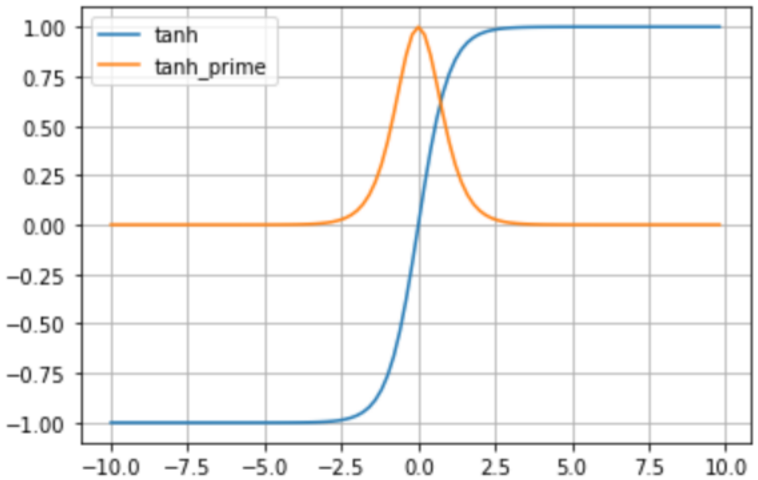
일석이조 조별보고서	
작성일 : 2023년 10월 10일	작성자 : 탁성재
조 모임 일시 : 10월 10일	모임 장소 : 구글미트
참석자 : 이준용, 유정훈, 김동규, 이학빈, 탁성재	조원 : 이준용, 유정훈, 김동규, 이학빈, 탁성재

구분	내용
학습 범위와 내용	1. 6주차 온라인 강의 내용 2. 교재 3장 내용
논의 내용 (조별 모임 전에 조장이 지시)	3.1.1. 인공신경망과 생물신경망 3.1.2. 신경망의 간략한 역사 3.1.3. 신경망의 종류 3.2.1. 구조 2.1.2. 동작
질문 내용 (모임 전 공지된 개별 학습 범위에서 이해된 것과 못한 것)	Q(1) 교수님강의에서 로지스틱 시그모이드 함수와 하이퍼볼릭 탄젠트 함수의 1 차 도함수를 구해보았다.

표 3-1 활성함수로 사용되는 여러 함수

함수 이름	함수	1차 도함수	범위
계단	$\tau(s) = \begin{cases} 1 & s \geq 0 \\ -1 & s < 0 \end{cases}$	$\tau'(s) = \begin{cases} 0 & s \neq 0 \\ \text{불가} & s = 0 \end{cases}$	-1과 1
로지스틱 시그모이드	$\tau(s) = \frac{1}{1+e^{-as}}$	$\tau'(s) = a\tau(s)\big(1-\tau(s)\big)$	(0,1)
하이퍼볼릭 탄젠트	$\tau(s) = \frac{2}{1+e^{-as}} - 1$	$\tau'(s) = \frac{a}{2}(1-\tau(s)^2)$	(-1,1)
소프트플러스	$\tau(s) = \log_e(1+e^s)$	$\tau'(s) = \frac{1}{1+e^{-s}}$	(0, ∞)
렉티파이어(ReLU)	$\tau(s) = \max(0, s)$	$\tau'(s) = \begin{cases} 0 & s < 0 \\ 1 & s > 0 \\ \text{불가} & s = 0 \end{cases}$	[0, ∞)

	A(1) 하이퍼볼릭 탄젠트 도함수 $\mathrm{f(x)} = \tanh(\mathrm{x})$ $= \dfrac{\sinh(\mathrm{x})}{\cosh(\mathrm{x})}$ $* \ \left(\dfrac{f}{g}\right)' = \dfrac{fg' - gf'}{g^2}$ $* \ \sinh(\mathrm{x})'' = \cosh(\mathrm{x})' = \sinh(\mathrm{x})$ $* \ cosh^2(x) - sinh^2(x) = 1$  하이퍼블릭 탄젠트 그래프

로지스틱 시그모이드의 도함수 추출

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

$$\frac{\mathrm{d}}{\mathrm{dx}}\mathrm{Sigmoid(x)} = \frac{\mathrm{d}}{\mathrm{dx}}(1+e^{-x})^{-1}$$

$$= (-1)\,\frac{1}{(1+e^{-x})^2}\frac{\mathrm{d}}{\mathrm{dx}}(1+e^{-x})$$

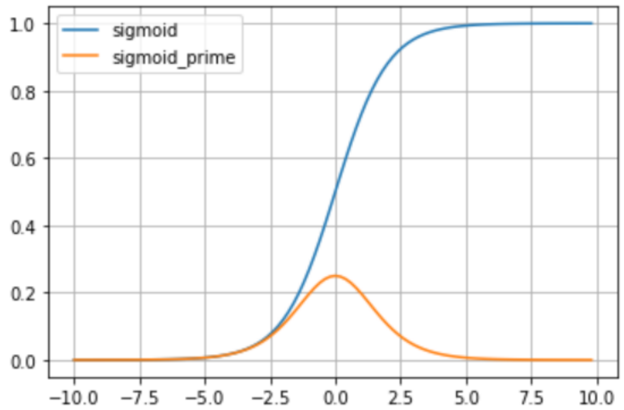
$$= (-1)\,\frac{1}{(1+e^{-x})^2}e^{-x}\frac{\mathrm{d}}{\mathrm{dx}}(-\mathrm{x})$$

$$= (-1)\,\frac{1}{(1+e^{-x})^2}-e^{-x}$$

$$= \frac{e^{-x}}{(1+e^{-x})^2} = \frac{1+e^{-x}-1}{(1+e^{-x})^2} = \frac{1+e^{-x}}{(1+e^{-x})^2} - \frac{1}{(1+e^{-x})^2}$$

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

$$= \mathrm{sigmoid(x)(1\text{-}sigmoid(x))}$$

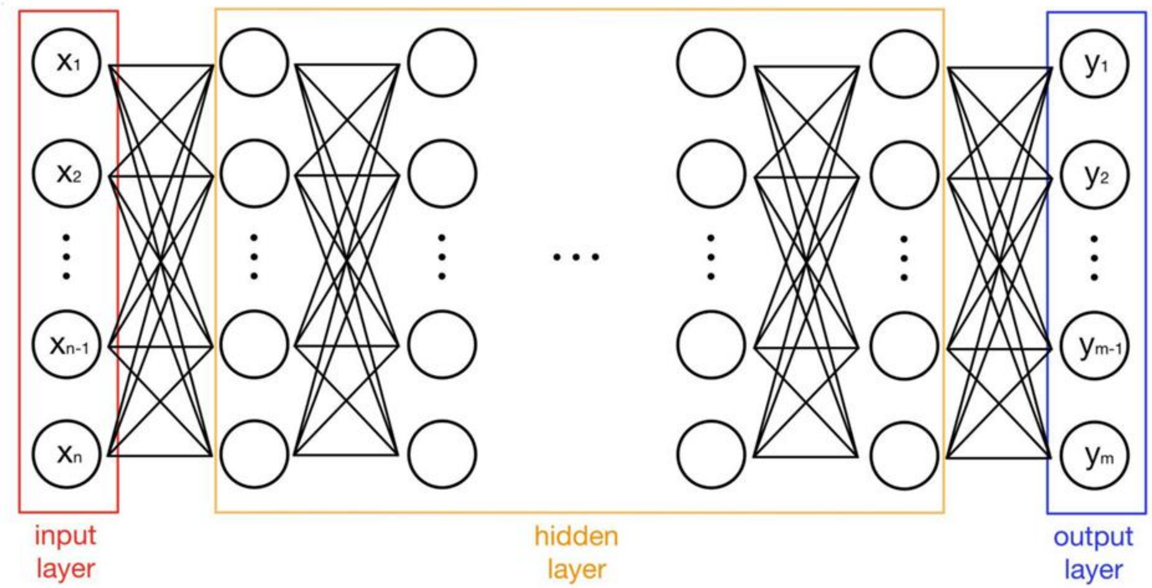


시그모이드

	β=1 일 때 softplus 를 미분하면 sigmoid 가 됩니다. $$\frac{\partial}{\partial x}\zeta(x) = \frac{\partial}{\partial x}\left(\frac{1}{\beta}\log(1+\exp(\beta\cdot x))\right)$$ $$= \frac{\partial}{\partial x}\log(1+\exp(x))$$ $$= \frac{\exp(x)}{1+\exp(x)} = \frac{1}{1+\exp(-x)}$$ Q(2) 다중 퍼셉트론의 구조 A(2) 입력층은 입력된 데이터를 표현하는 특징 벡터와 바이어스 노드를 통해 데이터가 입력됩니다. 입력된 데이터는 U1 이라는 입력층과 은닉층을 연결하는 가중치 벡터에 의해 계산되어 다음 은닉층으로 데이터를 이동시킵니다.

	은닉층은 계산값이 외부로 보여지지 않으며, 각 노드는 사용자가 갯수를 지정할 수 있으며, 이를 통해 해당 퍼셉트론의 용량이 규정될 수 있습니다. 해당 갯수를 적당히 조절해야 과잉적합, 과소적합을 예방할 수 있습니다. 출력층은 은닉층들의 결과를 합하여 활성함수의 조건에 따라 결과를 출력하는 층입니다. 마지막 은닉층과 출력층은 U2 라는 가중치 벡터를 사용해 규정됩니다. 은닉층의 노드 갯수와 가중치 벡터는 흔히 f(x)로 표현됩니다. 각 층은 서로 다른 f(x)를 가지고 있으며, 2 중 퍼셉트론인 경우 f2(f1(x)), 3 중 퍼셉트론인 경우 f3(f2(f1(x)))의 형태로 표현됩니다.

Q(3) 다층퍼셉트론은 무엇인가요?



A(3)

다층 퍼셉트론은 층이 2 개 이상 존재하는 신경망입니다. 입력층과 출력층을 제외한 층을 은닉층이라고 합니다. 이때 은닉층이 1 개만 있는 다층 퍼셉트론의 경우 얕은 신경망, 은닉층이 2 개 이상인 경우 깊은 신경망라고 합니다. 그리고 이 깊은 신경망을 학습시키는 것을 딥러닝이라고 합니다.

다층 퍼셉트론은 단층 퍼셉트론과 달리 비선형으로 분포하는 데이터들에 대해 학습이 가능합니다. 다층 퍼셉트론은 가중치에 대해 선형 방정식을 계산하기 때문에 층과 층 사이에서 선형으로 표현된 데이터를 비선형으로 바꿔주는 과정이 필요합니다. 이 과정을 담당하는 것이 바로 활성화 함수입니다.

Q(4) 다층퍼셉트론의 핵심 아이디어 보충 조사

A(4)

다층 퍼셉트론(Multilayer Perceptron, MLP)은 인공 신경망(ANN)의 한 유형으로, 여러 개의 뉴런(또는 노드)이 여러 층으로 구성된 구조를 가지고 있습니다. 다층 퍼셉트론의 핵심 아이디어는 복잡한 비선형 문제를 해결하기 위해 여러 층의 뉴런을 사용하여 입력 데이터로부터 유용한 표현을 학습하는 것입니다. 다음은 다층 퍼셉트론의 핵심 아이디어와 관련된 주요 개념과 구성 요소에 대한 자세한 설명입니다.

1. **입력층 (Input Layer)**:
   - 다층 퍼셉트론의 첫 번째 층으로, 입력 데이터를 받아들입니다.
   - 각 입력 변수는 입력 뉴런에 대응되며, 입력 뉴런의 개수는 입력 데이터의 특성 수에 따라 결정됩니다.
2. **은닉층 (Hidden Layers)**:
   - 하나 이상의 은닉층이 다층 퍼셉트론에 포함됩니다. 이 층들은 입력층과 출력층 사이에 있으며, 신경망의 핵심적인 역할을 합니다.
   - 은닉층은 여러 개의 뉴런으로 구성되며, 이 뉴런들은 각각 가중치와 활성화 함수를 가집니다.
   - 은닉층의 뉴런들은 입력 데이터로부터 중간 표현을 학습하고, 이 중간 표현은 점차적으로 더 복잡한 특징을 추출하기 위해 여러 은닉층을 거쳐 강화됩니다.
3. **가중치 (Weights)**:
   - 각 뉴런은 입력과 연결된 가중치를 가지고 있습니다. 이 가중치는 입력 신호에 대한 중요성을 나타내며, 학습 과정에서 조정됩니다.
   - 각 뉴런의 출력은 입력과 가중치의 선형 조합으로 계산됩니다.
4. **활성화 함수 (Activation Function)**:
   - 은닉층과 출력층의 각 뉴런은 활성화 함수를 사용하여 출력을 계산합니다.

- 활성화 함수는 비선형성을 도입하며, 신경망이 복잡한 함수를 모델링할 수 있게 합니다.
- 주로 사용되는 활성화 함수에는 시그모이드, 하이퍼볼릭 탄젠트, 렐루(Rectified Linear Unit, ReLU) 등이 있습니다.

5. **출력층 (Output Layer)**:
   - 출력층은 최종 예측값을 생성하는 역할을 합니다.
   - 출력층의 뉴런 수 및 활성화 함수는 특정 작업(회귀, 분류 등)에 따라 다르게 정의됩니다.
6. **학습 (Training)**:
   - 다층 퍼셉트론은 주어진 입력 데이터와 실제 출력 사이의 오차를 최소화하기 위해 학습됩니다.
   - 역전파(backpropagation) 알고리즘을 사용하여 오차를 각 뉴런과 가중치로 다시 전파하고, 가중치를 업데이트합니다.
   - 이러한 반복적인 과정을 통해 네트워크는 데이터에 대한 적절한 표현을 학습하게 됩니다.
7. **일반화 (Generalization)**:
   - 학습된 다층 퍼셉트론은 학습 데이터에 대한 예측 뿐만 아니라 새로운, 이전에 보지 못한 데이터에 대한 예측도 수행할 수 있어야 합니다.
   - 이를 통해 모델이 일반화된 특징을 학습하게 됩니다.

다층 퍼셉트론은 복잡한 비선형 문제를 해결하는 데 매우 강력한 도구로 사용되며, 딥러닝의 핵심 구성 요소 중 하나입니다. 그러나 많은 뉴런과 은닉층, 데이터 양, 학습 알고리즘의 조절이 필요할 수 있으며, 이를 튜닝하여 원하는 작업에 맞게 다층 퍼셉트론을 설계하고 학습시키는 것이 중요합니다.

	Q(5) 다층퍼셉트론으로 해결할 수 있는 문제의 예시
	
	A(5)
	
	다층 퍼셉트론(MLP)은 다양한 유형의 문제를 해결할 수 있는 강력한 기계 학습 모델입니다. 아래는 MLP 를 사용하여 해결할 수 있는 문제의 예시입니다:
	
	1. **이진 분류 (Binary Classification)**:
	    o MLP 는 두 개의 클래스로 데이터를 분류하는 이진 분류 문제에 사용될 수 있습니다. 예를 들어, 스팸 메일 여부를 판단하거나 의료 진단에서 질병의 발생 여부를 예측하는 등의 문제에 적용할 수 있습니다.
	2. **다중 클래스 분류 (Multiclass Classification)**:
	    o MLP 는 세 개 이상의 클래스로 데이터를 분류하는 다중 클래스 분류 문제에도 사용됩니다. 손글씨 숫자 인식, 이미지 분류, 자연어 처리에서의 텍스트 분류 등이 해당합니다.
	3. **회귀 (Regression)**:
	    o MLP 는 연속적인 출력 값을 예측하는 회귀 문제에도 사용됩니다. 주택 가격 예측, 주식 가격 예측, 날씨 예측 등이 회귀 문제의 예시입니다.
	4. **시계열 예측 (Time Series Prediction)**:
	    o MLP 는 시계열 데이터에서 다음 시간 단계의 값을 예측하는 데 사용됩니다. 주식 가격 예측, 기상 예측, 트래픽 예측 등의 문제에 적용 가능합니다.
	5. **이미지 처리 (Image Processing)**:
	    o 이미지 분류, 객체 검출, 얼굴 인식, 이미지 생성 등의 작업에 MLP 를 사용할 수 있습니다. 그러나 이를 위해 보다 복잡한 신경망 구조나 CNN(합성곱 신경망)과 같은 특별한 구조를 사용하는 것이 더 효과적입니다.

6. **자연어 처리 (Natural Language Processing, NLP)**:
   - 텍스트 분류, 감정 분석, 기계 번역, 질의 응답 시스템 등 다양한 NLP 작업에서 MLP 를 사용할 수 있습니다. 그러나 최근에는 순환 신경망(RNN)과 트랜스포머(Transformer)와 같은 특화된 구조가 더 일반적으로 사용됩니다.
7. **추천 시스템 (Recommendation Systems)**:
   - MLP 를 사용하여 사용자에게 개인화된 제품, 서비스 또는 콘텐츠를 추천하는 추천 시스템을 구축할 수 있습니다.
8. **생성 모델 (Generative Models)**:
   - Variational Autoencoders(VAE)나 Generative Adversarial Networks(GAN)과 같은 MLP 를 사용한 생성 모델은 이미지, 텍스트, 음악 등의 데이터를 생성하는 데 사용됩니다.

MLP 는 다양한 문제에 적용할 수 있지만 데이터의 복잡성과 다양성에 따라 모델 아키텍처와 하이퍼파라미터를 조정해야 할 수 있습니다. 또한 최근에는 딥러닝에서 다른 신경망 구조, 전이 학습, 강화 학습 등이 더 많이 사용되고 있으므로 문제의 특성에 따라 적절한 모델을 선택하는 것이 중요합니다.

	Q(6) 로지스틱 시그모이드, 하이퍼볼릭 탄젠트 시그모이드, softplus 와 rectifier 의 특징에 대해서 알아봤습니다. A(6) 1. 로지스틱 시그모이드 (Logistic Sigmoid): • 로지스틱 시그모이드 함수는 입력 값을 0 과 1 사이의 값으로 압축합니다. • 주로 이진 분류 문제에서 출력 뉴런의 활성화 함수로 사용됩니다. • 그러나 입력 값이 큰 절대값을 가질 때 기울기 소실 문제가 발생할 수 있으며, 이는 심층 신경망에서 학습을 어렵게 만들 수 있습니다. 1. 하이퍼볼릭 탄젠트 시그모이드 (Hyperbolic Tangent Sigmoid): • 하이퍼볼릭 탄젠트 함수는 입력 값을 -1 과 1 사이의 값으로 변환합니다. • 로지스틱 시그모이드와 유사하지만, 출력 범위가 -1 에서 1 로 확장되었습니다. • 여전히 기울기 소실 문제가 발생할 수 있지만, 입력 값이 0 주변에서 대칭인 특성을 가집니다. 1. Softplus: • Softplus 함수는 입력 값을 부드럽게 변환하여 양수 값을 출력합니다. • 이 함수는 신경망에서 ReLU 함수와 유사한 역할을 하며, 부드럽게 증가하는 특징을 가집니다. • 소프트플러스는 음수 입력 값에 대해 0 이 아닌 출력을 생성하므로 미분 가능한 특성을 가집니다. 1. ReLU (Rectified Linear Unit):

	• ReLU 함수는 입력 값이 양수인 경우에는 입력 값을 그대로 출력하고, 음수인 경우에는 0 으로 출력합니다. • 단순하면서도 계산이 효율적이며, 기울기 소실 문제를 어느 정도 해결합니다. • 심층 신경망에서 많이 사용되며, 특히 컴퓨터 비전과 관련된 작업에서 성능이 우수한 결과를 보이는 경향이 있습니다. 각 활성함수는 다른 종류의 문제나 신경망 아키텍처에 더 적합할 수 있습니다. 선택은 주어진 작업과 데이터에 따라 달라질 수 있으며, 실험과 조정을 통해 최적의 활성함수를 결정하는 것이 중요합니다.
질문내용	