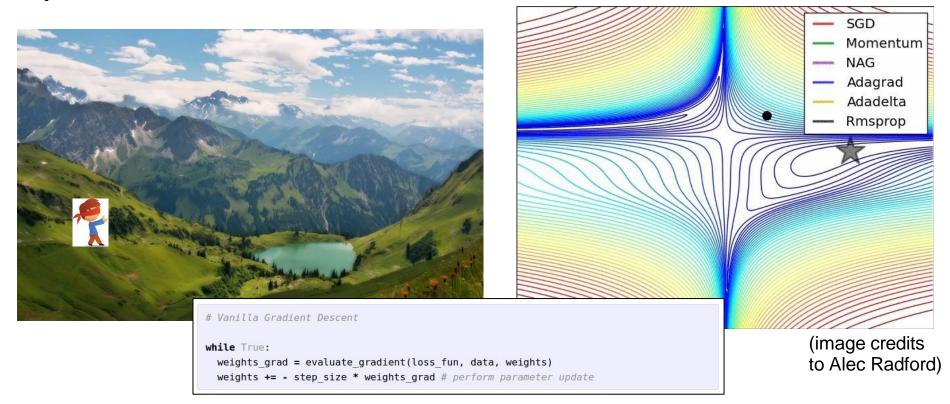
Lecture 6: Backpropagation Vector, Matrix and Tensor **Derivatives**

Where we are ...

$$s=f(x;W)=Wx$$
 scores function $L_i=\sum_{j
eq y_i}\max(0,s_j-s_{y_i}+1)$ SVM loss $L=rac{1}{N}\sum_{i=1}^NL_i+\sum_kW_k^2$ data loss + regularization

want $abla_W L$

Optimization



Gradient Descent

$$rac{df(x)}{dx} = \lim_{h o 0} rac{f(x+h) - f(x)}{h}$$

Numerical gradient: slow:(, approximate:(, easy to write:)
Analytic gradient: fast:), exact:), error-prone:(

In practice: Derive analytic gradient, check your implementation with numerical gradient

Overview of where we're going

- We want to evaluate the gradient of a Loss function L(x,W,...), with respect to the parameters (weights) of a neural network, at the "point" represented by the arguments to the function (x,W,...).
 - We are not interested in an algebraic expression for the gradient, but rather only in the evaluation of that gradient at the current value of the function arguments.

Consider the function

$$z(x,y) = x^2 + y^2,$$

and suppose we are interested in evaluating the gradient of this function at the point

$$(x,y) = (5,3).$$

Evaluate the gradient:

$$\frac{\partial z}{\partial x} = 2x.$$

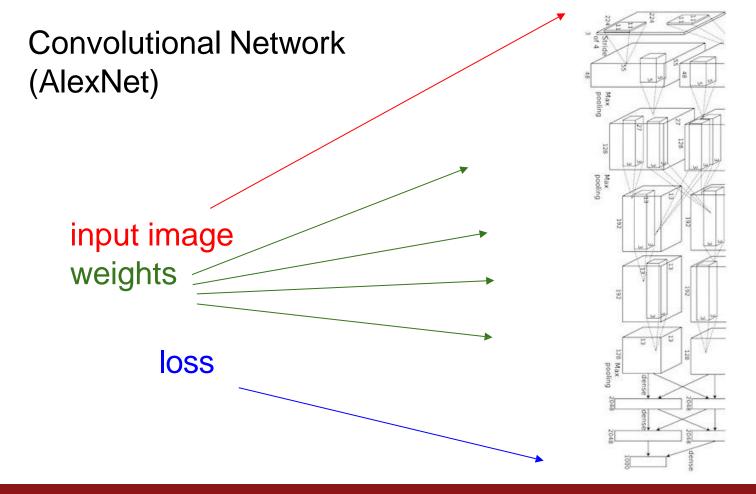
$$\frac{\partial z}{\partial y} = 2y.$$

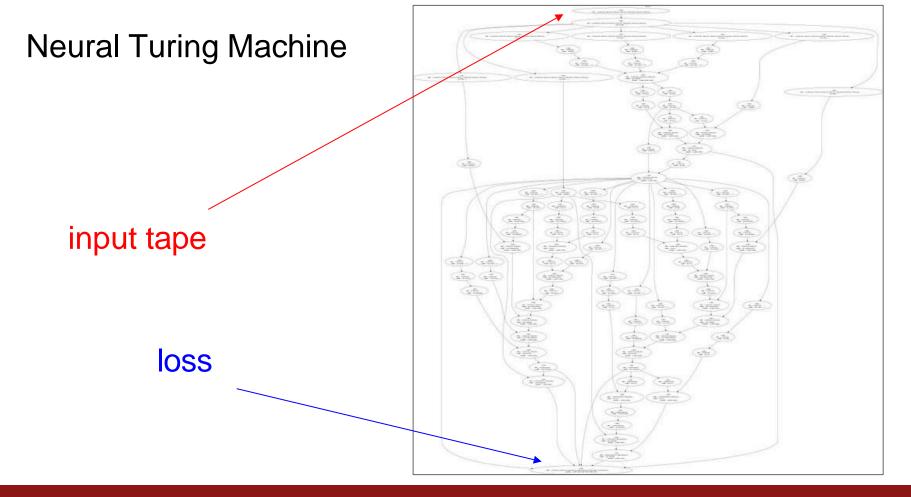
The algebraic expression of the gradient is just the collection of these partials into a "vector":

$$abla z = egin{bmatrix} 2x \\ 2y \end{bmatrix}$$
 . $lacksquare$ Don't care about this

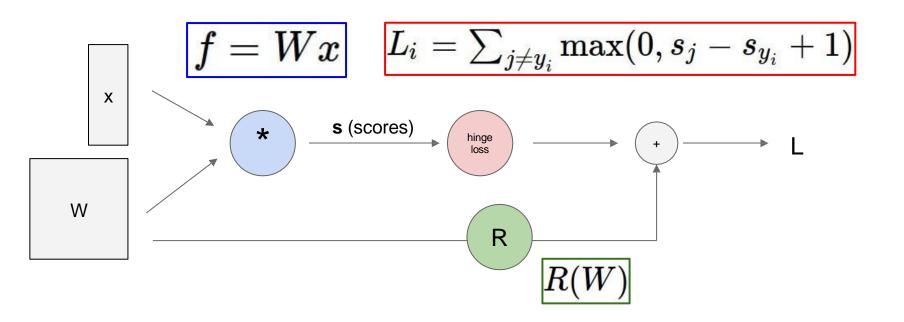
The evaluation of this gradient at the point (x, y) = (5, 3) is simply

$$\nabla z(5,3) = \begin{bmatrix} 2 \times 5 \\ 2 \times 3 \end{bmatrix} = \begin{bmatrix} 10 \\ 6 \end{bmatrix}$$
. Do care about this





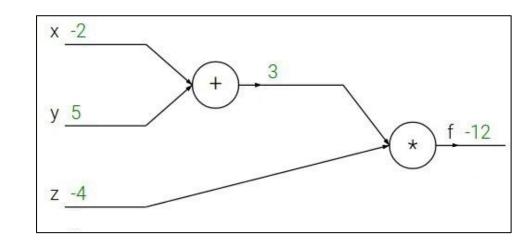
Computational Graph



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

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

Forward pass: evaluating each expression in the computational graph from the inputs to the final output (or outputs). The results of each forward step are shown in green.

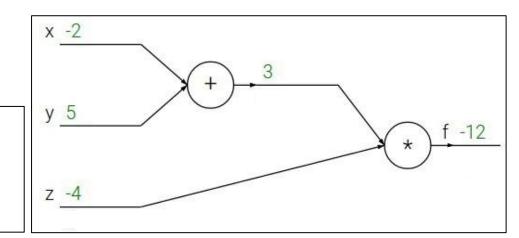


```
# set some inputs
x = -2; y = 5; z = -4
# perform the forward pass
q = x + y # q becomes 3
f = q * z # f becomes -12
# perform the backward pass (backpropagation) in reverse order:
# first backprop through f = q * z
dfdz = q \# df/dz = q, so gradient on z becomes 3
dfdg = z \# df/dg = z, so gradient on g becomes -4
# now backprop through q = x + y
dfdx = 1.0 * dfdq # dq/dx = 1. And the multiplication here is the chain rule!
dfdy = 1.0 * dfdq # dq/dy = 1
```

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

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

Backward pass: evaluating the partial derivative of each parameter or intermediate result in the computational graph from the outputs back to the inputs. The results of each backward step are shown in red.



Goal is to calculate

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

evaluated at the point

$$[x = -2, y = 5, z = -4].$$

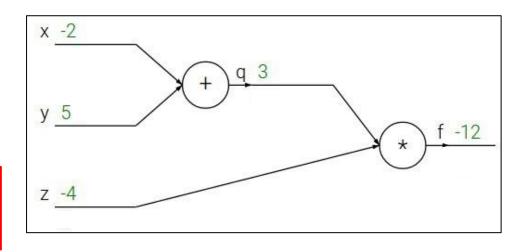
$$f(x,y,z)=(x+y)z$$

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

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

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

Want:



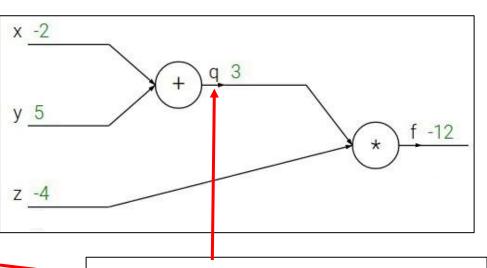
$$f(x,y,z) = (x+y)z$$

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

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

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

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



Important: name the intermediate quantities

Compute some **local partial derivatives.**These are derivatives of the outputs of a node with respect to the inputs....

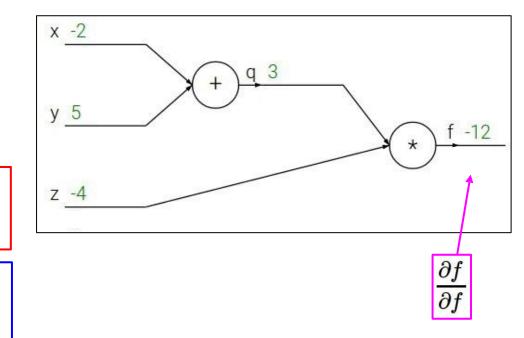
$$f(x,y,z)=(x+y)z$$

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:



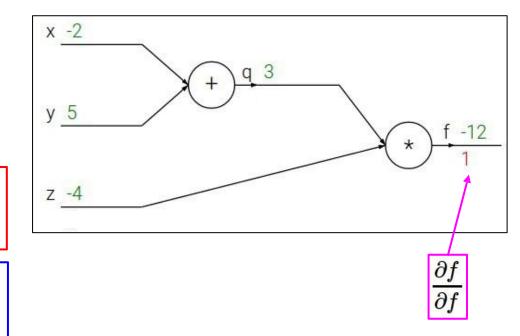
$$f(x,y,z) = (x+y)z$$

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



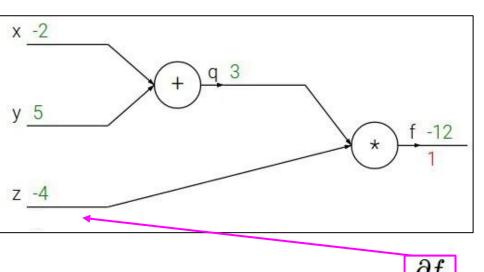
$$f(x,y,z) = (x+y)z$$

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



 $\frac{\partial f}{\partial z}$

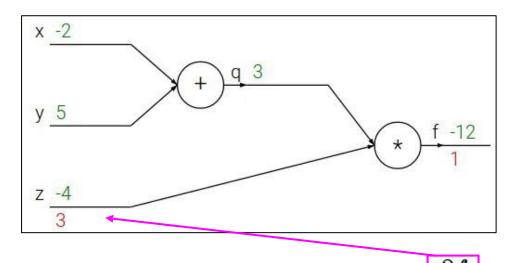
$$f(x,y,z)=(x+y)z$$

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



 $\frac{\partial f}{\partial z}$

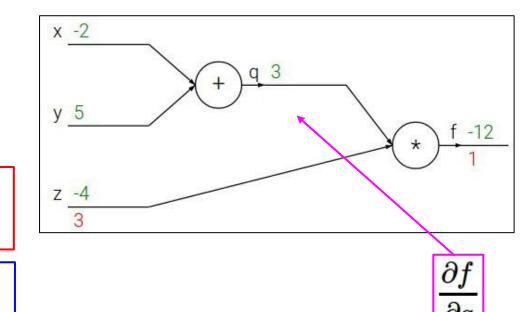
$$f(x,y,z)=(x+y)z$$

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



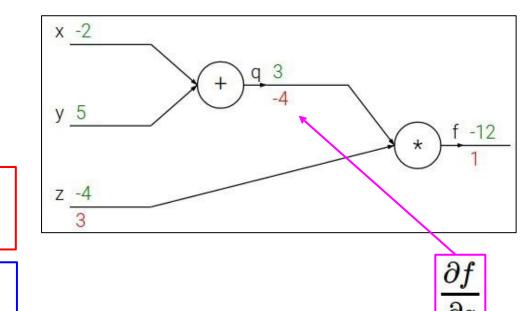
$$f(x,y,z) = (x+y)z$$

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:



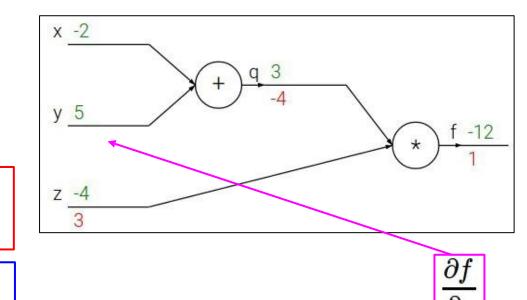
$$f(x,y,z)=(x+y)z$$

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



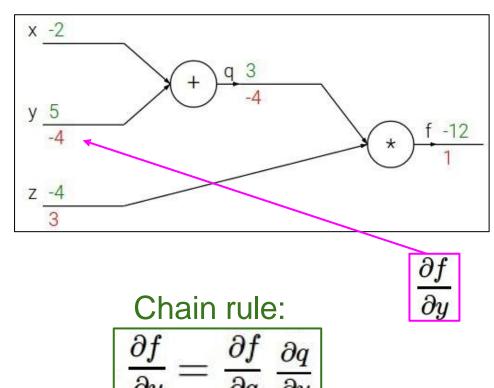
$$f(x,y,z) = (x+y)z$$

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

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

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

Want:



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

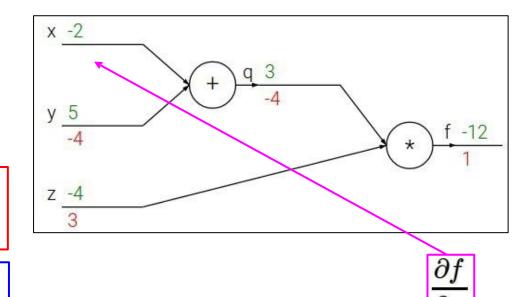
$$f(x,y,z) = (x+y)z$$

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



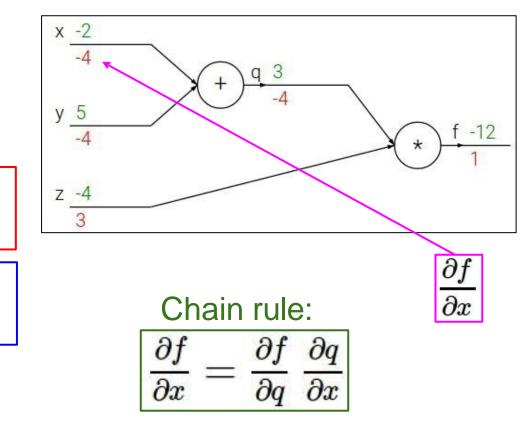
$$f(x,y,z) = (x+y)z$$

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

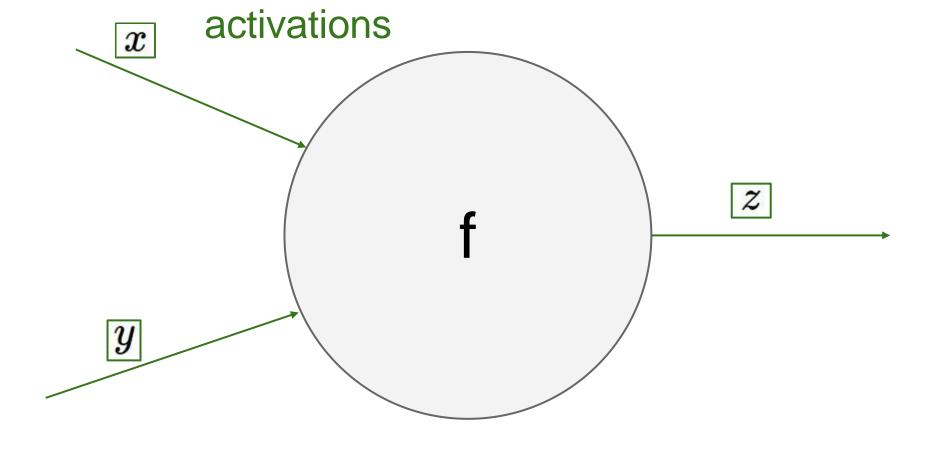
$$q=x+y \qquad rac{\partial q}{\partial x}=1, rac{\partial q}{\partial y}=1$$

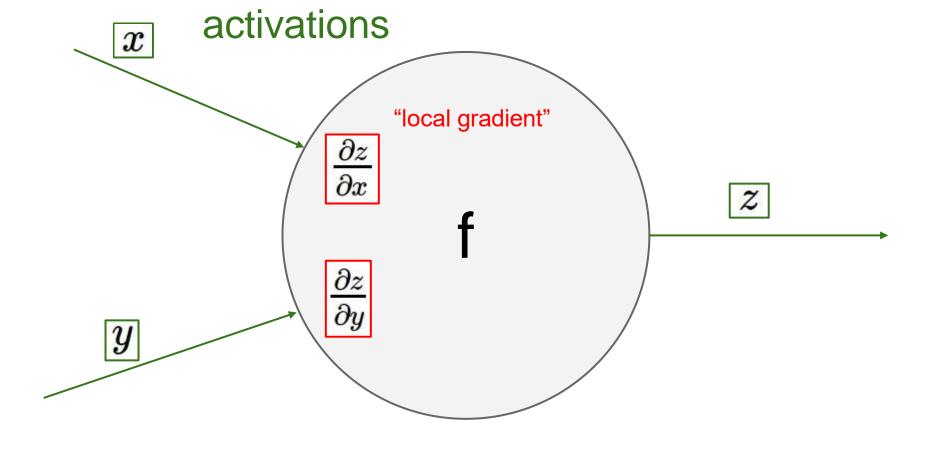
$$f=qz$$
 $rac{\partial f}{\partial q}=z, rac{\partial f}{\partial z}=q$

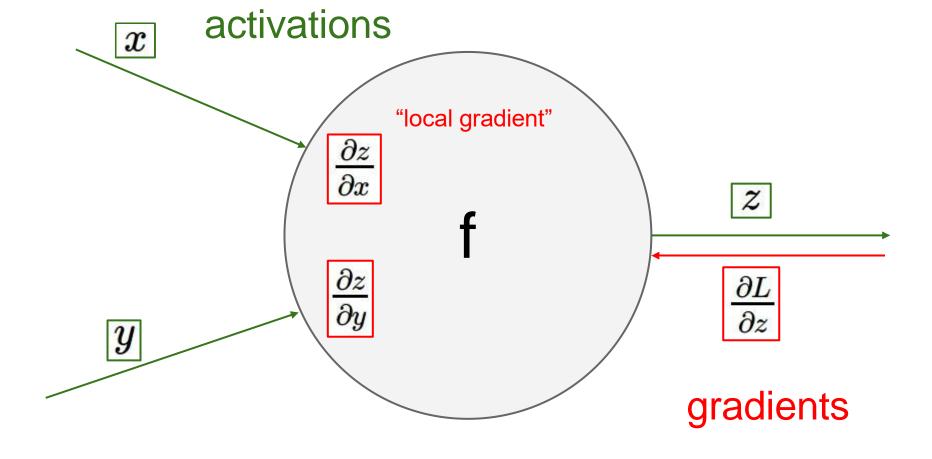
Want:

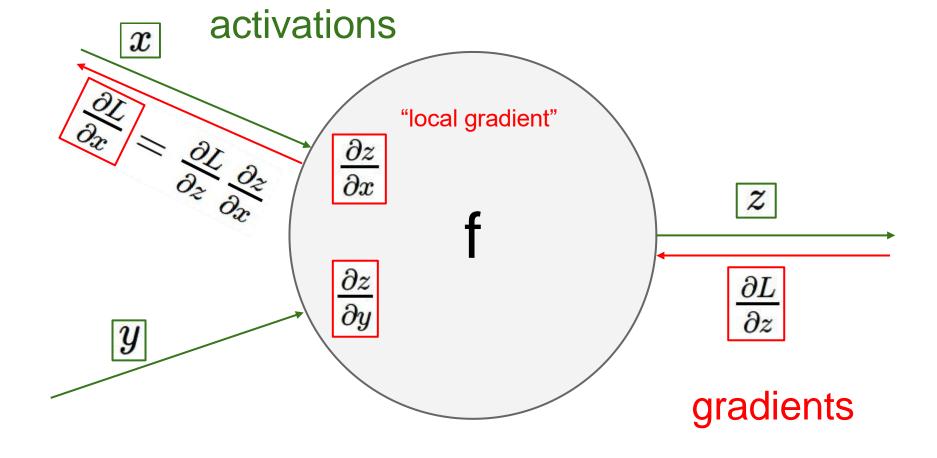


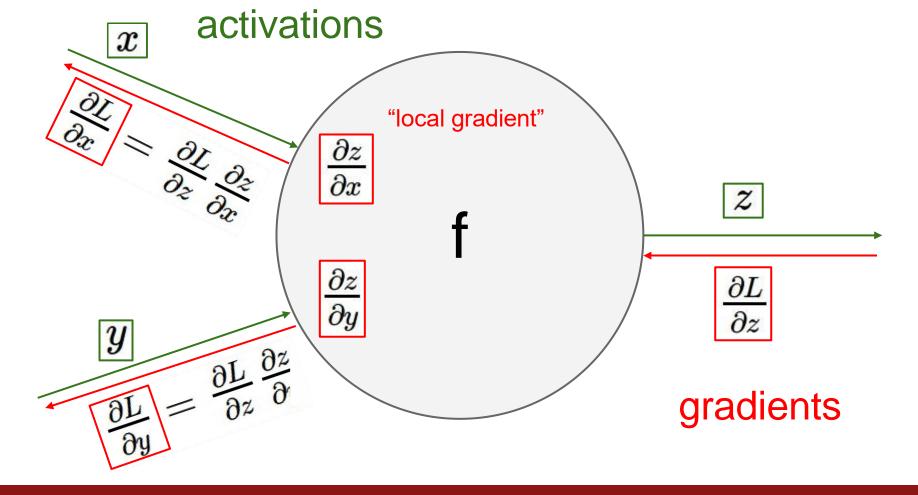
```
# set some inputs
x = -2; y = 5; z = -4
# perform the forward pass
q = x + y # q becomes 3
f = q * z # f becomes -12
# perform the backward pass (backpropagation) in reverse order:
# first backprop through f = q * z
dfdz = q \# df/dz = q, so gradient on z becomes 3
dfdg = z \# df/dg = z, so gradient on g becomes -4
# now backprop through q = x + y
dfdx = 1.0 * dfdq # dq/dx = 1. And the multiplication here is the chain rule!
dfdy = 1.0 * dfdq # dq/dy = 1
```

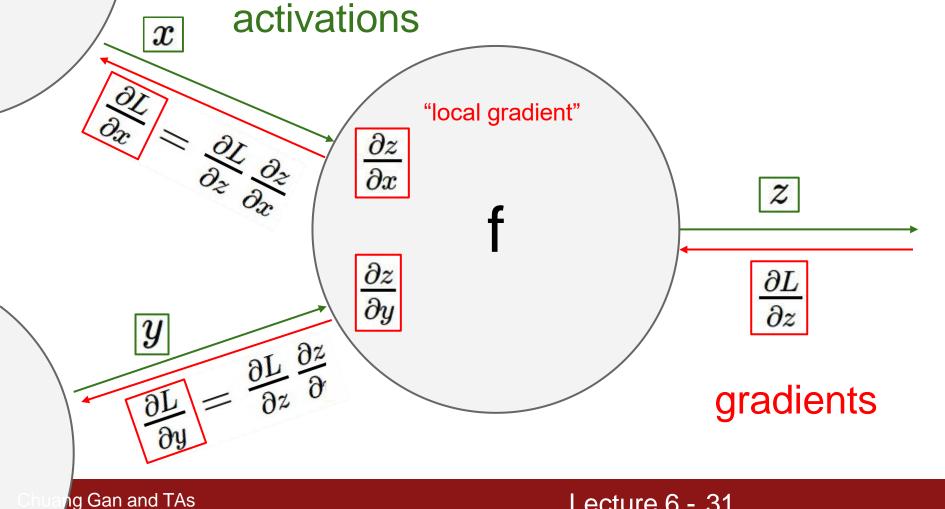










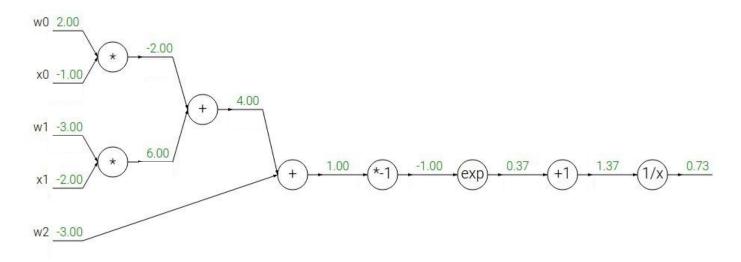


slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

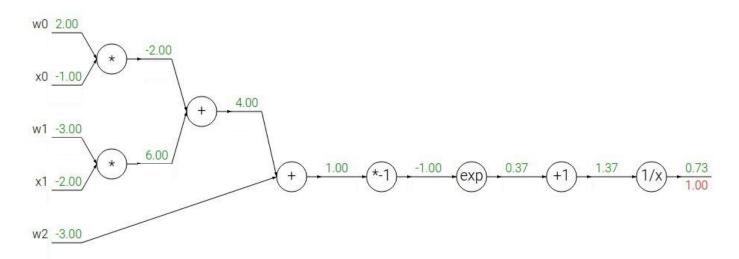
Lecture 6 - 31

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

"sigmoid function"



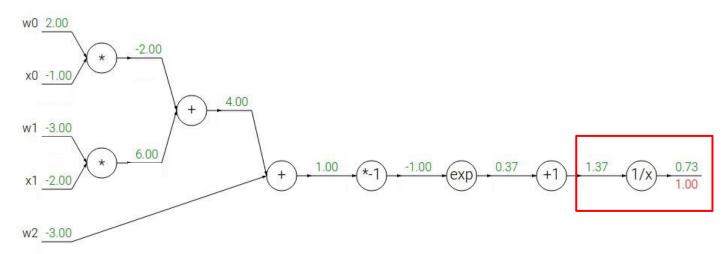
$$f(w,x) = rac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$$



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

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

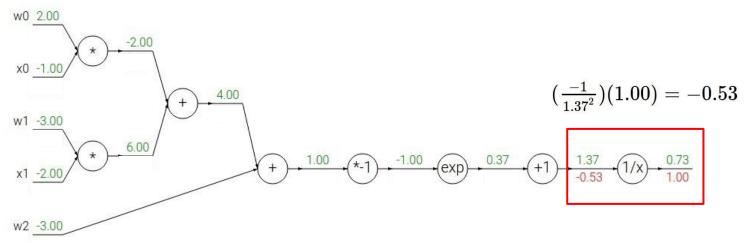
$$f(w,x) = rac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$$



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

$$f(x)=rac{1}{x} \qquad \qquad \qquad rac{df}{dx}=-1/x^2 \ f_c(x)=c+x \qquad \qquad \qquad \qquad rac{df}{dx}=1$$

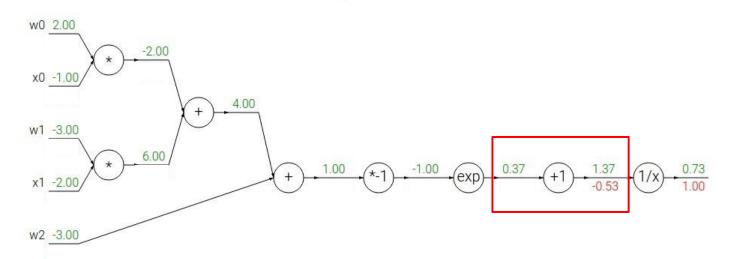
$$f(w,x) = rac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$$



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

$$f(x)=rac{1}{x} \qquad \qquad \qquad rac{df}{dx}=-1/x^2 \ f_c(x)=c+x \qquad \qquad \qquad \qquad rac{df}{dx}=1$$

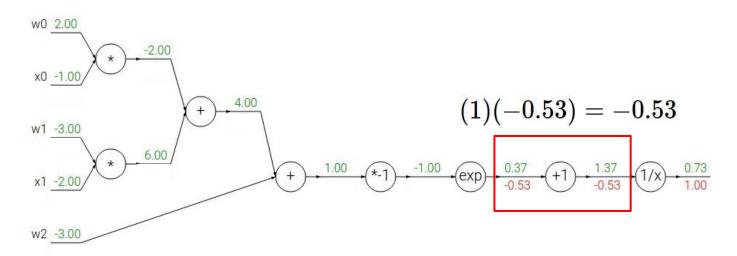
$$f(w,x) = rac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$$



$$egin{aligned} f(x) = e^x &
ightarrow & rac{df}{dx} = e^x \ f_a(x) = ax &
ightarrow & rac{df}{dx} = a \end{aligned}$$

$$f(x)=rac{1}{x} \qquad \qquad \qquad rac{df}{dx}=-1/x^2 \ f_c(x)=c+x \qquad \qquad \qquad \qquad rac{df}{dx}=1$$

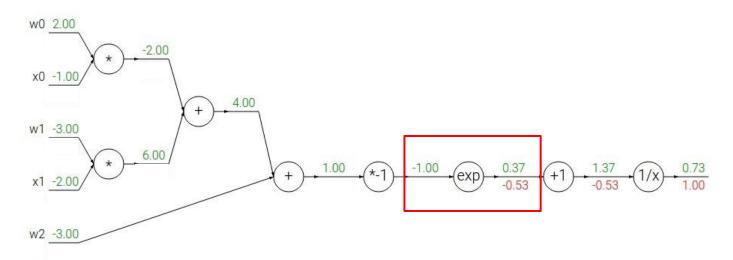
$$f(w,x) = rac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$$



$$egin{aligned} f(x) = e^x &
ightarrow & rac{df}{dx} = e^x \ f_a(x) = ax &
ightarrow & rac{df}{dx} = a \end{aligned}$$

$$f(x)=rac{1}{x} \qquad \qquad \qquad rac{df}{dx}=-1/x^2 \ f_c(x)=c+x \qquad \qquad
ightarrow \qquad rac{df}{dx}=1$$

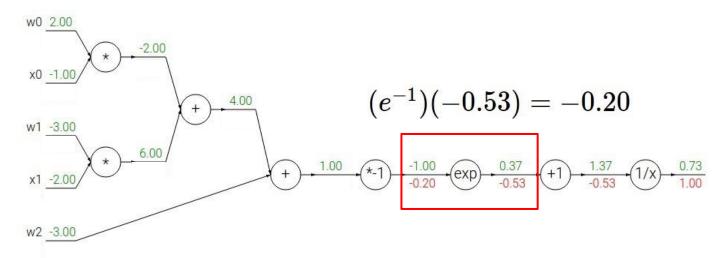
$$f(w,x) = rac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$$



$$f(x) = e^x \hspace{1cm} o \hspace{1cm} rac{df}{dx} = e^x \ f_a(x) = ax \hspace{1cm} o \hspace{1cm} rac{df}{dx} = a$$

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

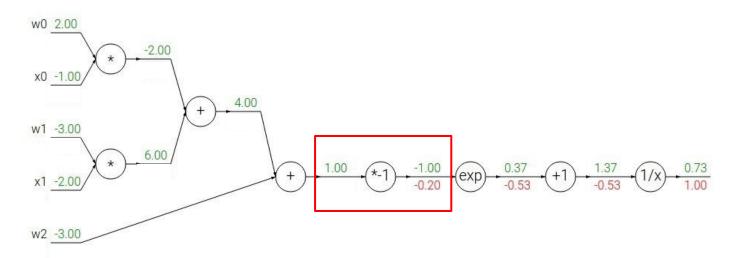
$$f(w,x) = rac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$$



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

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

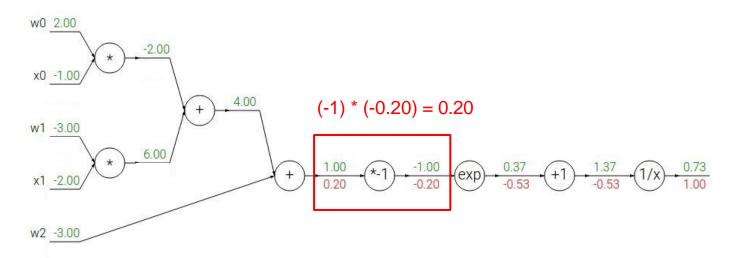
$$f(w,x) = rac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$$



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

$$f(x)=rac{1}{x} \qquad \qquad
ightarrow \qquad rac{df}{dx}=-1/x^2 \ f_c(x)=c+x \qquad \qquad
ightarrow \qquad rac{df}{dx}=1$$

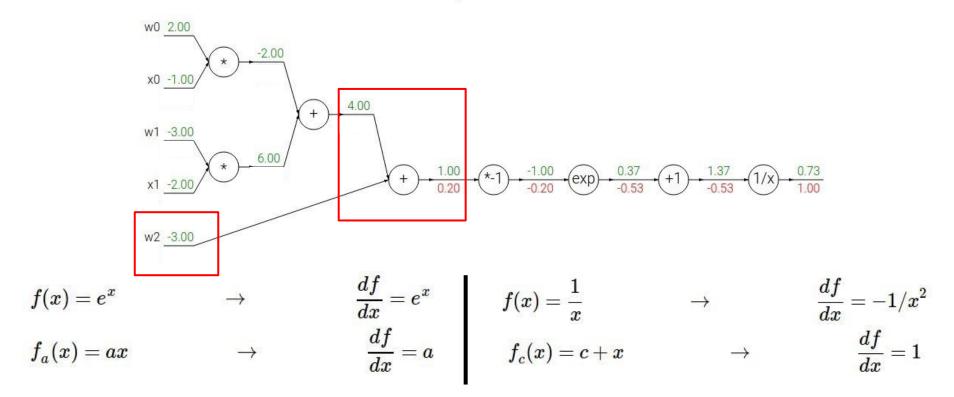
$$f(w,x) = rac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$$



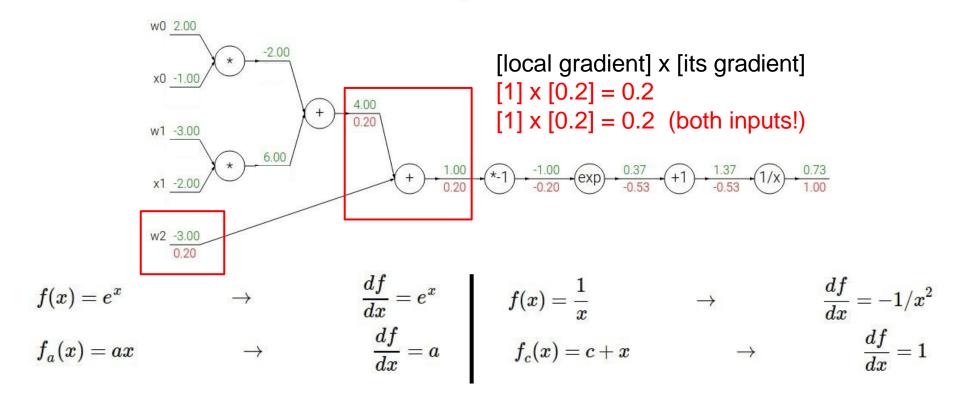
$$f(x) = e^x \qquad o \qquad rac{df}{dx} = e^x \ f_a(x) = ax \qquad o \qquad rac{df}{dx} = a$$

$$f(x)=rac{1}{x} \qquad \qquad
ightarrow \qquad rac{df}{dx}=-1/x^2 \ f_c(x)=c+x \qquad \qquad
ightarrow \qquad rac{df}{dx}=1$$

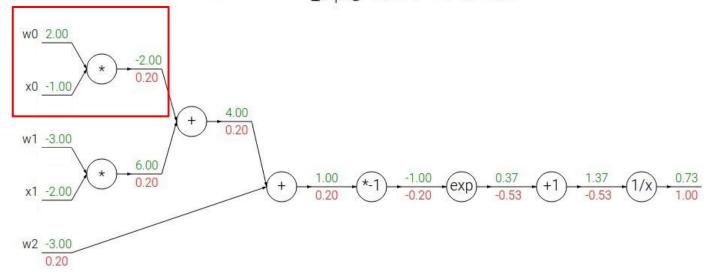
$$f(w,x) = rac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$$



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



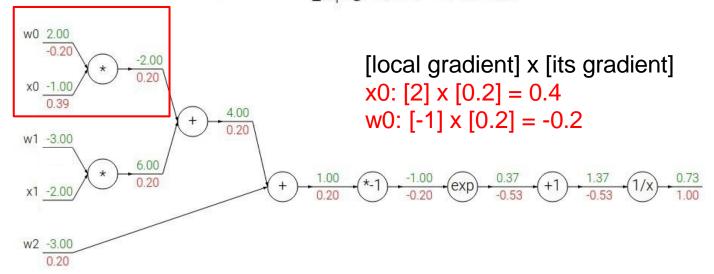
$$f(w,x) = rac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$$



$$egin{aligned} f(x) = e^x &
ightarrow & rac{df}{dx} = e^x \ f_a(x) = ax &
ightarrow & rac{df}{dx} = a \end{aligned}$$

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

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



$$egin{aligned} f(x) = e^x &
ightarrow & rac{df}{dx} = e^x \ f_a(x) = ax &
ightarrow & rac{df}{dx} = a \end{aligned}$$

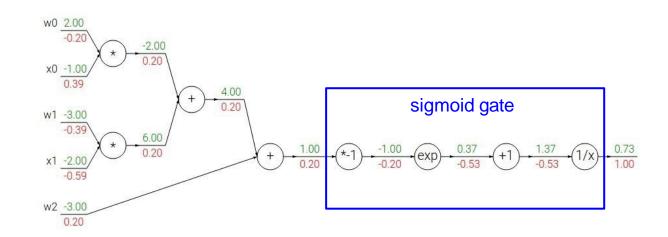
$$rac{df}{dx} = e^x \hspace{1cm} f(x) = rac{1}{x} \hspace{1cm}
ightarrow \hspace{1cm} rac{df}{dx} = -1/x \ rac{df}{dx} = a \hspace{1cm} f_c(x) = c + x \hspace{1cm}
ightarrow \hspace{1cm} rac{df}{dx} = 1$$

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

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

sigmoid function

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

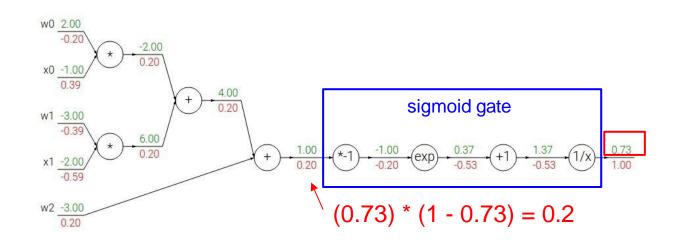


$$f(w,x)=rac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$

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

sigmoid function

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



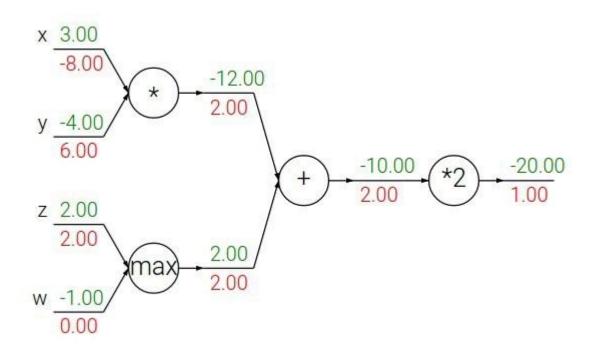
```
w = [2, -3, -3] # assume some random weights and data
x = [-1, -2]
# forward pass
dot = w[0]*x[0] + w[1]*x[1] + w[2]
f = 1.0 / (1 + math.exp(-dot)) # sigmoid function
# backward pass through the neuron (backpropagation)
ddot = (1 - f) * f # gradient on dot variable, using the sigmoid gradient derivation
dx = [w[0] * ddot, w[1] * ddot] # backprop into x
dw = [x[0] * ddot, x[1] * ddot, 1.0 * ddot] # backprop into w
# we're done! we have the gradients on the inputs to the circuit
```

Patterns in backward flow

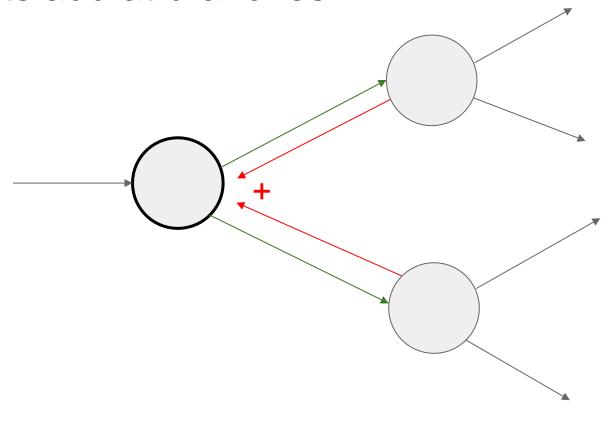
add gate: gradient distributor

max gate: gradient router

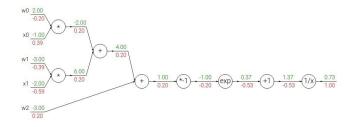
mul gate: gradient... "switcher"?



Gradients add at branches



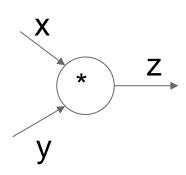
Implementation: forward/backward API



Graph (or Net) object. (Rough pseudo code)

```
class ComputationalGraph(object):
    # . . .
    def forward(inputs):
        # 1. [pass inputs to input gates...]
        # 2. forward the computational graph:
        for gate in self.graph.nodes topologically sorted():
            gate.forward()
        return loss # the final gate in the graph outputs the loss
    def backward():
        for gate in reversed(self.graph.nodes topologically sorted()):
            gate.backward() # little piece of backprop (chain rule applied)
        return inputs gradients
```

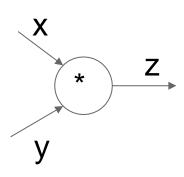
Implementation: forward/backward API



(x,y,z are scalars)

```
class MultiplyGate(object):
    def forward(x,y):
        z = x*y
        return z
    def backward(dz):
        \# dx = \dots \#todo
        \# dy = \dots \# todo
        return [dx, dy]
```

Implementation: forward/backward API



```
class MultiplyGate(object):
    def forward(x,y):
        z = x*y
        self.x = x # must keep these around!
        self.y = y
        return z
    def backward(dz):
       dx = self.y * dz # [dz/dx * dL/dz]
        dy = self.x * dz # [dz/dy * dL/dz]
        return [dx, dy]
```

(x,y,z are scalars)



Example: Torch Layers



R LosSecost.los	Add THNN conversion of [ELU, Lewis/FacLI, LogDigmoid, LogDidMax, Louisu	7 days ago
E LogSoltMax.tua	Add THNN conversion of (ELL), LeakyReLL), LogSigmoid, LogSoftMax, Looku	7 days ago
E LookupTatiin kae	Harmoniae Lookup Table eignature with curn impl	5 days ago
E Milio	Remote separal to table separal for Lue 5.3	E norths ago
E Margin Criterion has	Add Ston-Average to orientoris in the constructor modernized Marga-Criterius.	2 months ago a year ago
E Margin Flanking Grismon Aus	Fix book mode in Margin-Centring Citierion	A days ago
Marka	Merga pull reguest 9404 from vg/re-invaster	Z mooths ago
E Mean ha	Add support for requires dimension and both batch and non-batch input.	Z months ago
Mobile	Berge pull request P464 from vypolinaster	2 months ago
History/Tubis Ass	consul sinused variable and usaless expression	20 days ago
F Module has	Revert "Don't re-Satien parameters if they are already Satiened"	15 hours ago
Mallon	removing the requirement for jetoviding size in m.Mul.	a year ago
NoContent has	Ignore (prode/Goodings) if the Egypotropic to reli	2 months ago
E MultiCriterion has	asserts in MJSO/belon and PassleOrbeton add	2 months ago
MultiLabetMarginCriterion lua	sitted revenue of large? Inve	# years ago
MultiMarginCriterion lus	multimargin supports pr2	Tf months ago
Narrow lus	typeAs in Narrow not done in place.	6 months ago
€ NarravTable.loa	NanosTatie	Enorthiag
Normaliza has	Remove brom and heddfrom from Normalize, because they allocate memory,	25 days agr
PReLI/Ion	Suffers for PReCD code implementation.	Emmis ag
Pedding lue	fixed broken on Pekiding Input was returned in backgrop	5 months ago
PairwiseDistance lus	Merge pull request #532 from swgengimenter	29 days ago
Parate/lux	ts a bug is conditional expression	a north agr
ParafelCitation lux	accerts in MultiCriterion and ParallelCriterion add	Z mooths ago
Parate/Table.lus	Paradel aptinization, Farate/Table inherbs Container, unit tests	a year age
El Power tus	Size UNIX line andings	T roochs ago
READINE and	doc madifiedoca	S morths ago
WiteLU law	Add randomized leaky rectified linear unit (ffffaLU)	2 months ago
ReLU law	white in-place ReCU and fixes a potential divide-by-zero in nr. Sign.	I months ago
E Replicate has	Replicate betchlicke	Il months ago
Restope lus	Added more informative pretty printing.	a year ago
Select lus	initial reviewp of torch? tree	4 years ago
SelectTable lus	rn Module preserve type sharing semantics (RTST); add in Module apply	4 months ago
Sequential has	Taling Sequential reviews comer cases initial revience of torch? Inse	fi months ago
Signoid Iua		A years ago
E Snootil 10 terminae	Add SpeAverage to orderions in the constructor Fit carrous shaded natiables in m	2 months ago a year ago
SoftMedian	Fix validate unlated nariables in mi	a year ago
D Soffhalus	fire Carramonial issues in the SoffPlus module (it inmaks for most p	2 years ago
SoftStork.lux	avital revents of longs? Inse	4 years ag
SotSign ka	milial owarra of territy tree	4 years ago
Sparseulacoban-kus	Fix various unused variables in nn	is year ago
Spared Insur Na	Using spanse implementation of zeroDradPoremeters for SpanseLinear	a month ago
Spatial Adaptive Man Pooling		A year ago
Spatial Average Proling los	Spatia/kerepePacing supports padding, call mode and excitate paid thr	29 days ago
Spetalist/Normalization to	Add C implementation of Spatialbeth Normalization	7 days agr
Spatial Contrastivationnair	Make type() truly recursive.	S months ago
Speta/Convolution/up	Fis type() in SpalasConvolution	3 months ago
SpetialConvolutionAM has	Fix type() in SpatialConvolution	3 months ago
Spetia/ConsolutionMapilus	Remove unused and expensive initialization logic from in SpatialCooks.	Emething
SpatialCrossMapLRNLtus	coda consistency	16 days ago
Spatial Division Normalization	Spatial/Constructive,Director, Subtractive/Normalization work with but	Emortis agr
Spatia/Dropout/ua	arrull fix on error message	6 months ago
SpetaFracionatMerPools	Adding Friedenal Max Pooling	2 months ago
SprinFulConvolution/us	Add adjustment term in SpatisFulConvolution to control the size of	5 days ago
SpetaFulConvolutorMap ha		3 years ago
Spetal.PPoologius	Spela/Average/Fooling divides by kill/fillH	10 months ago
Spatia Max Pooling Lia	SpatisMaxFooling supports padding and cell mode	E months ago
SpetaMarUrpooling.ka	Add SpatisMarLingcoling	26 days ago
Spetial SoftMax.lua	Update SntMax to work in spatial mode.	4 months ago
SpelieSubSempling laid	Merge branch 'on_fast_weet'	3 years ago
Spatial Submictive Normaliz	Spalia@Constructive.Dinaive,Subtractive(Normalization work with but	8 months ago
Spatia/UplianpingNearest.L.	Use UNIX line endings	7 months ago
Speta@eroPatring.kos	Added non-informative prelty-printing	a year ago
D Solffable has	Add a conset for consetion indices in on Solit Fable	Toronto con

Example: Torch Layers



® Logispodia	Add THNN conversion of IELU, Leaty, Fal.U, LogSigmoid, LogSigMax, Looks	7 days ago
E LogSofMax ton	Add THNN comercion of (ELU, LeatyReLU, LogSigneid, LogSofMax, Looku	7 days ago
El LookupTable Nas	Harmoniae Lookup Tablis signature with core impl.	5 days ago
E Milia	Reserve unpack to table unpack for Lue 5.3	Enords ago
(E) MSECriterion lux	Add Ston-Average to orderions in the constructor	2 months ago
E Margin Criterion has	modernized MarginCriterius	a year ago
MarginRankingCelaron.ka	Fix botch mode in Margin-RankingCriterion	4 days ago
€ Macta	Merge pull request 8464 from vigins insister	Z morths ago
E Manha	Add support for requires dimension and both batch and non-batch input	Z months ago
E Mahe	Marga piul respect #464 fron vyjochruster	2 months ago
E MinneyTable kee	control strugged variable and useless expression	29 days ago
E Nodule has	Revert 'Don't re-faster parameters if they are already flattened'	15 hours ago
© Matter	removing the requirement for providing size in mr. Mul.	it year ago:
E NoConstant has	Sprore update Goodings of self-gradings for self-	2 months ago
El MultiCrismo has	asserts in MUSC theter and PassilelCriterion sold	2 months ago
Multi_abstNargirCriteron lue	sitted revenue of large? tree	#years ago
€ MultMarginCriterion lus	multimargin supports pr2	TT months ago-
El Narrow lust	typeAs in Narrow not done in place.	6 months ago
El NantowTable.loa	NarosTable	E norths ago
IP Normaliza has	Ranges from and haddings from Normalize, because they allocate memory	20 days ago
El PReLU/sus	Suffers for PfleCD code implementation.	E months ago
E Fedding lus	familianium on Fedding: Irgut was returned in bestigrap	5 months age
E ParwiseDatance lue	Merge pull request #512 from yappengimaster	29 days ago
E Facilities	fix a bug is conditional expression	a north age
R ParallelCitation has	asserts in Multi-Orderion and Parallel-Ortedon add	Z months ign
© Parale/Table has	Paradel optimization: ParadelTable intents Contained, until tests	a year ago
El Power has	Clea UNIX line androgs	7 months ago
IR READINE and	the malffedies	Sinundha ago.
St Strattler	Add and mand leavy wolfed linear unit (MfeLU)	3 months ago
E Relifice	wide in-place ReCU and fees a potential divide-by-pero in not Sign.	Fronth ago
E Redom ka	Section betalling	E months ago
E Festage Las	Added more informative prefly proting.	A year ago
R. Select ka	eltal reveno of tech? tee	4 9990 900
El SelectTable lus	m Module preserve type sharing semantics (#167); add m Module apply	4 months ago
Separtid ha	firing Seguential ventries comer case	6 months ago
E Signoid tas	initial revenue of toruth? Inse	4 years ago
E Seprential Seprential	Add Spakwege to orienous in the constructor	2 months ago
Si Sottler ha	Fig. springs properly satisfies in m	
El Sofffenius		a year ago
El Soffinius	Fix various unused variables in on	a year ago
	fixed a numerical issue in the SoffPlus module (it breaks for most $\mathfrak g$.	2 years ago
El SotShrekke	aritial reviewp of loach? Inne	4 years ago
B SotSign.ka	initial reverse of toron? tree	if years ago
€ SparseJacobian kua	Fix various unused variables in m	is year ago
E SparseLinear has	Using sparse implementation of zeroDradParameters for SparseLinear	a rooth ago
E Spatial Adaptive Main Pooling	Attled Spatial-Septivalities Pooling	k year ago
Spetal/veragePooling-los	Spatia/LesspePsoing supports padding, call mode and austable, pad div	29 days age
Speta/Eat/Normalization Ave	Add C implementation of Spatial Batch Normalization	7 days ago
E Spetis ContractiveNormals	Make type() truly recursive.	S months ago
E Spatia/Convolution lue	Fir type() in SpalaiConvolution	3 months age
Spela/ConvolutionMitting	Fix type() in SpatistConvolution	7 months ago
Spette/ConsolutionMapilips	Ramove unused and expensive initialization logic from in: SpatialConvis	8 months ago
■ SpatialCrossMapLRNLtus	code consistency	16 days ago
Spatia/DivisiveNermalizatio	Spalle(Constructive, Director, Subtractive)Normalization work with two	E months ago
E Spatia Dropout/Lea	and fain ero mesage	6 months ago
■ SpetaFracionalMeiPools	Adding Frederical Max Posing	2 months ago:
E SprinFulConvolution.lus	Add adjustment term in SpatialFullConvolution to control the size of	5 days ago
□ SpetisFulConvolctorMap has		3 years ago.
Spetial_PPoologius	Spella/Average/Pooling divides by kill/NH	10 months ago
■ SpatiaNewPooling.lue	SpatialNetPooling supports padding and cell mode	E months ago
SpataMartingcoling.tua	Att SpatisMarLispooling	26 days ago
Spetia/SoffMax.lus	Update SntMax to work in spatial mode.	4 months ago
E SpelieSubSemplinglue	Merge branch 'on_fast_reset'	3 years ago
E Spatial Submicherbornals	Spatial(Constraintee, Division, Subtractive)Normalization work with but.	8 months ago
E Spatial/plamping/learnet/	Use UNIX ine endings	7 months ago:
SpetalZeroPadding kas	Added non-informative pretty-printing	A year ago
El SoltTable his	Add account for connection in front in on Sold Fathin	Tunnelly and







local MulConstant, parent = torch.class('nn.MulConstant', 'nn.Module') function MulConstant: init(constant scalar,ip) parent.__init(self) assert(type(constant_scalar) == 'number', 'input is not scalar!') self.constant_scalar = constant_scalar -- default for inplace is false self.inplace = ip or false if (ip and type(ip) ~= 'boolean') then error('in-place flag must be boolean') function MulConstant:updateOutput(input) if self.inplace then input:mul(self.constant_scalar) self.output = input else self.output:resizeAs(input) self.output:copy(input) self.output:mul(self.constant_scalar) return self.output function MulConstant:updateGradInput(input, gradOutput) if self.gradInput then if self.inplace then gradOutput:mul(self.constant_scalar) self.gradInput = gradOutput -- restore previous input value input:div(self.constant_scalar) else self.gradInput:resizeAs(gradOutput) self.gradInput:copy(gradOutput) self.gradInput:mul(self.constant_scalar) return self.gradInput

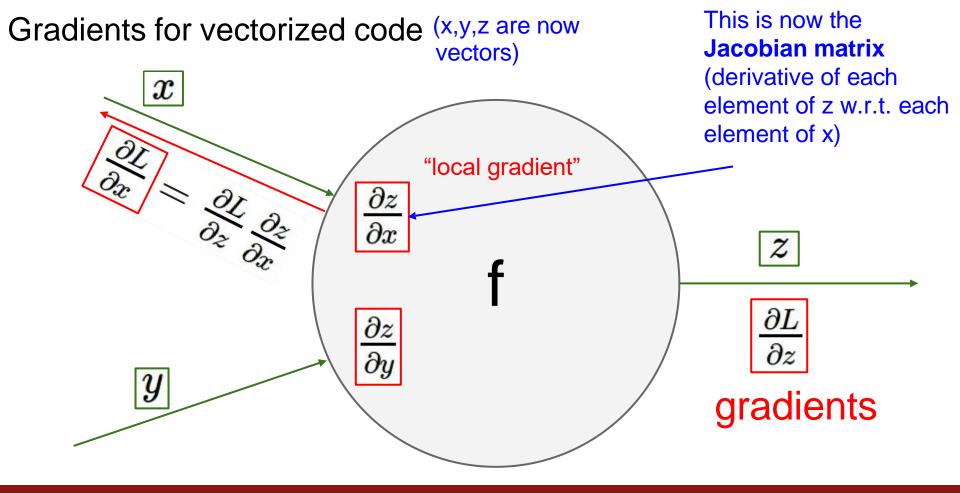
Example: Torch MulConstant

$$f(X) = aX$$

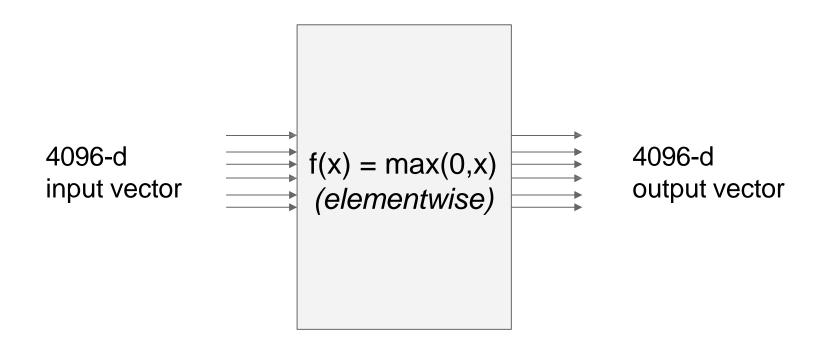
initialization

forward()

backward()

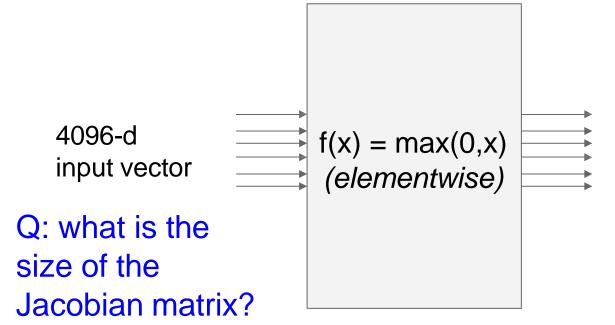


[slides] [backprop notes] [Efficient BackProp] (optional) related: [1], [2], [3] (optional) [slides] handout 1: Vector, Matrix, and Tensor Derivatives handout 2: Derivatives, Backpropagation, and Vectorization Deep Learning [Nature] (optional) [slides] tips/tricks: [1], [2] (optional)



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

Jacobian matrix



4096-d output vector

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

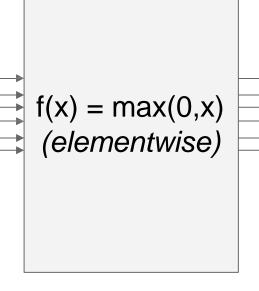
Jacobian matrix

4096-d input vector

Q: what is the size of the

Jacobian matrix?

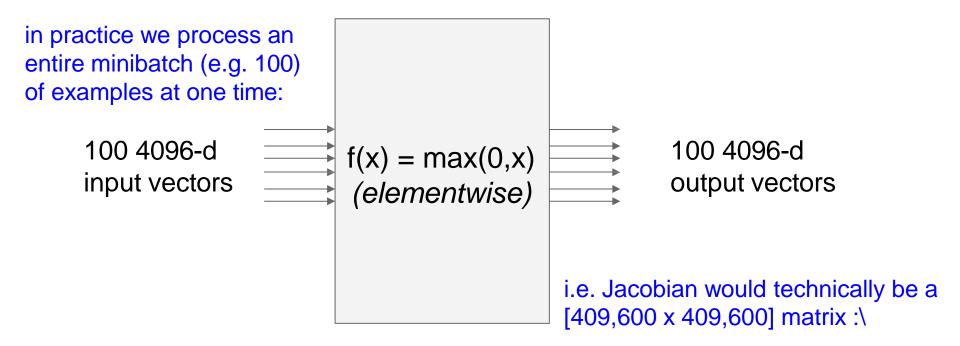
[4096 x 4096!]



output vector

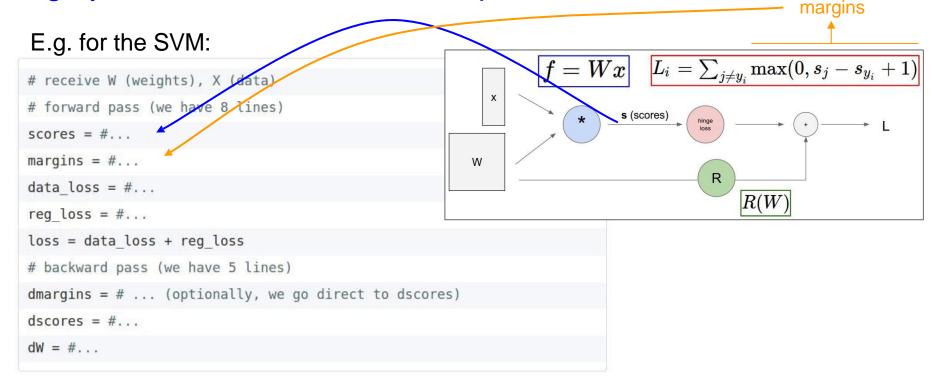
4096-d

Q2: what does it look like?



Assignment: Writing SVM/Softmax

Stage your forward/backward computation!



Summary so far

- neural nets will be very large: no hope of writing down gradient formula by hand for all parameters
- backpropagation = recursive application of the chain rule along a computational graph to compute the gradients of all inputs/parameters/intermediates
- implementations maintain a graph structure, where the nodes implement the forward() / backward() API.
- forward: compute result of an operation and save any intermediates needed for gradient computation in memory
- **backward**: apply the chain rule to compute the gradient of the loss function with respect to the inputs.