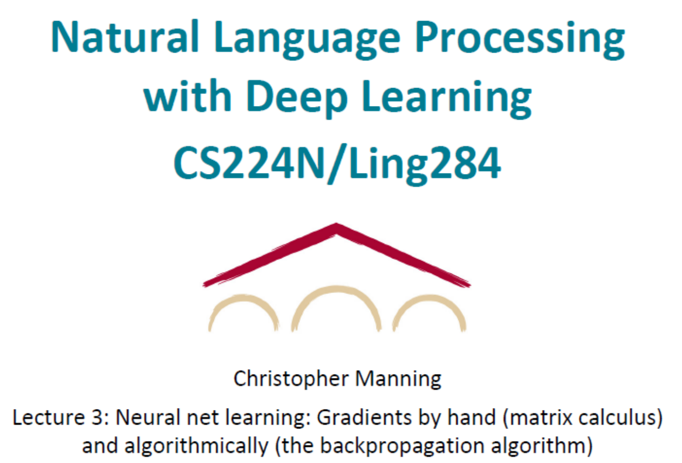
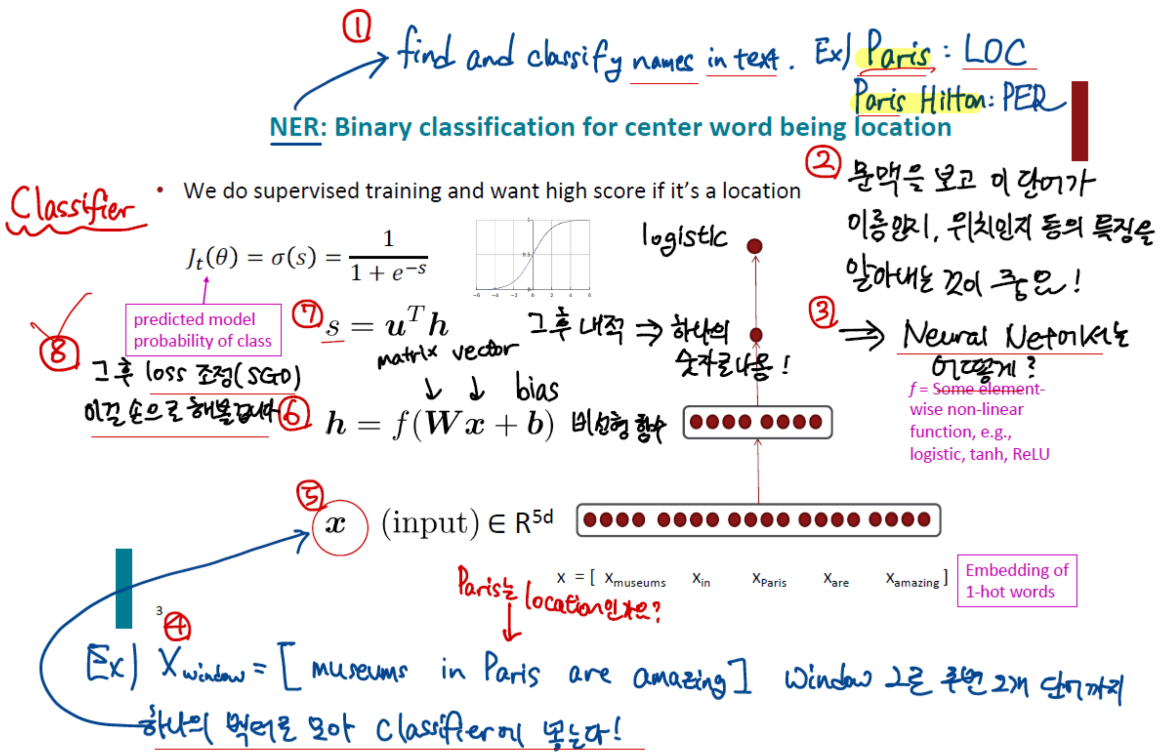
Week 3

발제자 : @손승진

> 
>
>
> 신경망 수학적으로 계산해볼것!
> 수학적인 요소 多

이번 주는 자연어처리에서 사용하는 신경망의 기본 개념에 대해서 배워보겠습니다

신경망의 수학적인 원리를 다룰 것이기 때문에 수학적인 요소가 많습니다



이 슬라이드는 앞으로 자연어처리에서 신경망을 어떻게 사용하는지 그 과정을 전반적으로 보여주는 중요한 슬라이드입니다

먼저 두 가지 문장을 보여드리겠습니다

1. I love Paris which is the capital city of France
2. Paris Hilton lives in London

첫 번째 문장의 Paris와 두 번째 문장의 Paris는 서로 다른 의미를 가진 단어입니다. 첫 번째 문장의 Paris는 위치를 나타내는 파리이고, 두 번째 문장의 Paris는 패리스 힐튼의 이름인 패리스를 나타내고 있습니다.

같은 단어이더라고 문맥에 따라 그 의미가 달라질 수 있는 것이죠. 그래서 자연어처리 모델은 문맥을 보고 이 언어가 이름인지, 위치인지 등의 특징을 알아내는 것이 중요합니다!

이를 신경망에서는 어떻게 할까요?

신경망에서는 문장을 하나의 벡터로 모아 분류기에 넣고, 비선형 함수에 넣은 다음 내적의 과정을 거쳐 예측을 진행한 후 loss를 조정하는 방식으로 학습을 진행하게 됩니다.

이 과정은 전반적인 과정을 보여준 것이므로 틈틈이 보면서 공부하면 좋을 것 같습니다

네 이제 그러면 이번주 강의 내용으로 다시 돌아와서

본격적으로 행렬 연산을 해보겠습니다

Jacobian Matrix: Generalization of the Gradient

일반화

input n개
output m개
⇒ Jacobian

- Given a function with ***m* outputs** and *n* inputs

$$\boldsymbol{f}(\boldsymbol{x}) = [f_1(x_1, x_2, ..., x_n), ..., f_m(x_1, x_2, ..., x_n)]$$

- It's Jacobian is an ***m* x *n* matrix** of partial derivatives

Jacobian Matrix

m × n matrix

↑ ↑

output input

$$\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

f_1을 x_1에 대해 미분

f_m을 x_n에 대해 미분

$$\boxed{\left(\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}\right)_{ij} = \frac{\partial f_i}{\partial x_j}}$$

13

Chain Rule

- For composition of one-variable functions: **multiply derivatives**

일변수 chain rule
도함수 곱하면 됨

$$z = 3y$$
$$y = x^2$$
$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx} = (3)(2x) = 6x$$

3 2x

- For multiple variables functions: **multiply Jacobians**

다변수 (벡터, 행렬)
일변수랑 똑같음.
Jacobian 곱하면 됨.

$$\boldsymbol{h} = f(\boldsymbol{z})$$
$$\boldsymbol{z} = \boldsymbol{W}\boldsymbol{x} + \boldsymbol{b}$$
$$\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{z}}\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}} = \dots$$

14

Jacobian Matrix를 계산하고 Chain rule을 이용해 원하는 도함수를 얻는 과정입니다

이 과정은 추후 오차역전파 알고리즘에서 많이 쓰이게 됩니다.

오차역전파

3. Backpropagation 오차역전파 계산은 앞에서 배웠다

We've almost shown you backpropagation

It's taking derivatives and using the (generalized, multivariate, or matrix) chain rule

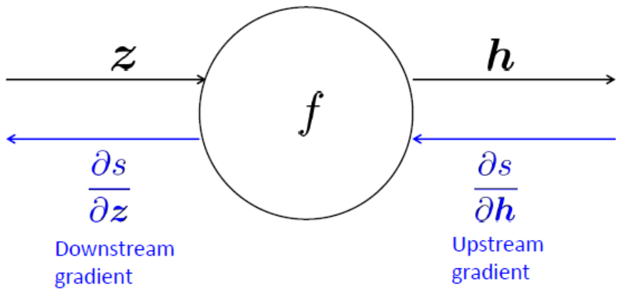
Other trick:

We **re-use** derivatives computed for higher layers in computing derivatives for lower layers to minimize computation

Backpropagation: Single Node

- Node receives an "upstream gradient"
- Goal is to pass on the correct "downstream gradient"

$$\boxed{\boldsymbol{h} = f(\boldsymbol{z})}$$



계산하고 싶은 것 노드가 받은 것

$\frac{\partial h}{\partial z}$? local gradient

51

→ Chain rule로 풀자!

$\frac{ds}{dz} = \frac{ds}{dh} \frac{dh}{dz}$ → upstream, → local gradient!) 둘 다 계산 가능

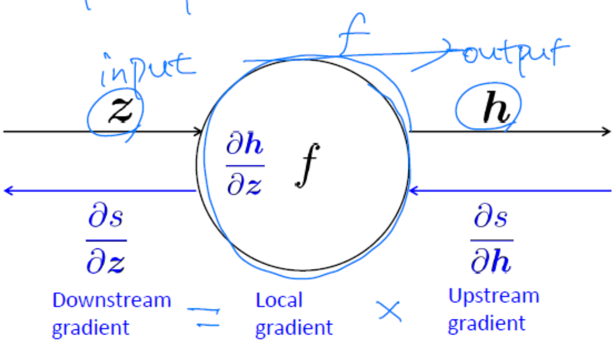
Backpropagation: Single Node

- Each node has a **local gradient**
  - The gradient of its output with respect to its input

$$\boxed{h = f(z)}$$

output을 input으로미분한것

h=f(z)를 z에 대해 미분

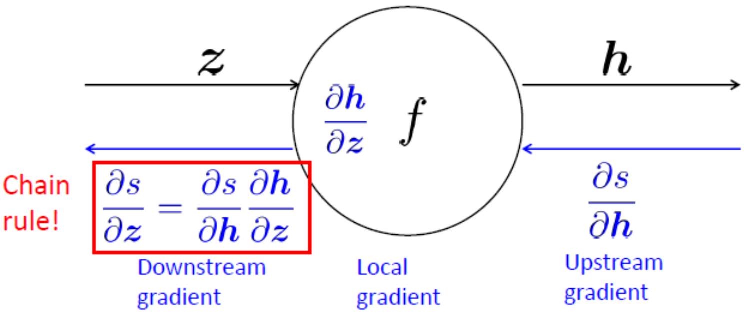


52

Backpropagation: Single Node

- Each node has a **local gradient**
  - The gradient of its output with respect to its input

$$\boxed{h = f(z)}$$



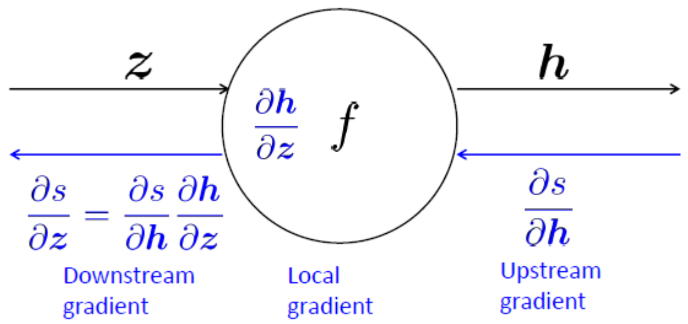
53

Backpropagation: Single Node

- Each node has a **local gradient**
  - The gradient of its output with respect to its input

$$\boxed{\boldsymbol{h} = f(\boldsymbol{z})}$$

- [downstream gradient] = [upstream gradient] x [local gradient]

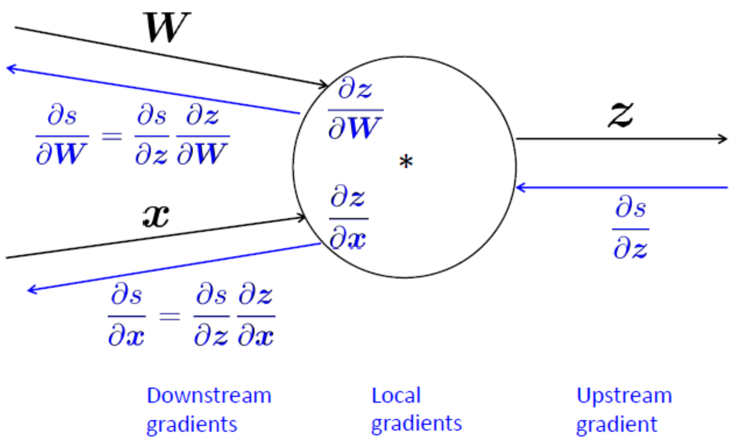


54

Backpropagation: Single Node

- Multiple inputs → multiple local gradients

$$\boxed{\boldsymbol{z} = \boldsymbol{W}\boldsymbol{x}}$$



56

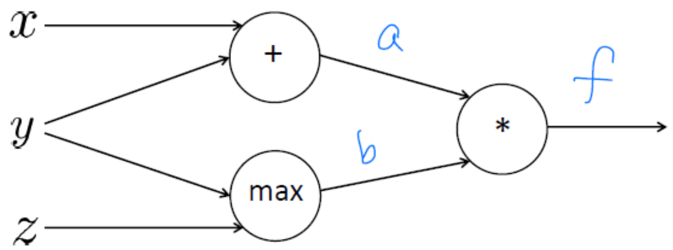
오차역전파 예제

An Example

$$f(x,y,z) = \overbrace{(x+y)}^{a}\,\overbrace{\max(y,z)}^{b} = ab$$
$$x = 1, y = 2, z = 0$$

Forward prop steps

$$a = x + y$$
$$b = \max(y, z)$$
$$f = ab$$



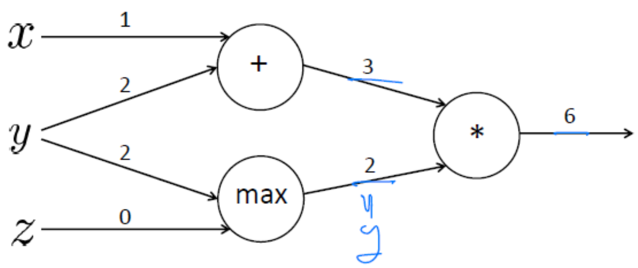
58

An Example

$$f(x,y,z) = (x+y)\max(y,z)$$
$$x = 1, y = 2, z = 0$$

Forward prop steps

$$a = x + y$$
$$b = \max(y, z)$$
$$f = ab$$



59

An Example

$$f(x,y,z) = (x+y)\max(y,z)$$
$$x=1, y=2, z=0$$

Forward prop steps

$a = x + y$

$b = \max(y, z)$

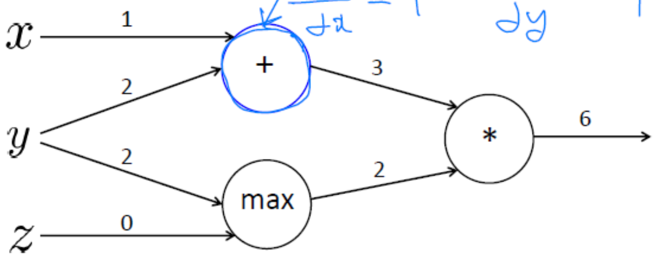
$f = ab$

Local gradients

$\frac{\partial a}{\partial x} = 1 \quad \frac{\partial a}{\partial y} = 1$

$a = x + y$

$\frac{\partial a}{\partial x} = 1 \quad \frac{\partial a}{\partial y} = 1$



60

An Example

$$f(x,y,z) = (x+y)\max(y,z)$$
$$x=1, y=2, z=0$$

Forward prop steps

$a = x + y$

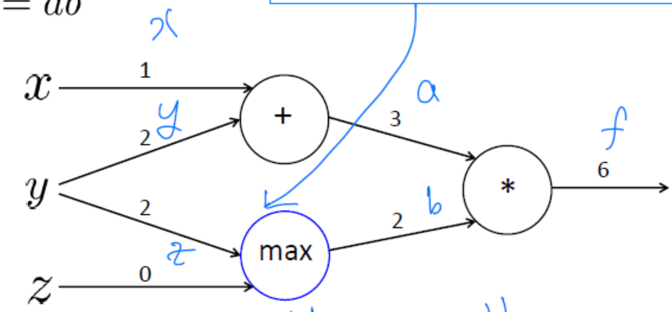
$b = \max(y, z)$

$f = ab$

Local gradients

$\frac{\partial a}{\partial x} = 1 \quad \frac{\partial a}{\partial y} = 1$

$\boxed{\frac{\partial b}{\partial y} = \mathbf{1}(y > z) = 1 \quad \frac{\partial b}{\partial z} = \mathbf{1}(z > y) = 0}$



$\frac{\partial b}{\partial y} = 1 \quad \frac{\partial b}{\partial z} = 0 \quad (y > z)$

61

An Example

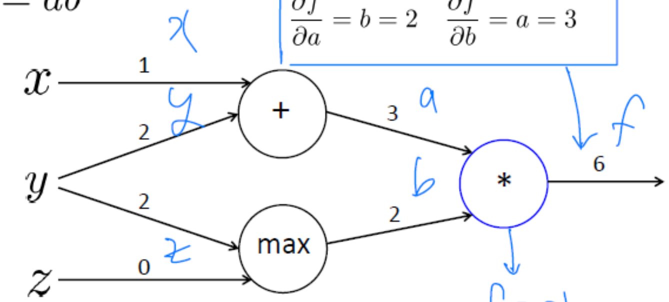
$$\boxed{\begin{aligned} &f(x,y,z) = (x+y)\max(y,z) \\ &x = 1, y = 2, z = 0 \end{aligned}}$$

Forward prop steps

$$a = x + y$$
$$b = \max(y,z)$$
$$f = ab$$

Local gradients

$$\frac{\partial a}{\partial x} = 1 \quad \frac{\partial a}{\partial y} = 1$$
$$\frac{\partial b}{\partial y} = \mathbf{1}(y > z) = 1 \quad \frac{\partial b}{\partial z} = \mathbf{1}(z > y) = 0$$
$$\boxed{\frac{\partial f}{\partial a} = b = 2 \quad \frac{\partial f}{\partial b} = a = 3}$$



$f = ab$

$$\frac{\partial f}{\partial a} = b = 2 \quad \frac{\partial f}{\partial b} = a = 3$$

62

예제 계산

An Example

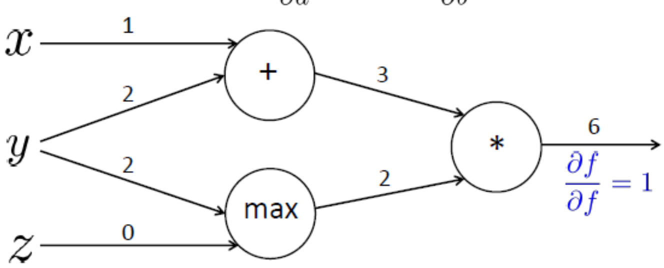
$$\boxed{\begin{aligned} &f(x,y,z) = (x+y)\max(y,z) \\ &x = 1, y = 2, z = 0 \end{aligned}}$$

Forward prop steps

$$a = x + y$$
$$b = \max(y,z)$$
$$f = ab$$

Local gradients

$$\frac{\partial a}{\partial x} = 1 \quad \frac{\partial a}{\partial y} = 1$$
$$\frac{\partial b}{\partial y} = \mathbf{1}(y > z) = 1 \quad \frac{\partial b}{\partial z} = \mathbf{1}(z > y) = 0$$
$$\frac{\partial f}{\partial a} = b = 2 \quad \frac{\partial f}{\partial b} = a = 3$$



63

An Example

$$\boxed{\begin{array}{l} f(x,y,z) = (x+y)\max(y,z) \\ x=1, y=2, z=0 \end{array}}$$

Forward prop steps

$$a = x + y$$

$$b = \max(y,z)$$

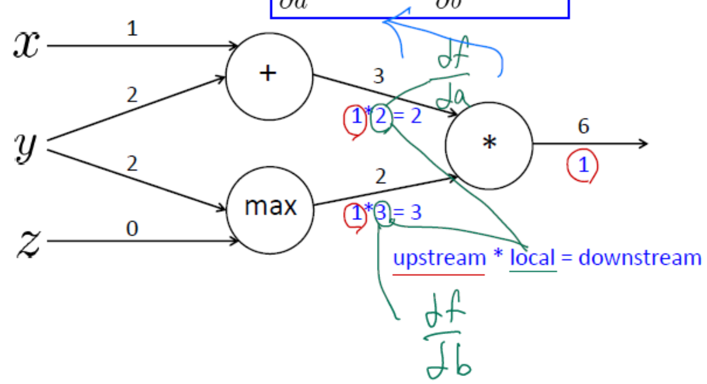
$$f = ab$$

Local gradients

$$\frac{\partial a}{\partial x} = 1 \quad \frac{\partial a}{\partial y} = 1$$

$$\frac{\partial b}{\partial y} = \mathbf{1}(y > z) = 1 \quad \frac{\partial b}{\partial z} = \mathbf{1}(z > y) = 0$$

$$\boxed{\frac{\partial f}{\partial a} = b = 2 \quad \frac{\partial f}{\partial b} = a = 3}$$



64

An Example

$$\boxed{\begin{array}{l} f(x,y,z) = (x+y)\max(y,z) \\ x=1, y=2, z=0 \end{array}}$$

Forward prop steps

$$a = x + y$$

$$b = \max(y,z)$$

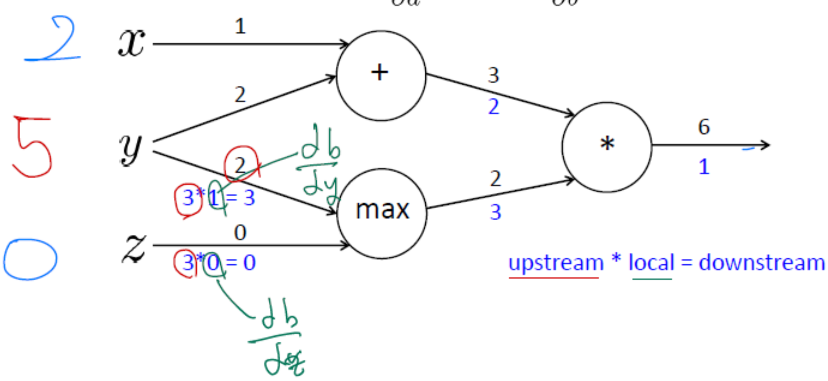
$$f = ab$$

Local gradients

$$\frac{\partial a}{\partial x} = 1 \quad \frac{\partial a}{\partial y} = 1$$

$$\boxed{\frac{\partial b}{\partial y} = \mathbf{1}(y > z) = 1 \quad \frac{\partial b}{\partial z} = \mathbf{1}(z > y) = 0}$$

$$\frac{\partial f}{\partial a} = b = 2 \quad \frac{\partial f}{\partial b} = a = 3$$



65

An Example

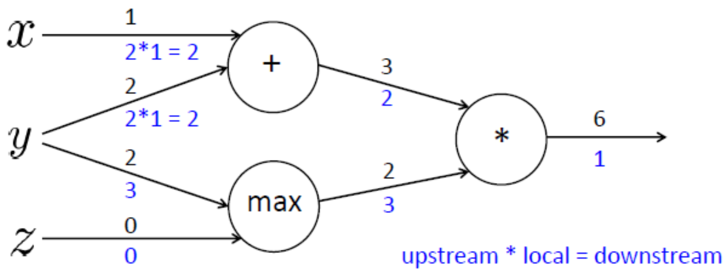
$$f(x,y,z) = (x+y)\max(y,z)$$
$$x = 1, y = 2, z = 0$$

Forward prop steps

$$a = x + y$$
$$b = \max(y,z)$$
$$f = ab$$

Local gradients

$$\boxed{\frac{\partial a}{\partial x} = 1 \quad \frac{\partial a}{\partial y} = 1}$$
$$\frac{\partial b}{\partial y} = \mathbf{1}(y > z) = 1 \quad \frac{\partial b}{\partial z} = \mathbf{1}(z > y) = 0$$
$$\frac{\partial f}{\partial a} = b = 2 \quad \frac{\partial f}{\partial b} = a = 3$$



upstream * local = downstream

66

An Example

$$f(x,y,z) = (x+y)\max(y,z)$$
$$x = 1, y = 2, z = 0$$

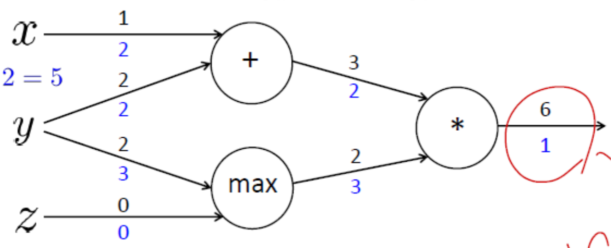
Forward prop steps

$$a = x + y$$
$$b = \max(y,z)$$
$$f = ab$$

Local gradients

$$\frac{\partial a}{\partial x} = 1 \quad \frac{\partial a}{\partial y} = 1$$
$$\frac{\partial b}{\partial y} = \mathbf{1}(y > z) = 1 \quad \frac{\partial b}{\partial z} = \mathbf{1}(z > y) = 0$$
$$\frac{\partial f}{\partial a} = b = 2 \quad \frac{\partial f}{\partial b} = a = 3$$

$$\frac{\partial f}{\partial x} = 2$$
$$\frac{\partial f}{\partial y} = 3 + 2 = 5$$
$$\frac{\partial f}{\partial z} = 0$$



67

x 1→1.1로 바꾸면
f 0.2만큼 바뀐다!

x 변화량 × $\frac{\partial f}{\partial x}$ 만큼 f가 바뀐다
0.1
변화량

이번 주에는 오차역전파를 손으로 직접 계산해보는 시간을 가졌습니다