

Cs231n Lecture 4 Summary
Introduction to Neural Networks

- 1. Backpropagation**
- 2. Neural Networks**
- 3. Artificial Neural Network**

1. Backpropagation

Backpropagation 은 한국말로 역전파라고 하며, 알고리즘의 일종이다. 먼저 내가 뽑고자 하는 타겟값과 모델이 계산한 아웃풋값이 얼마나 차이가 나는지 오차를 계산한다. 그리고 그 오차를 다시 뒤로 전파해나가면서 각 노드가 가지고 있는 weight(가중치)값을 업데이트 하는 과정이다.

Weight 값을 계산할 때 미적분의 chain rule 이 사용된다. Chain rule 을 사용해서 local gradient 값을 계산하고 앞에서 미리 계산되어서 넘어온 global gradient 값을 곱해서 최종적으로 gradient 값을 계산한다. Global gradient 는 upstream gradient 라고도 부른다.

이때 일일이 계산하지 않고 활성화함수를 이용하면 더 빠르는데, 활성화함수의 종류에는 대표적으로 시그모이드 함수가 있다.

계산 과정에서 3 개의 gate 가 사용되는데, add gate 는 gradient 를 그대로 전해주고 max gate 는 두 값 중 큰 값을 전해준다. 마지막으로 mul gate 는 서로의 값을 교환한다.

만약 변수가 숫자가 아닌 벡터라면 gradient 는 자코비안 행렬 방식으로 나타내진다.

2. Neural Networks

W 가 행렬이고 x 가 입력 벡터일 때, $s = Wx$ 라는 식으로 class score 를 계산했다. 이것을 Linear score function 이라고 부른다.

만약 layer 가 두개인 Neural Network 라면 parameter 2 개는 gradient 로 학습시키고 그 gradient 는 chain rule 을 이용해서 계산해서 구한다.

3. Artificial Neural Network

인공 뉴런에서 정보를 받아들일 때에는 Input 을 받고 가중치 w 를 곱해서 활성화 함수를 지나 output 으로 나가는 형태를 띄게 된다.

하지만 실제 생물학적 뉴런은 훨씬 더 복잡한 연산을 수행하기 때문에 인공 뉴런과 실제 뉴런이 유사하다고 말하는 것에는 경계심이 필요하다.

이 활성화 함수에는 많은 종류가 있는데, sigmoid(시그모이드), tanh, ReLU 등등이 있지만 현재 가장 많이 사용하는 함수는 ReLU 이다.

참고로 Neural Network 에서 layer 의 기준은 weight 를 갖고 있는지 여부이다. 따라서 input layer 에는 weight 이 없으므로 총 3 단계여도 2-layer Neural Net 이라고 부른다. 더하여 모든 Node(노드)가 연결되어 있는 layer 를 Fully-connected layer, 줄여서 FC layer 라고 부른다. Network 를 layer 형태로 구성하는 이유는 계산이 효율적이기 때문이다.