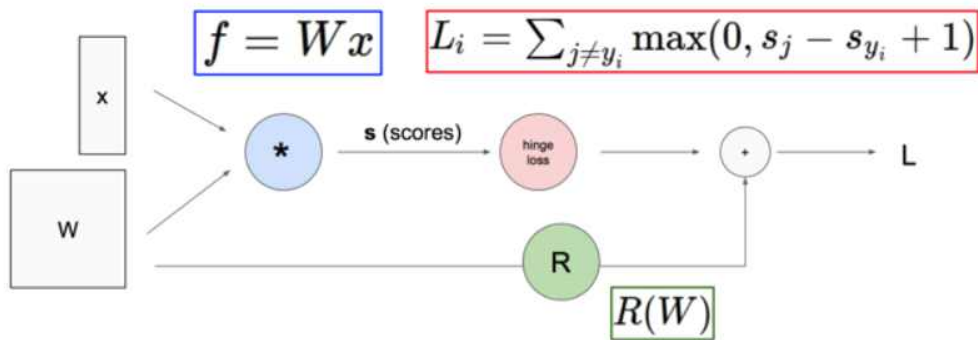


CS231n(4강)

최민우

1. Computational Graph(계산 그래프)

계산 그래프 (Computational Graphs)



- 어떤 함수를 표현하기 위해서, 이런 종류의 계산 그래프를 사용할 수 있다.
- 일단 계산 그래프로 함수를 표현하면, 역전파(back propagation)라고 불리는 테크닉을 사용할 수 있다.
- 반복적으로 체인 룰(chain rule)을 사용해서, 계산 그래프의 모든 변수에 대해 경사를 계산한다.

2. Back Propagation(역전파)

역전파 (backpropagation): 간단한 예

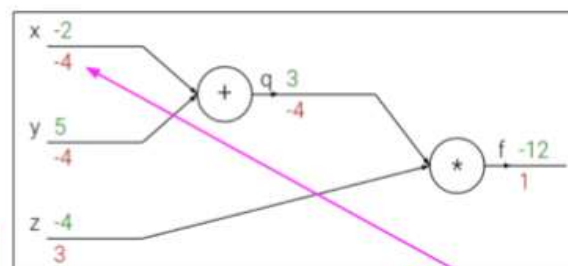
$$f(x, y, z) = (x + y)z$$

e.g. $x = -2, y = 5, z = -4$

$$q = x + y \quad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$

$$f = qz \quad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

원하는 것: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$

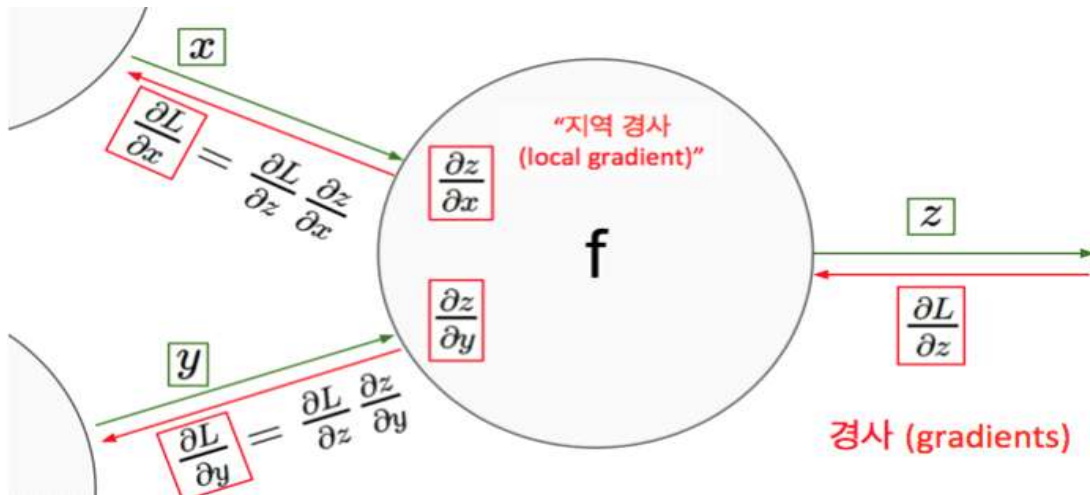


체인 룰 (Chain rule):

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial q} \frac{\partial q}{\partial x}$$

$$\frac{\partial f}{\partial x}$$

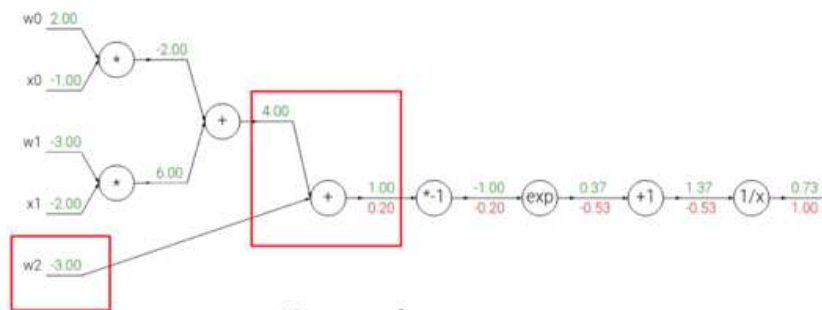
- 신경망의 각 노드가 가지고 있는 가중치(weight)와 편향(bias)을 학습시키기 위한 알고리즘



- 지역 경사(local gradient): x에 대한 z의 경사, y에 대한 z의 경사
- 업스트림 경사(upstream gradient): 각 노드에서 직접적인 출력에 대한 거꾸로 가는 경사

또 다른 예:

$$f(w, x) = \frac{1}{1 + e^{-(w_0x_0 + w_1x_1 + w_2x_2)}}$$



$f(x) = e^x$	\rightarrow	$\frac{df}{dx} = e^x$		$f(x) = \frac{1}{x}$	\rightarrow	$\frac{df}{dx} = -1/x^2$
$f_a(x) = ax$	\rightarrow	$\frac{df}{dx} = a$		$f_c(x) = c + x$	\rightarrow	$\frac{df}{dx} = 1$

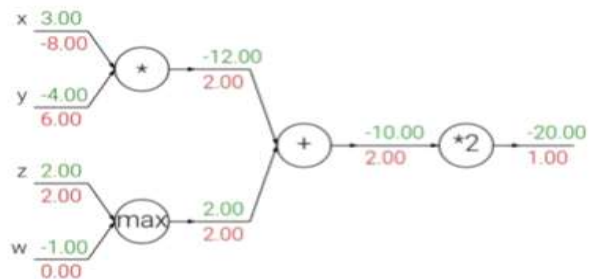
- 역전파(back propagation) 심화

역방향 흐름 속의 패턴

add 게이트:
경사 분배기 (gradient distributor)

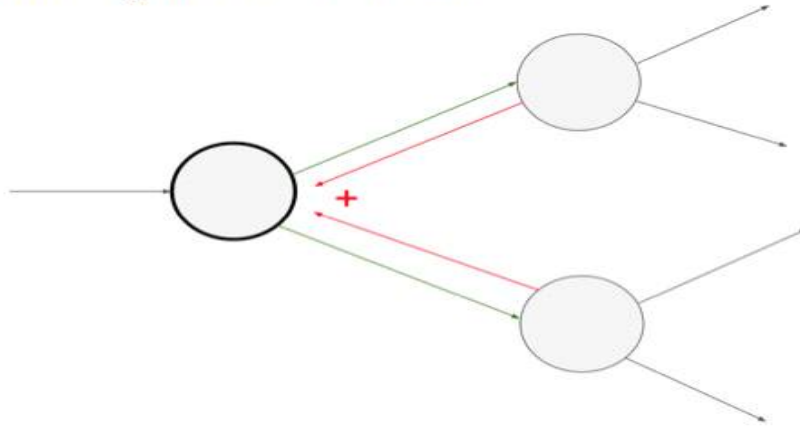
max 게이트:
경사 라우터 (gradient distributor)

mul 게이트:
경사 스위처 (gradient switcher)



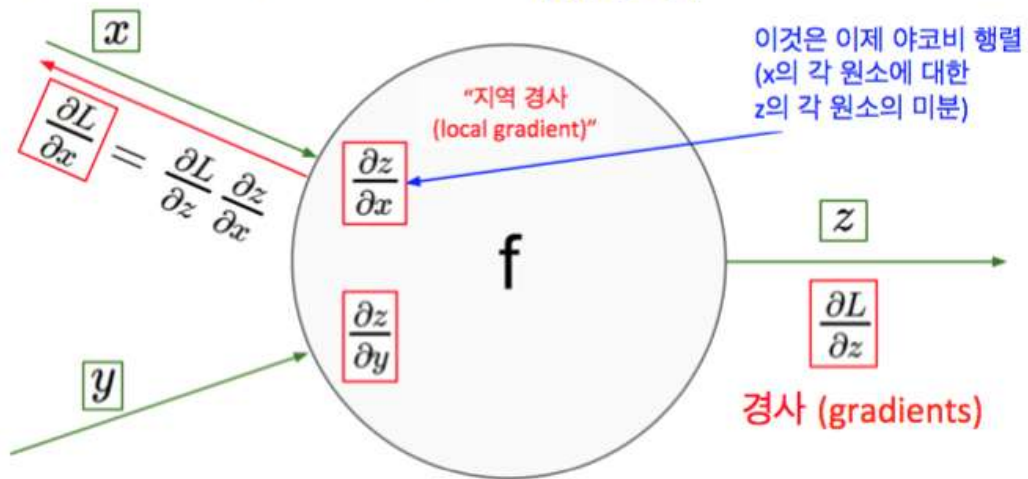
- add 게이트, max 게이트, mul 게이트(multiply: 곱하기)

가지 (branches)에서 경사 더하기



- 하나의 노드가 여러 개의 노드로 연결될 때, 이 노드에서 경사들이 더해진다.
- 가지(branches)에서 다변수 체인 룰(multi-variable chain rule)을 사용해서, 이 노드들에서부터 올라오는 업스트림 경사의 값을 계산한다.
- 업스트림 경사를 모두 더하여, 총 업스트림 경사(total upstream gradient)를 얻는다.

벡터화된 코드 (vectorized code)를 위한 경사 (x, y, z는 벡터)

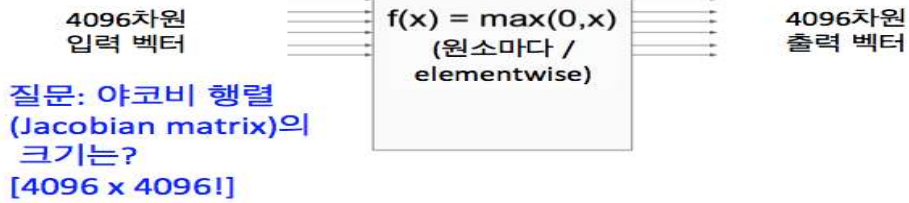


- 스칼라(scalar)인 경우는 다 봤고, 벡터(vector)인 경우를 살펴본다.
- 변수가 x, y, z는 단지 숫자가 아니라, 각각 벡터를 가진다.
- 전체적인 흐름은 동일하지만, 경사가 야코비 행렬(Jacobian matrix)이 된다.
- 모든 항목들의 미분을 포함하는 행렬이 된다.

벡터화된 연산 (vectorized operations)

$$\frac{\partial L}{\partial x} = \boxed{\frac{\partial f}{\partial x}} \frac{\partial L}{\partial f}$$

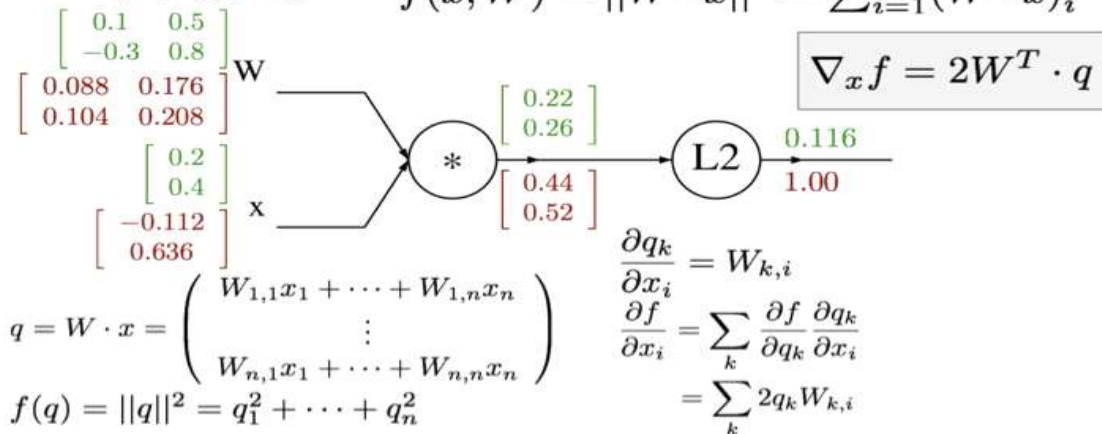
야코비 행렬



- 야코비 행렬의 계산

벡터화된 예:

$$f(x, W) = \|W \cdot x\|^2 = \sum_{i=1}^n (W \cdot x)_i^2$$



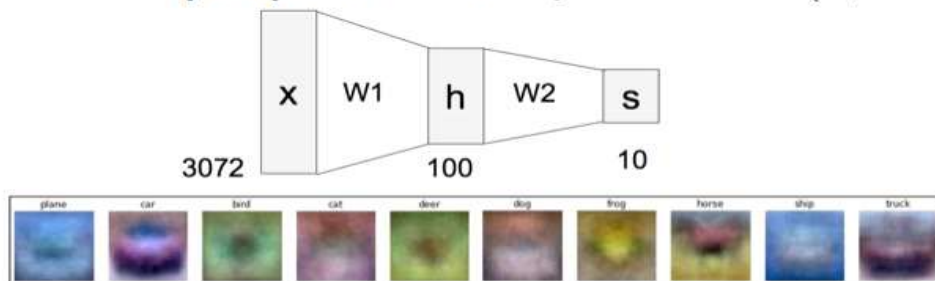
- 벡터화된 예

3. Neural Network(신경망)

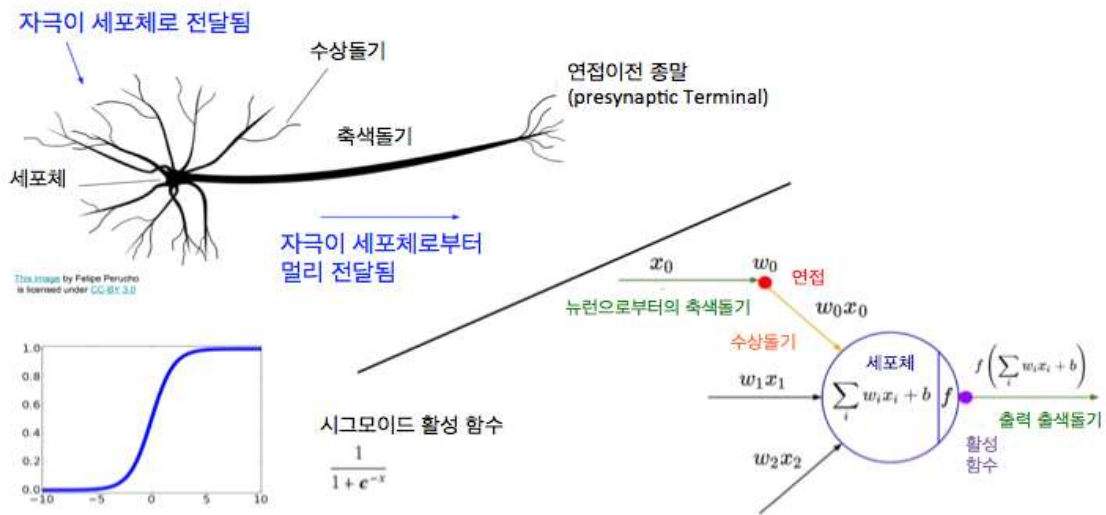
신경망: 뇌와 상관 없이

(이전) 선형 점수 함수: $f = Wx$

(이제) 2-층 신경망: $f = W_2 \max(0, W_1 x)$



- 생물학의 신경망 개념을 빌려왔다.
- 생물학적 지식을 필요로 하지 않는다.(함수로서의 신경망)
- 2개의 계층(Layer)으로 된 신경망

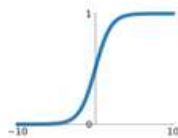


- 비유

활성 함수 (activation functions)

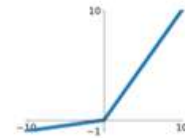
Sigmoid

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



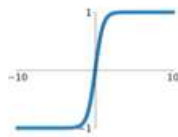
Leaky ReLU

$$\max(0.1x, x)$$



tanh

$$\tanh(x)$$



Maxout

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

ReLU

$$\max(0, x)$$



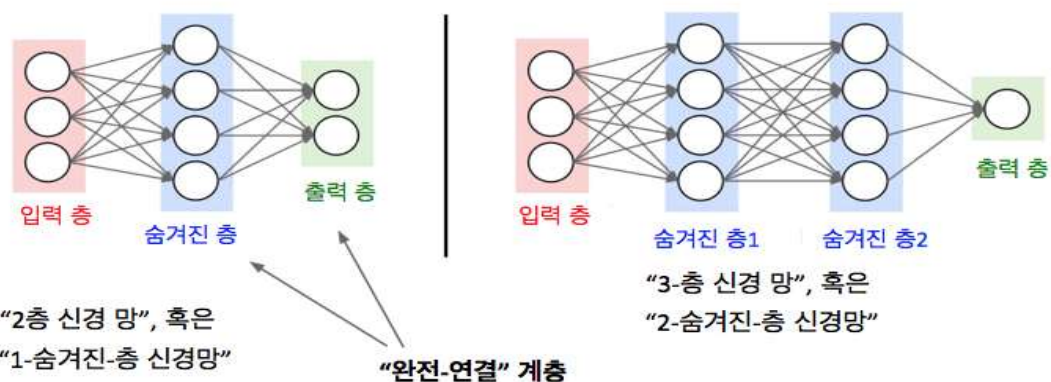
ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



- 활성화 함수

신경망: 아키텍처



- 완전 연결층(fully connected layer) : 2층 신경망, 3층 신경망