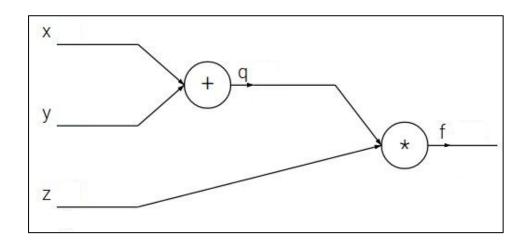
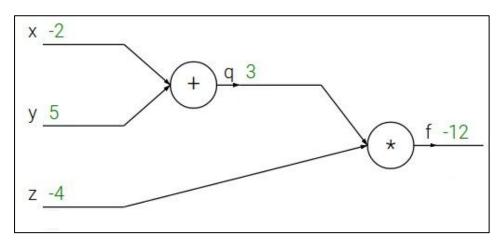
$$f(x,y,z)=(x+y)z$$

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



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

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

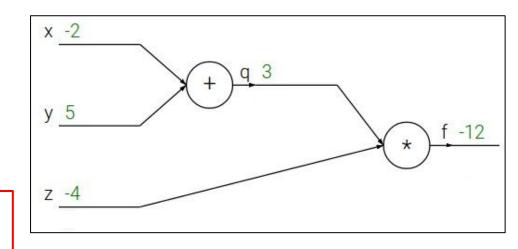


$$f(x,y,z)=(x+y)z$$

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

$$q=x+y \qquad rac{\partial q}{\partial x}=1, rac{\partial q}{\partial y}=1$$

$$f=qz$$
 $rac{\partial f}{\partial q}=z, rac{\partial f}{\partial z}=q$

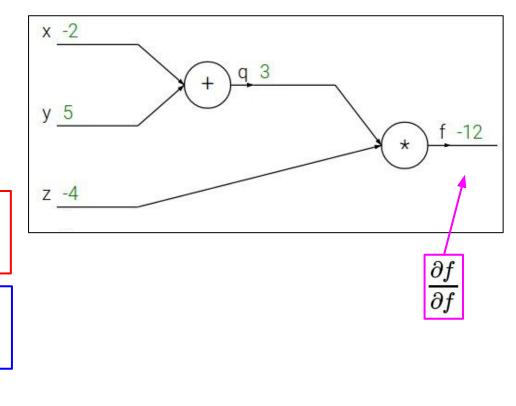


$$f(x,y,z)=(x+y)z$$

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

$$q=x+y \qquad rac{\partial q}{\partial x}=1, rac{\partial q}{\partial y}=1$$

$$f=qz$$
 $rac{\partial f}{\partial q}=z, rac{\partial f}{\partial z}=q$

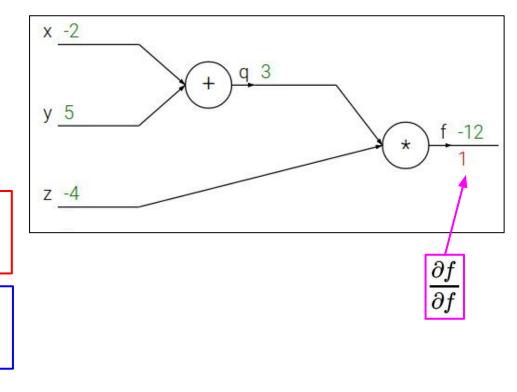


$$f(x,y,z)=(x+y)z$$

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

$$q=x+y \qquad rac{\partial q}{\partial x}=1, rac{\partial q}{\partial y}=1$$

$$f=qz$$
 $rac{\partial f}{\partial q}=z, rac{\partial f}{\partial z}=q$

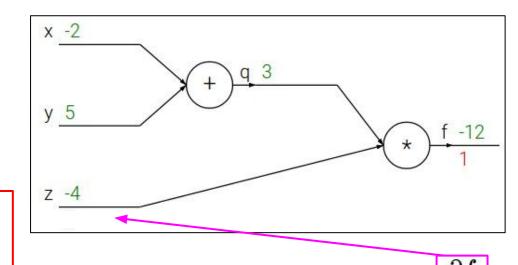


$$f(x,y,z)=(x+y)z$$

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

$$q=x+y \qquad rac{\partial q}{\partial x}=1, rac{\partial q}{\partial y}=1$$

$$f=qz$$
 $rac{\partial f}{\partial q}=z, rac{\partial f}{\partial z}=q$



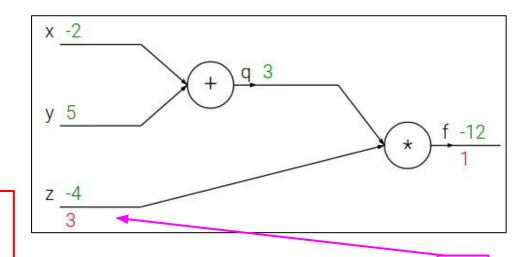
$$f(x,y,z)=(x+y)z$$

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

$$q=x+y \qquad rac{\partial q}{\partial x}=1, rac{\partial q}{\partial y}=1$$

$$f=qz$$
 $rac{\partial f}{\partial q}=z, rac{\partial f}{\partial z}=q$

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



Fei-Fei Li & Justin Johnson & Serena Yeung

Lecture 4 - 18

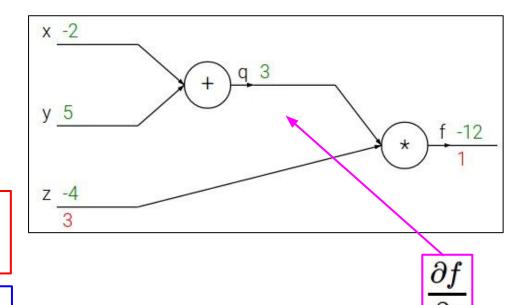
April 13, 2017

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

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

$$q=x+y \qquad rac{\partial q}{\partial x}=1, rac{\partial q}{\partial y}=1$$

$$f=qz$$
 $rac{\partial f}{\partial q}=z, rac{\partial f}{\partial z}=q$

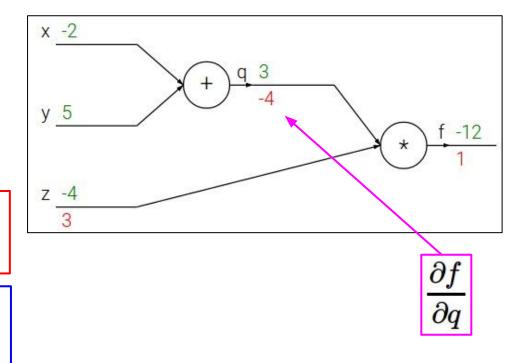


$$f(x,y,z)=(x+y)z$$

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

$$q=x+y \hspace{0.5cm} rac{\partial q}{\partial x}=1, rac{\partial q}{\partial y}=1$$

$$f=qz$$
 $rac{\partial f}{\partial q}=z, rac{\partial f}{\partial z}=q$

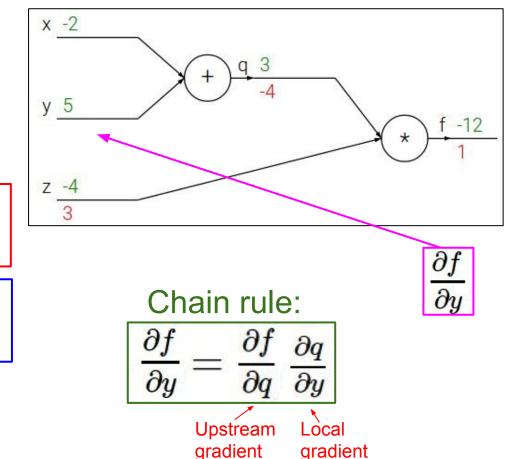


$$f(x,y,z)=(x+y)z$$

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

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

$$f=qz$$
 $rac{\partial f}{\partial q}=z, rac{\partial f}{\partial z}=q$

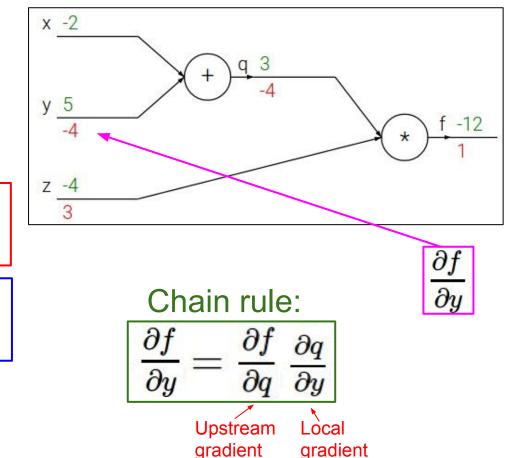


$$f(x,y,z)=(x+y)z$$

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

$$q=x+y \qquad rac{\partial q}{\partial x}=1, rac{\partial q}{\partial y}=1$$

$$f=qz$$
 $rac{\partial f}{\partial q}=z, rac{\partial f}{\partial z}=q$

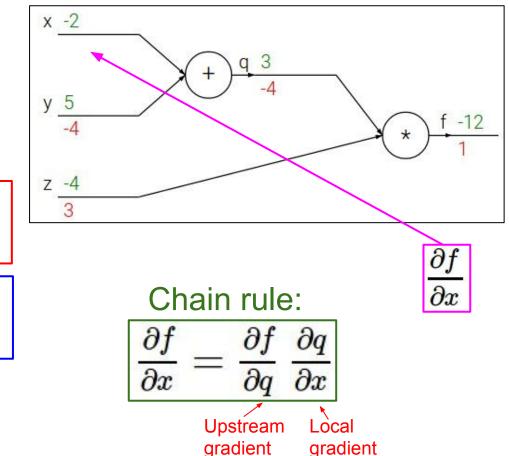


$$f(x,y,z)=(x+y)z$$

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

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

$$f=qz$$
 $rac{\partial f}{\partial q}=z, rac{\partial f}{\partial z}=q$

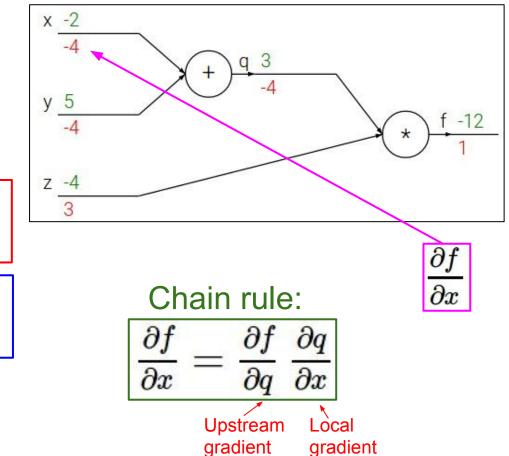


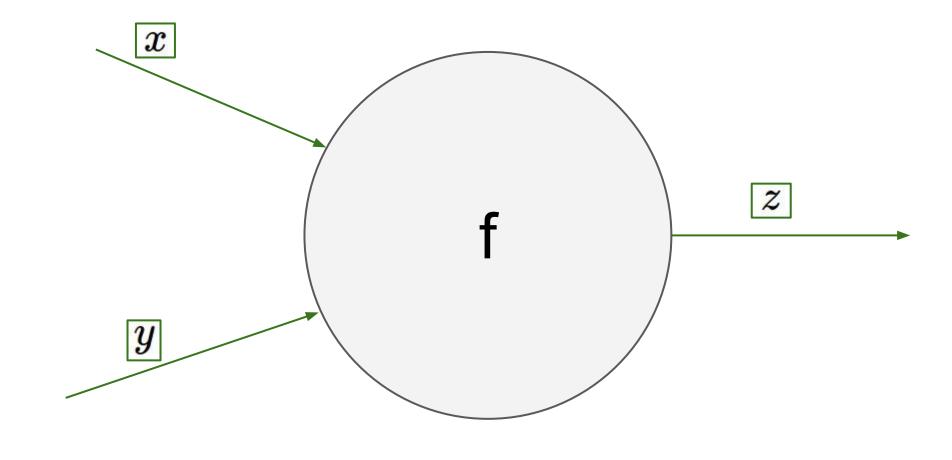
$$f(x,y,z)=(x+y)z$$

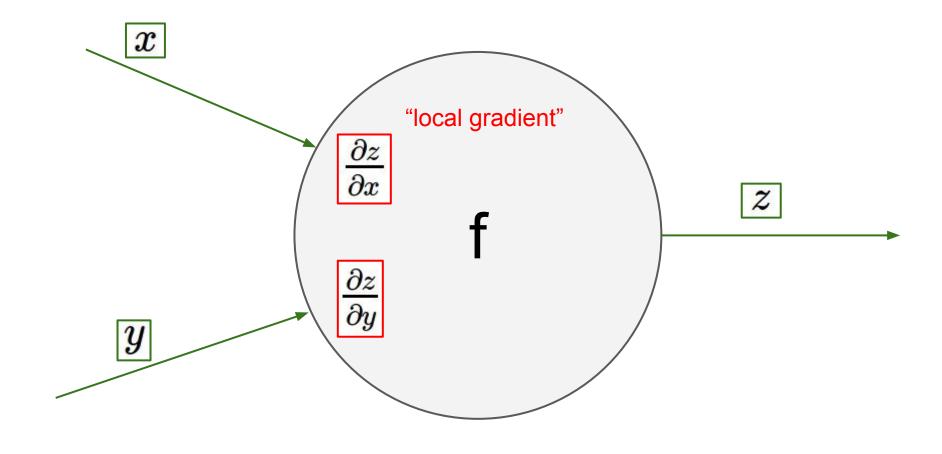
e.g.
$$x = -2$$
, $y = 5$, $z = -4$

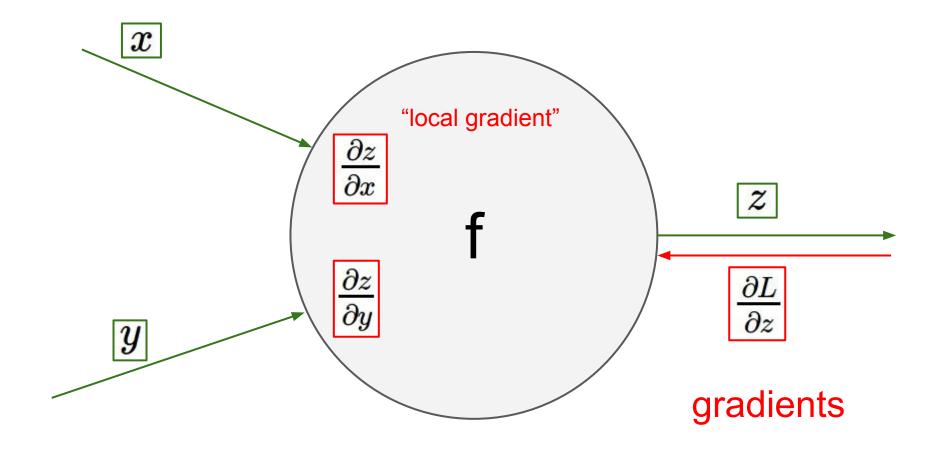
$$q=x+y$$
 $\frac{\partial q}{\partial x}=1, \frac{\partial q}{\partial y}=1$

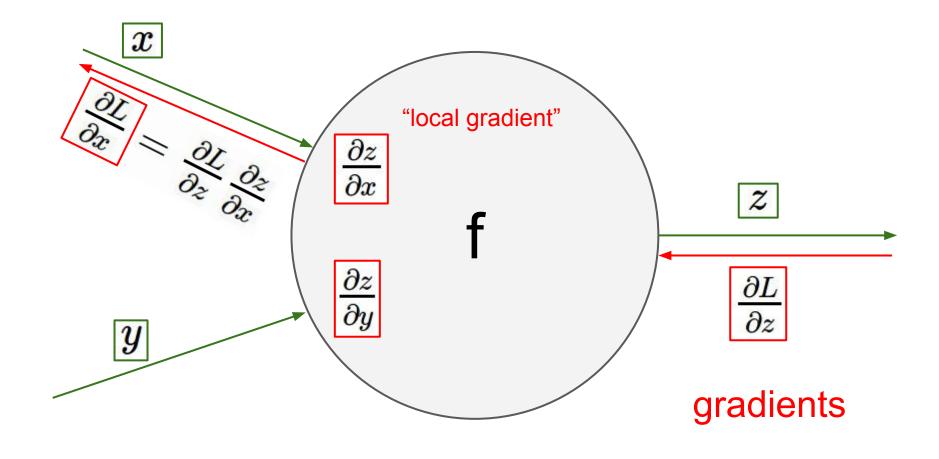
$$f=qz$$
 $rac{\partial f}{\partial q}=z, rac{\partial f}{\partial z}=q$

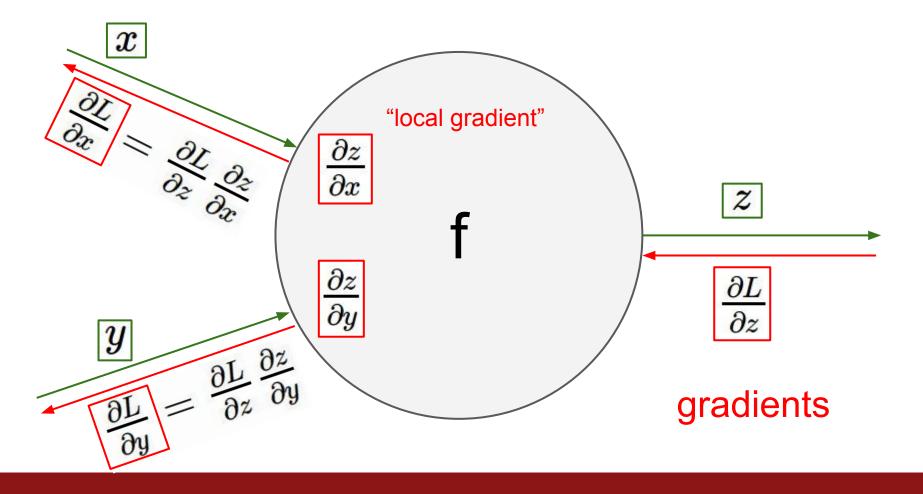


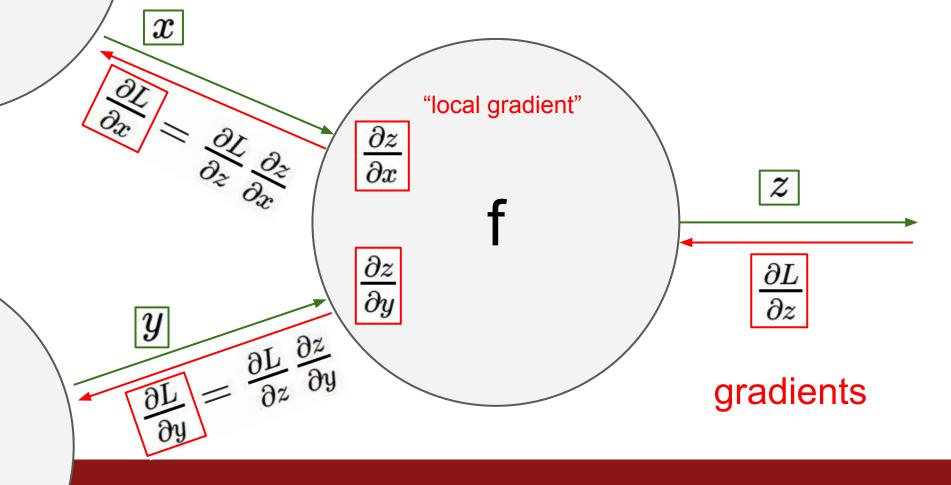




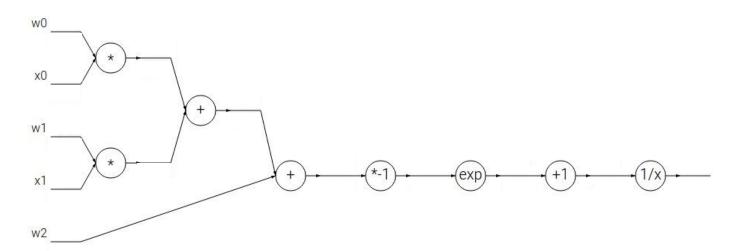




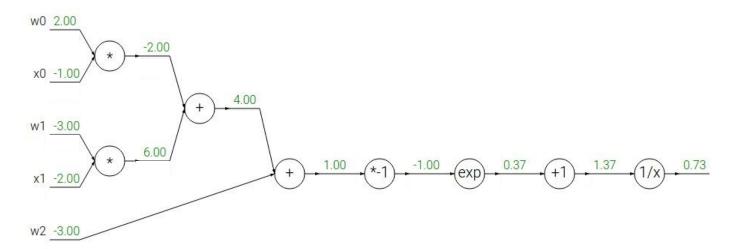




Another example: $f(w,x) = \frac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$



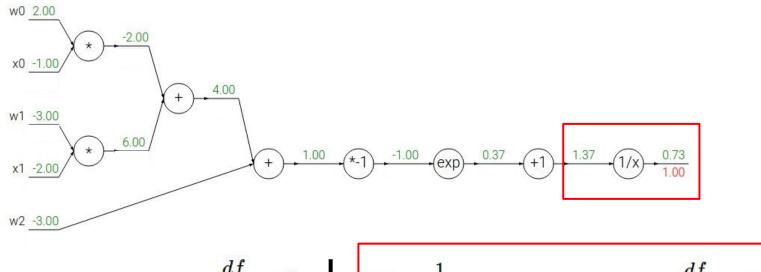
Another example:
$$f(w,x) = \frac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$$



Another example:
$$f(w,x) = \frac{1}{1 + e^{-(w_0)}}$$

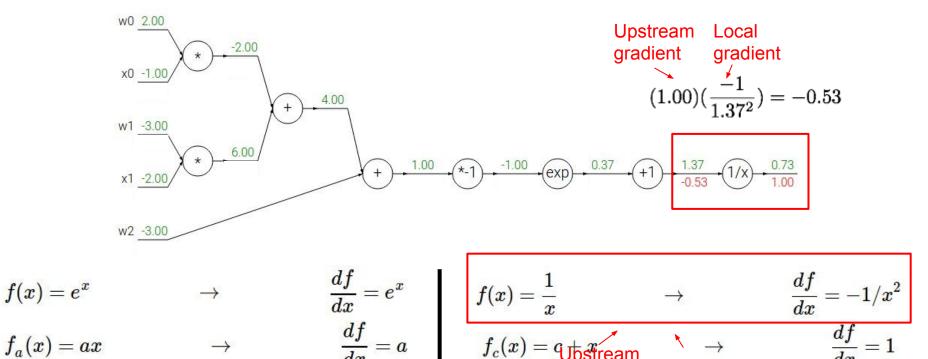
$$egin{aligned} f(x) = e^x &
ightarrow & rac{df}{dx} = e^x & f(x) = rac{1}{x} &
ightarrow & rac{df}{dx} = -1/x^2 \ f_a(x) = ax &
ightarrow & rac{df}{dx} = a & f_c(x) = c + x &
ightarrow & rac{df}{dx} = 1 \end{aligned}$$

Another example:
$$f(w,x) = \frac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$$



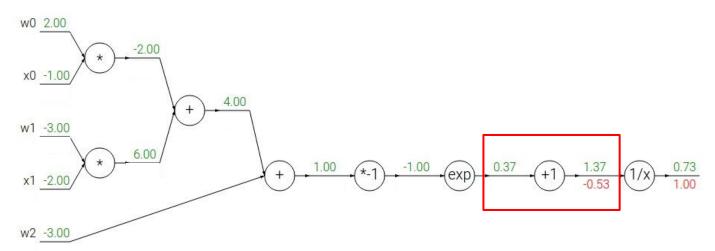
$$f(x) = e^x \hspace{1cm} o \hspace{1cm} rac{df}{dx} = e^x \hspace{1cm} f(x) = rac{1}{x} \hspace{1cm} o \hspace{1cm} rac{df}{dx} = -1/x^2 \ f_c(x) = ax \hspace{1cm} o \hspace{1cm} rac{df}{dx} = 1$$

$$f(w,x)=rac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$

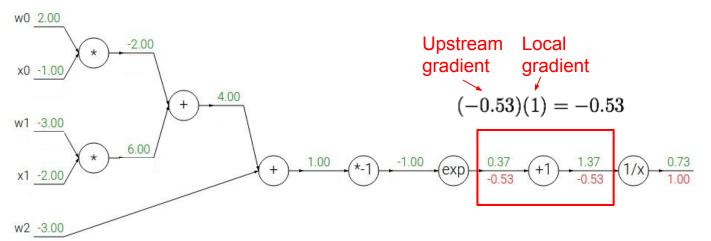


gradient

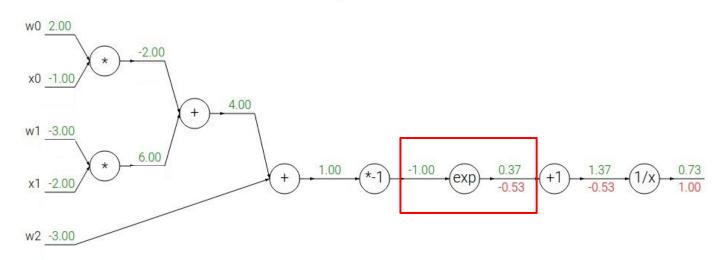
Another example:
$$f(w,x) = \frac{1}{1 + e^{-(w_0x_0 + w_1x_1)}}$$



Another example:
$$f(w,x) = \frac{1}{1+e^{-(u)}}$$



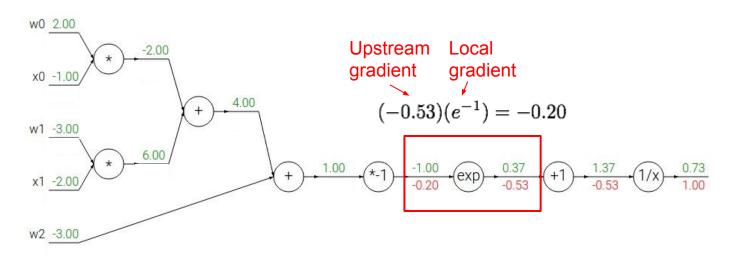
$$f(w,x) = rac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_1 x_2 + w_2 x_1 + w_2 x_2 + w_2 x_2$$



$$f(x)=e^x \qquad o \qquad rac{df}{dx}=e^x \ f(x)=rac{1}{x} \ f_c(x)=c+x \$$

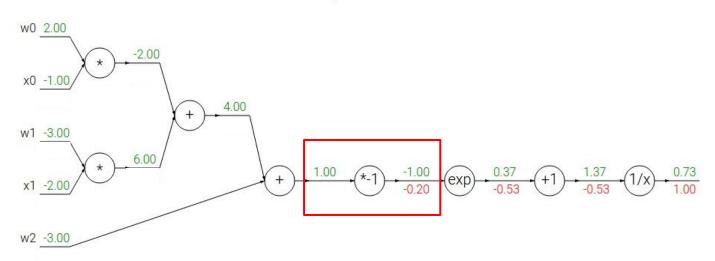
$$egin{array}{ll}
ightarrow & rac{df}{dx} = -1/x \
ightarrow & rac{df}{dx} = 1/x \
ightarrow & 1/x \
ightarrow & 1/x \
ightarrow & 1/x \
ightarrow & 1/x \
ho = 1/x \
h$$

$$f(w,x)=rac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$

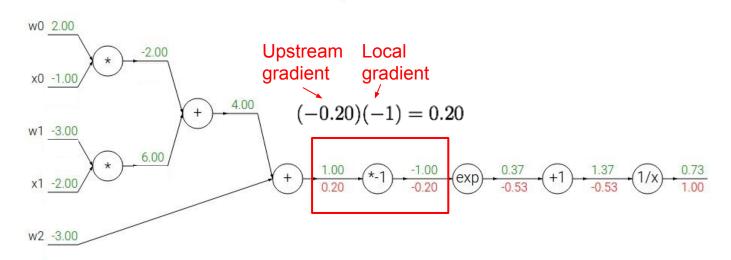


$$egin{aligned} f(x) = e^x &
ightarrow & rac{df}{dx} = e^x \ \hline f_a(x) = ax &
ightarrow & rac{df}{dx} = a \ \hline \end{pmatrix} egin{aligned} f(x) = rac{1}{x} &
ightarrow & rac{df}{dx} = -1/x^2 \ \hline f_c(x) = c + x &
ightarrow & rac{df}{dx} = 1 \ \hline \end{pmatrix}$$

$$f(w,x) = rac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_1 x_2 + w_2 x_1 + w_2 x_2 + w_2 x_2$$



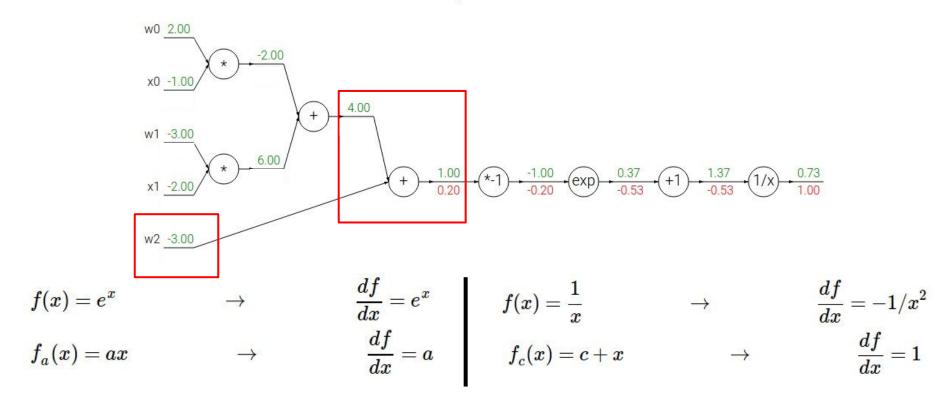
$$f(w,x) = rac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2 x_1 + w_2 x_2 + w_1 x_2 + w_2 x_2$$



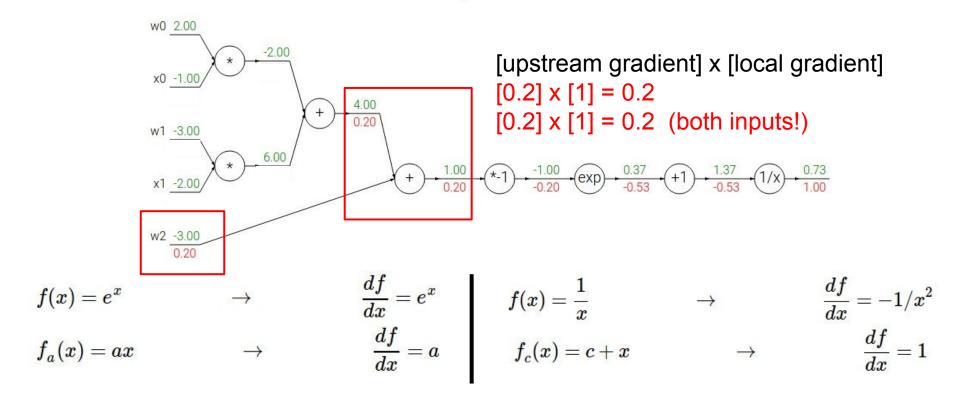
$$f(x) = e^x \qquad o \qquad rac{df}{dx} = e^x \ f_a(x) = ax \qquad o \qquad rac{df}{dx} = a$$

$$f(x) = rac{1}{x}
ightarrow rac{df}{dx} = -1/x
ightarrow f_c(x) = c + x
ightarrow rac{df}{dx} = 1$$

$$f(w,x) = rac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_1 x_2 + w_2 x_1 + w_2 x_2 + w_2 x_2$$



$$f(w,x)=rac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$

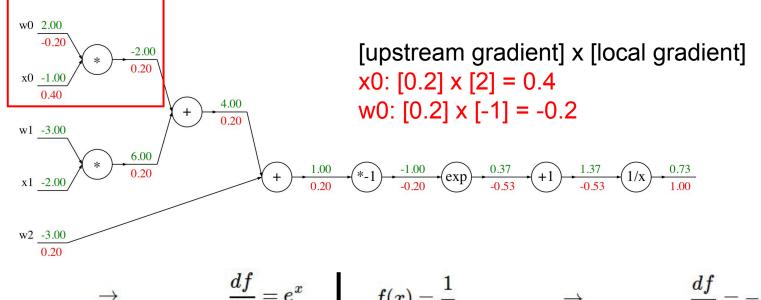


Another example:
$$f(w,x) = \frac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$$

$$f(x) = e^{x} \qquad \rightarrow \qquad \frac{df}{dx} = e^{x} \qquad f(x) = ax \qquad \rightarrow \qquad \frac{df}{dx} = a \qquad f(x) = c + x \qquad \rightarrow \qquad \frac{df}{dx} = 1$$

Another example:

$$f(w,x)=rac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$



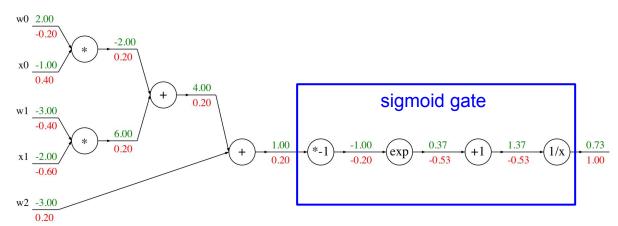
$$egin{aligned} rac{df}{dx} = e^x & f(x) = rac{1}{x} &
ightarrow & rac{df}{dx} = -1/x \ rac{df}{dx} = a & f_c(x) = c + x &
ightarrow & rac{df}{dx} = 1/x \ rac{df}{dx} =$$

$$f(w,x)=rac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$

Computational graph representation may not be unique. Choose one where local gradients at each node can be easily expressed!

$$\sigma(x) = rac{1}{1+e^{-x}}$$
 sigmoid function

$$rac{d\sigma(x)}{dx} = rac{e^{-x}}{\left(1 + e^{-x}
ight)^2} = \left(rac{1 + e^{-x} - 1}{1 + e^{-x}}
ight) \left(rac{1}{1 + e^{-x}}
ight) = \left(1 - \sigma(x)
ight)\sigma(x)$$

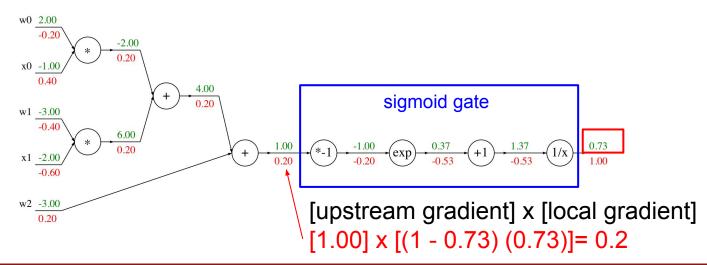


$$f(w,x)=rac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$

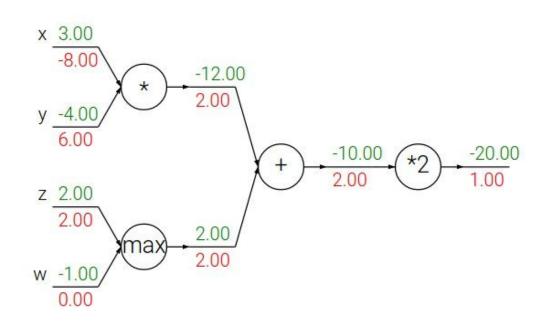
Computational graph representation may not be unique. Choose one where local gradients at each node can be easily expressed!

$$\sigma(x) = rac{1}{1 + e^{-x}}$$
 sigmoid function

$$rac{d\sigma(x)}{dx} = rac{e^{-x}}{(1+e^{-x})^2} = \left(rac{1+e^{-x}-1}{1+e^{-x}}
ight) \left(rac{1}{1+e^{-x}}
ight) = \left(1-\sigma(x)
ight)\sigma(x)$$

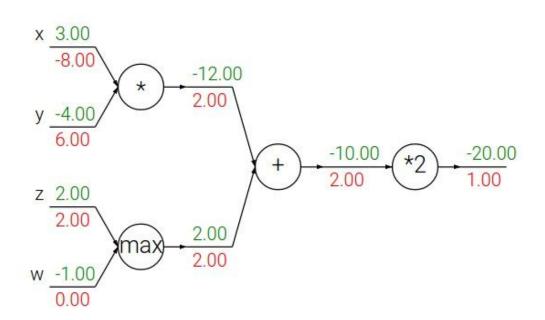


add gate: gradient distributor



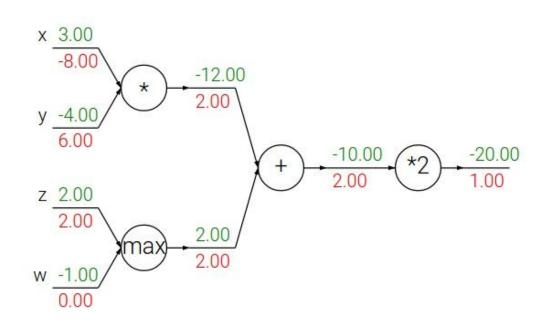
add gate: gradient distributor

Q: What is a max gate?



add gate: gradient distributor

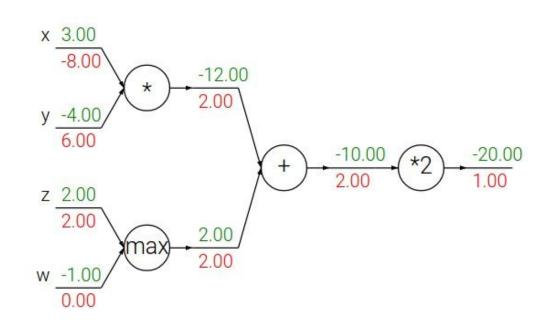
max gate: gradient router



add gate: gradient distributor

max gate: gradient router

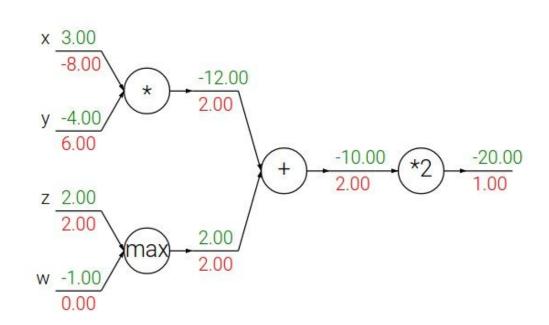
Q: What is a **mul** gate?



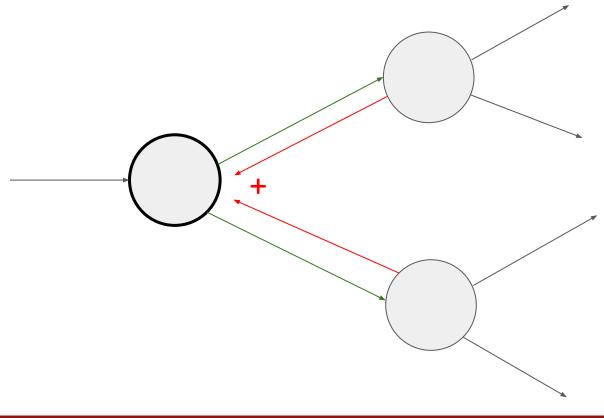
add gate: gradient distributor

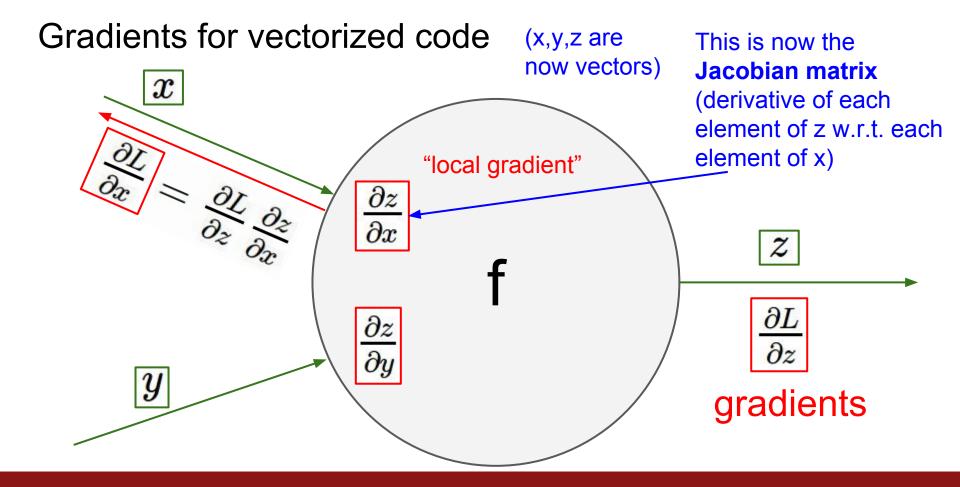
max gate: gradient router

mul gate: gradient switcher

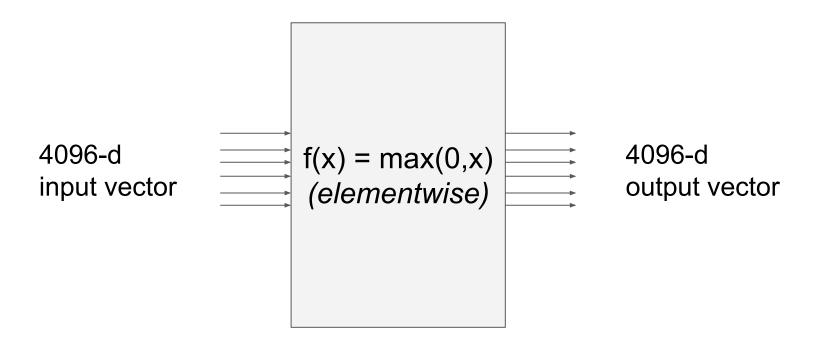


Gradients add at branches





Vectorized operations



Vectorized operations

$$rac{\partial L}{\partial x} = egin{bmatrix} rac{\partial f}{\partial x} rac{\partial L}{\partial f} \ & ext{Jacobian matrix} \end{cases}$$

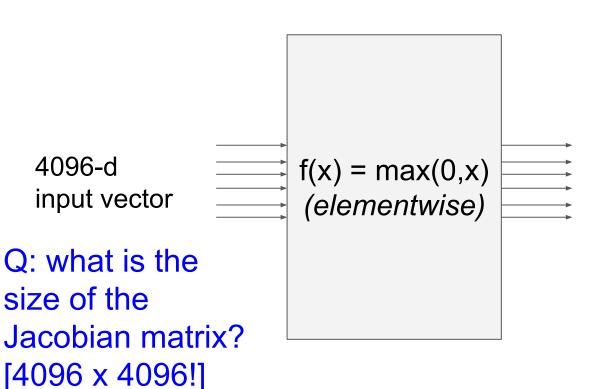
4096-d input vector

f(x) = max(0,x) (elementwise)

4096-d output vector

Q: what is the size of the Jacobian matrix?

Vectorized operations



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

Jacobian matrix

4096-d output vector