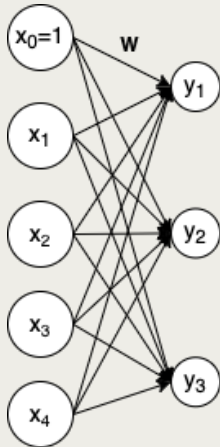
사실

출력층에서 j 번째 단위의 가중치 벡터를 $\vec{w}_{(j)}$ 라고 쓰면 ($1 \leq j \leq k$) 출력층의 가중치는

행렬 $\mathbf{W} = \begin{bmatrix} \vec{w}_{(1)} \\ \vdots \\ \vec{w}_{(k)} \end{bmatrix}$ 로 나타낼 수 있다.

단층(1층) 신경망

소프트맥스 회귀분석



신경 단위가 나올 때마다 가중합(Σ)과 활성화함수를 표시하기는 귀찮으므로 입력 노드와 출력 노드만 쓰자.

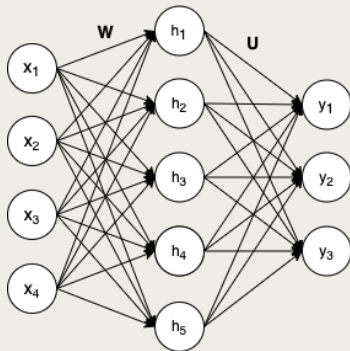
완전연결계층

이전 계층의 단위들이 다음 계층의 단위들과 모두 연결되어 있다.

- Fully-connected layer
- Dense layer
- Affine layer

2층 순방향 신경망

입력층은 층의 개수로 세지 않는다.



순방향신경망

단위들 사이의 순환(cycle)이 존재하지 않는 다층 신경망.

- 이전(낮은) 계층 단위의 출력은 다음(높은) 계층 단위로만 들어간다.

한 층의 가중치를 다음 층에서 재사용하지 않습니다.

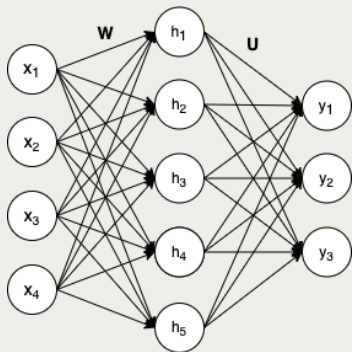
비교: 순환신경망

단위들 사이의 순환이 존재하는 다층 신경망.

하나의 가중치 행렬 W 를 반복해서 사용합니다.

2층 순방향 신경망

입력층은 층의 개수로 세지 않는다.



은닉층: 다층 신경망의 핵심

입력층과 출력층 사이를 연결하는 계층

2층 신경망의 구조

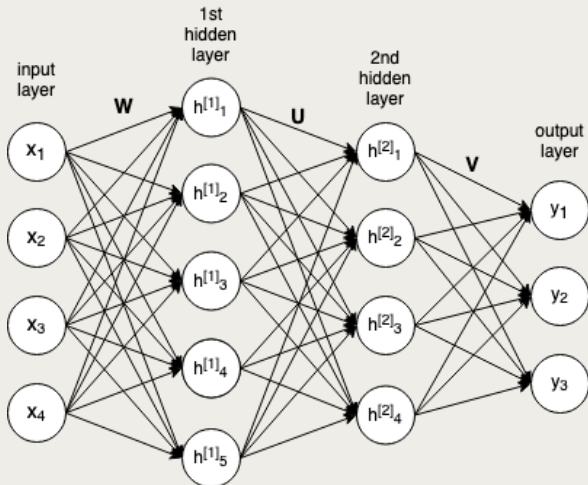
입력층 \vec{x}

은닉층 $\vec{h} = f(\mathbf{W}\vec{x})$

출력층 $\vec{y} = g(\mathbf{U}\vec{h})$

연습: 왼쪽 그림에서 가중치 행렬 \mathbf{W} 과 \mathbf{U} 의 크기는?

3층 순방향 신경망



3층 신경망의 구조

입력층 \vec{x}

은닉층1 $\vec{h}^{[1]} = g^{[1]}(\mathbf{W}\vec{x})$

은닉층2 $\vec{h}^{[2]} = g^{[2]}(\mathbf{U}\vec{h}^{[1]})$

출력층 $\vec{y} = g^{[3]}(\mathbf{V}\vec{h}^{[2]})$

활성화함수의 선택

은닉층 ReLU, tanh 등

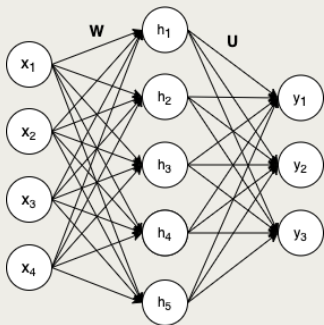
출력층 softmax (분류 문제)

다층 신경망의 활성화함수

- 1 분류 문제에서 출력층의 활성화함수는 일반적으로 softmax를 사용한다.
- 2 회귀 문제에서 출력층의 활성화함수는 선형함수가 될 수 있다.
- 3 은닉층의 활성화함수는 비선형함수만 유효하다.

비선형 활성화함수의 필요성

선형함수들끼리 합성하면 선형함수가 되기 때문이다.



- 1층의 활성화함수 $g^{[1]}(\vec{z}^{[1]})$ 가 선형이라고 가정하자.
- 그러면 $g^{[1]}(\vec{z}^{[1]}) = \mathbf{A}\vec{z}^{[1]}$ 를 만족하는 행렬 \mathbf{A} 가 존재한다.
(편향 \vec{b} 는 편의상 생략)
- 그러면 아래의 순서로 계산이 진행된다.

1층 입력 \vec{x}

1층 가중합 $\vec{z}^{[1]} = \mathbf{W}\vec{x}$

1층 활성화값 $\vec{a}^{[1]} = g^{[1]}(\vec{z}^{[1]}) = \mathbf{A}\vec{z}^{[1]}$

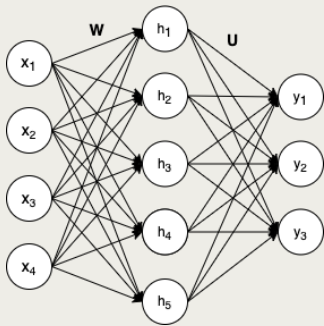
2층 가중합 $\vec{z}^{[2]} = \mathbf{U}\vec{a}^{[1]}$

2층 활성화값 $\vec{a}^{[2]} = g^{[2]}(\vec{z}^{[2]})$

- 4 2층 가중합 $\vec{z}^{[2]} = \mathbf{U}\vec{a}^{[1]}$ 을 자세히 살펴보자.
- $$\vec{z}^{[2]} = \mathbf{U}\vec{a}^{[1]} = \mathbf{U}(\mathbf{A}\vec{z}^{[1]}) = \mathbf{U}(\mathbf{A}(\mathbf{W}\vec{x})) = (\mathbf{UAW})\vec{x}$$

비선형 활성화함수의 필요성 (계속)

선형함수들끼리 합성하면 선형함수가 되기 때문이다.



4 2층 가중합 $\vec{z}^{[2]} = \mathbf{U}\vec{a}^{[1]}$ 을 자세히 살펴보자.

$$\vec{z}^{[2]} = \mathbf{U}\vec{a}^{[1]} = \mathbf{U} \left(\mathbf{A}\vec{z}^{[1]} \right) = \mathbf{U} \left(\mathbf{A} (\mathbf{W}\vec{x}) \right) = (\mathbf{UAW}) \vec{x}$$

(행렬의 곱셈은 결합법칙이 성립하기 때문)

5 2층 가중합은 사실상 1층 입력 \vec{x} 에 하나의 가중치 행렬 \mathbf{UAW} 를 곱한 1층의 가중합이 된다.

6 그러므로 1층 은닉층이 활성화함수로 선형함수를 사용하는 경우 1층과 2층이 합쳐져서 하나의 층이 된다.

다음 주에 배울 것

- 순방향신경망 모델을 훈련시키는 방법
- 순방향신경망 모델을 자연어처리에 활용하는 방법