

14강. Deep Learning 1.

◆ 담당교수 : 김 동 하

■ 학습개요

본 강의에서는 인공지능의 핵심 기술인 딥러닝에 대해 학습한다. 딥러닝이란 무엇인지에 대해 배우고, 딥러닝의 기반이 되는 모형인 (심층)인공신경망 모형에 대해 학습한다. 더 나아가, (심층)인공신경망 모형을 학습할 때 사용하는 목적함수와 목적함수를 최소화하는 최적화 알고리즘인 역전파 알고리즘에 대해서도 다루도록 한다.

■ 학습목표

1	딥러닝의 개념에 대해 학습한다.
2	(심층)인공신경망모형에 대해 학습한다.
3	(심층)인공신경망모형을 학습할 때 사용하는 목적함수와 최적화 알고리즘에 대해 학습한다.

■ 주요용어

용어	해설
(심층)인공신경망	다양한 딥러닝 모형의 기반이 되는 모형으로, 입력값을 여러 단계의 비선형 변환을 통해 최종 함수값을 출력하는 형태로 이루어져 있다.
활성함수	(심층)인공신경망 함수에 비선형성을 추가하기 위해 사용하는 함수. 주로 sigmoid, tanh, relu 함수 등을 사용한다.
역전파 알고리즘	(심층)인공신경망을 학습하기 위해 목적함수를 최소화할 때 사용하는 최적화 알고리즘.
Keras 모듈	파이썬에서 딥러닝 모형을 구축하고 학습하여 활용할 수 있게 해주는 대표적인 모듈 중 하나.

■ 학습하기

01. 딥러닝이란

기계 학습 (Machine learning)

- 컴퓨터에 인공적인 학습 가능한 지능을 부여하는 것을 연구하는 분야.
 - > 아서 사무엘 : “기계 학습이란 기계가 일일이 코드로 명시하지 않은 동작은 데이터로부터 학습하여 실행할 수 있도록 하는 알고리즘을 개발하는 연구 분야.”

인공 신경망

- 인공 신경망(Artificial Neural Network)
 - > 기계 학습 분야에서 연구되고 있는 학습 모델 중 하나.
 - > 주로 패턴 인식에 쓰이는 기술로, 인간의 뇌의 뉴런과 시냅스의 연결을 프로그램으로 재현하는 것.
- 딥러닝(Deep Learning)은 심층 인공 신경망 모형을 기초로 해서 발전하였음.
 - > 심층 모형을 학습한다. -> 딥러닝

딥러닝의 응용 분야

- 이미지 분석
 - > 이미지 인식, 압축, 복원, 생성 등
 - > AlphaGo (Silver et al., 2016 & 2017)
 - > DeepDream (그림을 그리는 인공지능), Artistic style transfer (Gatys et al., 2015)
 - > 무인 자동차
- 텍스트 분석
 - > 파파고, 구글 번역기
 - > 챗봇
- 오디오 분석
 - > 음성 인식, 복원, 생성, 분해 등
- 인공지능 스피커, 챗봇, 비서



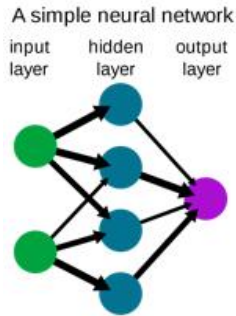
딥러닝의 구현

- Python 기반으로 다양한 모듈을 제공하고 있음.
- Tensorflow (Google)
- Keras (Google) - tensorflow, CNTK 기반.
- CNTK (Microsoft)
- Pytorch (Facebook)

02. 인공 신경망

인공 신경망 모형

- 인공 신경망 모형 (Artificial neural networks)
 - > 생물학의 뇌 구조 (신경망) 를 모방하여 만든 모형.
 - > 인간의 뇌가 문제를 해결하는 방식과 유사하게 구현.



인공 신경망 구조

- 1개의 중간층 혹은 은닉층 (hidden layer)을 갖는 (fully connected) neural networks

$R^P \rightarrow R^K$ 모형을 생각하자.

$$z_m^{(0)} = b_m^{(0)} + x^T w_m^{(0)}, \quad h_m = \sigma(z_m^{(0)}), \quad m = 1, \dots, M$$

$$z_k^{(1)} = b_k^{(1)} + h^T w_k^{(1)}, \quad f_k(x) = g_k(z_k^{(1)}), \quad k = 1, \dots, K$$

- $\sigma(\cdot)$: 활성화 함수 (activation function)

-> 비선형성을 갖도록 하는 함수

-> Sigmoid: $\sigma(v) = \frac{1}{1 + \exp(-v)}$

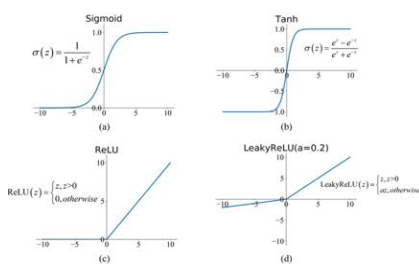
-> Tanh: $\sigma(v) = \frac{\exp(v) - \exp(-v)}{\exp(v) + \exp(-v)}$

- $g_k(\cdot)$: 출력 함수 (output function)

-> 회귀 분석: $g_k(t) = t$

-> 분류 문제: Softmax 함수 사용. $g_k(t) = \frac{\exp(t_k)}{\sum_{l=1}^K \exp(t_l)}$

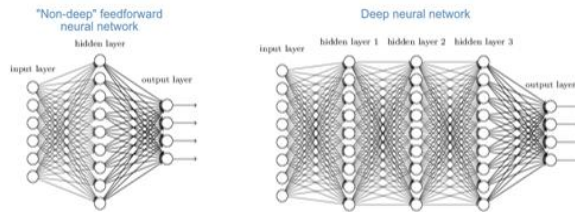
- 다양한 활성화 함수의 형태



03. 심층 인공 신경망

심층 인공 신경망 모형

- 심층 인공 신경망 모형 (Deep neural networks)
 - > 2개 이상의 중간층을 가지고 있는 신경망 모형



- L개의 중간층 (hidden layer) 을 갖는 (fully connected) neural networks 모형을 생각하자.
- $h^{(l)} \in R^{n_l}, l=0, \dots, (L+1)$: l번째 중간층의 노드 벡터
 - > $h^{(0)} = x, h^{(L+1)} = f(x)$
- 모형의 출력값은 다음과 같이 계산되어짐.
 - $z^{(l+1)} = b^{(l)} + W^{(l)}h^{(l)}, l=0, \dots, L$
 - $h^{(l+1)} = \sigma(z^{(l+1)}), l=0, \dots, L$
 - $f(x) = h^{(L+1)} = g(z^{(L+1)})$
- $\sigma(\cdot)$: 활성화 함수 (activation function), 비선형성을 갖도록 하는 함수.
 - > 은닉층의 수가 많을 경우 sigmoid보다는 ReLU를 많이 사용.
- $g(\cdot)$: 출력 함수 (output function)
 - > 회귀 분석: $g_k(t) = t$
 - > 분류 문제: softmax 함수 사용.

DNN vs. SVM

- 공통점
 - > Hidden feature를 사용.
 - > Hidden space 위에서 linear decision boundary를 이용하여 분류.
- 차이점
 - > SVM : 정해진 feature mapping을 사용하고, linear decision boundary만 학습.
 - > NN : linear decision boundary 뿐만 아니라 feature mapping도 함께 학습.

04. 인공신경망모형의 학습

목적 함수

- 학습해야 하는 모수: $(b^{(l)}, W^{(l)}), l=0, \dots, L$
 - > $\theta = \{(b^{(l)}, W^{(l)})\}_{l=0}^L$

- 훈련 데이터: $\{(x_i, y_i)\}_{i=1}^n$
- 본 강의에서는 지도 학습만을 다룰 예정.
- 분류 문제의 경우: Cross-entropy function

$$L(\theta) = - \sum_{i=1}^n \sum_{k=1}^K y_{ik} \cdot \log f_k(x_i)$$

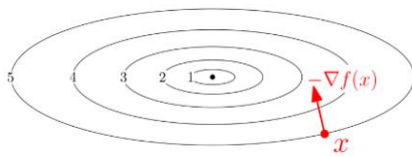
-> $y_i \in R^K$: one-hot vector

- 회귀 문제의 경우: Squared loss function

$$L(\theta) = - \sum_{i=1}^n \sum_{k=1}^K (y_{ik} - f_k(x_i))^2$$

목적 함수의 최적화: 기울기 강하 알고리즘

- Gradient descent algorithm
- 특정 목적 함수 $L(\theta)$ 를 최소화하는 θ 를 한 번에 찾기 힘든 경우에 사용하는 대표적인 반복 알고리즘.
- $-\frac{\partial L(\theta)}{\partial \theta} \big|_{\theta=\theta_0}$ 은 $\theta=\theta_0$ 일 때 목적 함수 $L(\theta)$ 를 가장 빠르게 감소시키는 방향이라는 점에서 착안.



- 알고리즘:
 - > 초기값 설정: $\theta = \theta_0$
 - > 현재의 값을 $\theta^{(t)}$ 라 할 때, 학습률 (learning rate) ϵ_t 를 이용하여

$$\theta^{(t+1)} \leftarrow \theta^{(t)} - \epsilon_t \cdot \frac{\partial L(\theta)}{\partial \theta} \big|_{\theta=\theta^{(t)}}$$
 으로 업데이트
 - > 수렴할 때까지 위의 업데이트 과정을 반복.

역전파 알고리즘

- Back propagation 알고리즘
- Deep neural network의 모수들의 gradient를 구하는 알고리즘
- 인공 신경망의 목적 함수 $L(\theta)$ 를 최소화하기 위해 기울기 강하 알고리즘을 적용하면, 인공 신경망 모형의 특수한 형태 때문에 $\frac{\partial L(\theta)}{\partial \theta^{(l+1)}}$ 의 계산에 필요한 값을 알고 있으면 $\frac{\partial L(\theta)}{\partial \theta^{(l)}}$ 이 자동적으로 계산됨.
- > $\theta^{(l)}$ 은 l 층에서의 모수

- 미분값이 위에서 아래로 계산되어짐
-> Back-propagation

■ 연습문제

(객관식)1. (심층)인공망모형에 대해 잘못 설명한 것을 고르시오.

- ① 생물학의 뇌 구조를 모방하여 만든 수학적 모형이다.
- ② 입력값에서 한 개 이상의 은닉층을 통과하여 최종 함수값이 계산된다.
- ③ 각 은닉층마다 비선형성을 추가하기 위해 활성화함수를 사용한다.
- ④ 분류 문제의 경우 보통 제곱 손실 함수를 목적함수로 사용한다.

정답 : ④

해설 : 분류 문제의 경우 주로 교차 엔트로피를 목적함수로 사용한다.

(주관식)2. 기울기 강하 알고리즘이 (심층)인공신경망 모형의 학습을 위해 사용될 때 모수의 미분값의 정보가 위에서 아래로 전파된다는 성질로부터 '이것'으로 불리기도 한다. '이것'은 무엇인가?

정답) 역전파 (Back-propagation) 알고리즘

해설) 미분값의 정보가 역전파된다고 하여 역전파 알고리즘으로 불린다.

(O/X)3. 은닉층의 수가 많은 심층인공신경망 모형에서 relu 활성화함수보다는 sigmoid 활성화함수가 일반적으로 더 좋은 성능을 가진다.

정답 : X

해설 : ReLU 활성화함수가 일반적으로 더 좋은 성능을 갖는다.

■ 정리하기

- 1. (심층)인공신경망모형은 입력값을 반복된 비선형 변환을 통해 최종 함수값을 출력한다.
- 2. 각 은닉층마다 비선형성을 추가하기 위해 활성화함수를 합성해주며, 주로 sigmoid, tanh, relu 등의 함수를 사용한다.

3. (심층)인공신경망모형의 학습을 위해 반복적으로 모수를 학습하는 역전파 알고리즘을 사용한다.

■ 참고자료 (참고도서, 참고논문, 참고사이트 등)

없음.