Lecture 6: Backpropagation Vector, Matrix and Tensor Derivatives

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller Lecture 6 - 1 Sept. 19, 2024

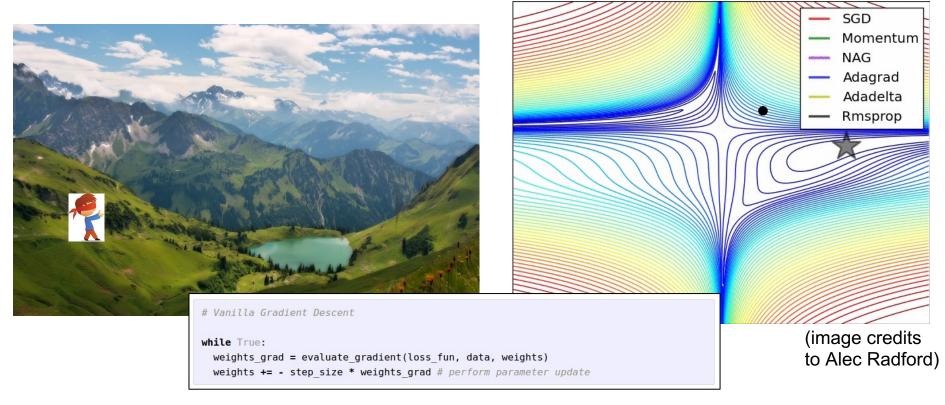
Where we are ...

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



Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller Lecture 6 - 2 Sept. 19, 2024

Optimization



Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 6 - 3 Sept. 19, 2024

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

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller Lecture 6 - 4 Sept. 19, 2024

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.

Lecture 6 - 5

Sept. 19, 2024

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller 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":

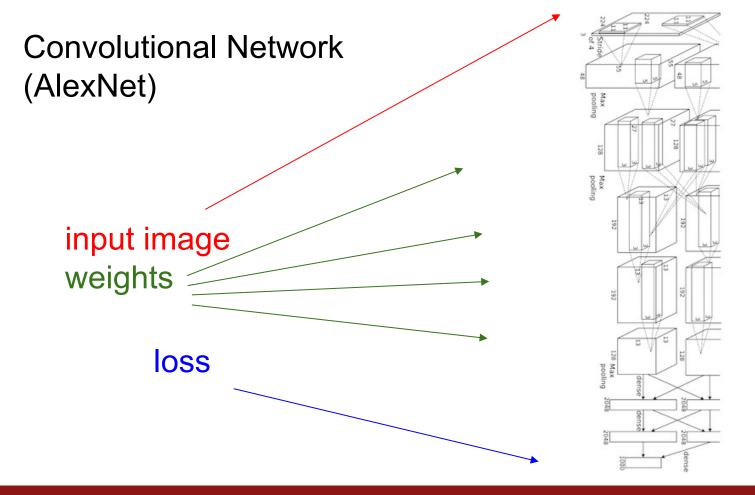
$$abla z = \begin{bmatrix} 2x \\ 2y \end{bmatrix}$$
. **Don't care about this**

2024

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

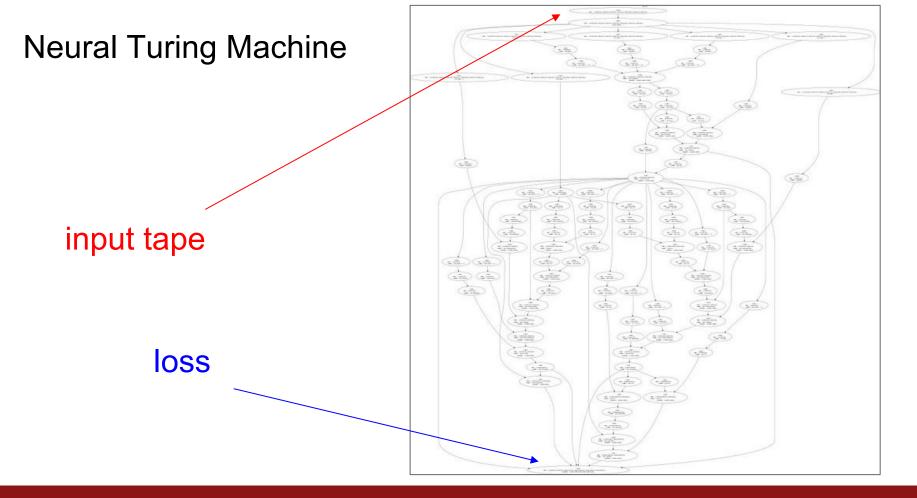
$$abla z(5,3) = \begin{bmatrix} 2 \times 5 \\ 2 \times 3 \end{bmatrix} = \begin{bmatrix} 10 \\ 6 \end{bmatrix}.$$
Do care about this





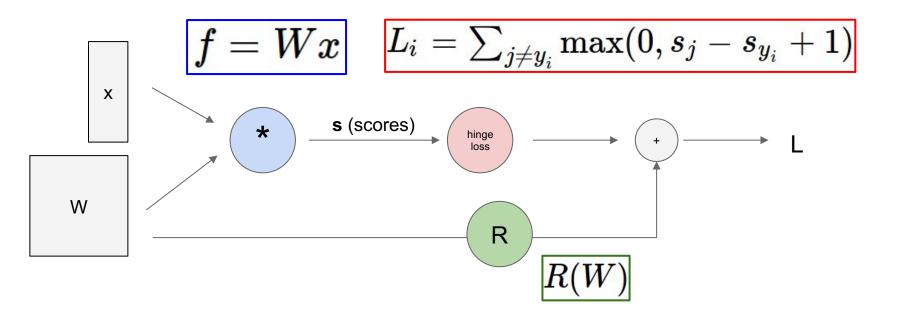
Lecture 6 - 7 Sept. 1

Sept. 19, 2024



Lecture 6 - 8 Sept. 19, 2024

Computational Graph



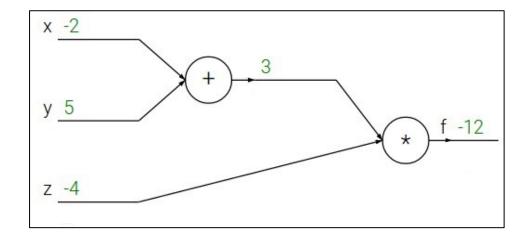
Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 6 - 9 Sept. 19, 2024

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

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.



Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 6 - 10 Sept. 19, 2024

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 dfdq = z # df/dq = z, so gradient on q 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

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 6 - 11 Sept. 19, 2024

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

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.

$$x -2$$

$$y -2$$

$$y -5$$

$$x -2$$

$$x -2$$

$$x -2$$

$$x -2$$

$$x -2$$

$$x -12$$

Goal is to calculate

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

evaluated at the point

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

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller Lecture 6 - 12 Sept. 19, 2024

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

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

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

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

$$\frac{\partial f}{\partial t} \quad \partial f$$

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

Lecture 6 - 13 Sept. 19, 2024

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

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

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

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

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

$$x = -2, y = 5, z = -4$$

$$y = \frac{1}{y} =$$

Lecture 6 - 14 Sept. 19, 2024

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

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

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

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

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

Lecture 6 - 15 Sept. 19, 2024

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

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

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

$$f = qz \qquad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$
Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$

Lecture 6 - 16 Sept. 19, 2024

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

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

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

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

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

Lecture 6 - 17 Sept. 19, 2024

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

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

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

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

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

Lecture 6 - 18 Sept. 19, 2024

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

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

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

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

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

Lecture 6 - 19 Sept. 19, 2024

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

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

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

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

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

Lecture 6 - 20 Sept. 19, 2024

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

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

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

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

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

Lecture 6 - 21 Sept. 19, 2024

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

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

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

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

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

$$\overset{x -2}{4} \qquad (x + y) = \frac{q}{4} \qquad (x + y) = \frac{q}{4$$

Lecture 6 - 22 Sept. 19, 2024

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

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

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

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

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

Lecture 6 - 23 Sept. 19, 2024

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

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

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

$$f = qz \qquad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$
Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$
Chain rule:

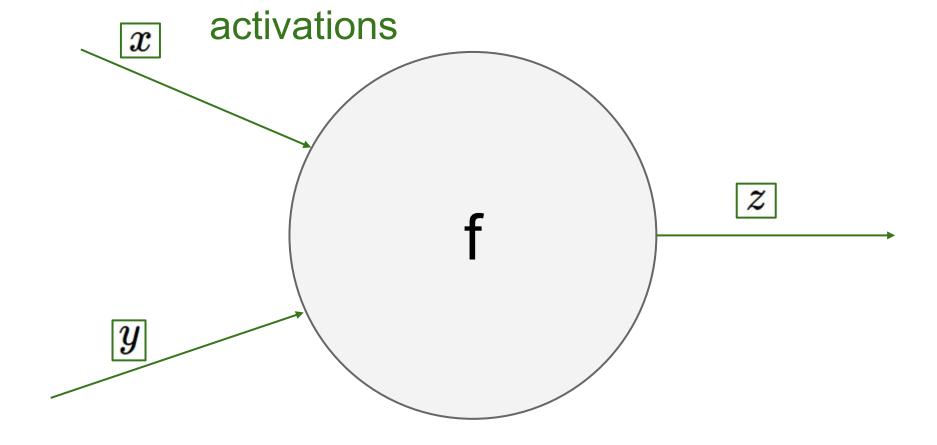
$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial q} \frac{\partial q}{\partial x}$$

Lecture 6 - 24 Sept. 19, 2024

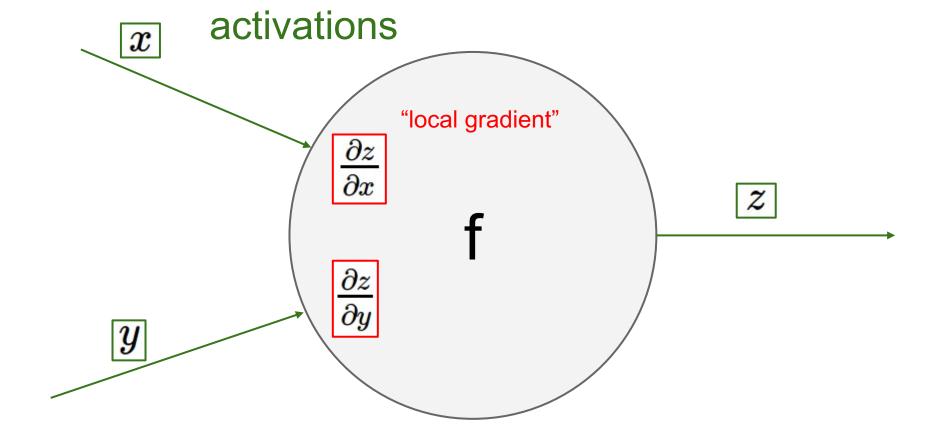
set some inputs
x = -2; y = 5; z = -4
perform the forward pass
q = x + y # q becomes 3
f = g * z # f becomes -12

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 dfdq = z # df/dq = z, so gradient on q 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

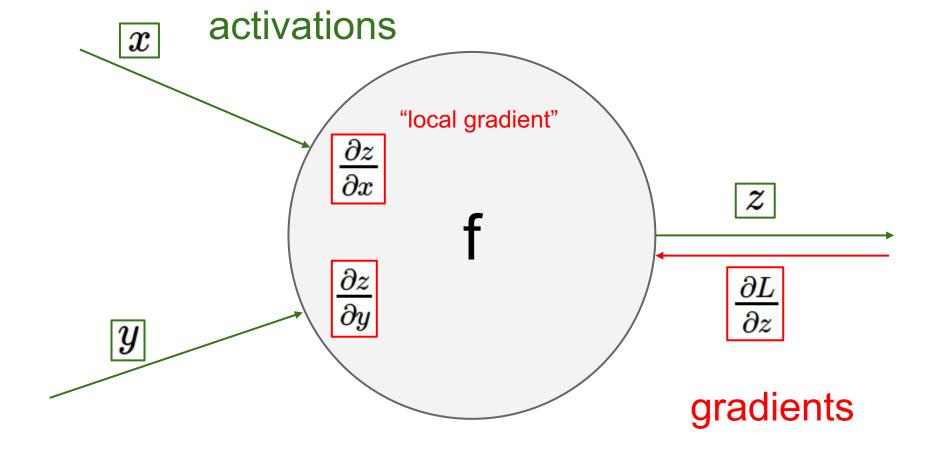
Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller Lecture 6 - 25 Sept. 19, 2024



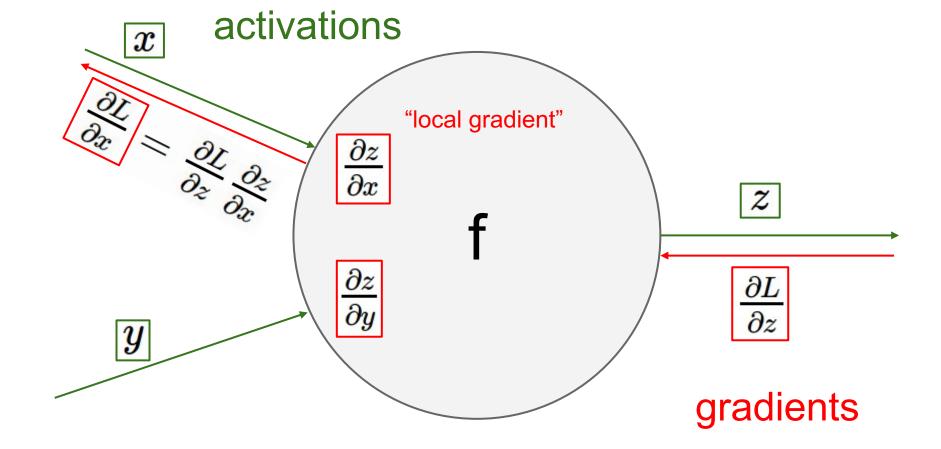
Lecture 6 - 26 Sept. 19, 2024



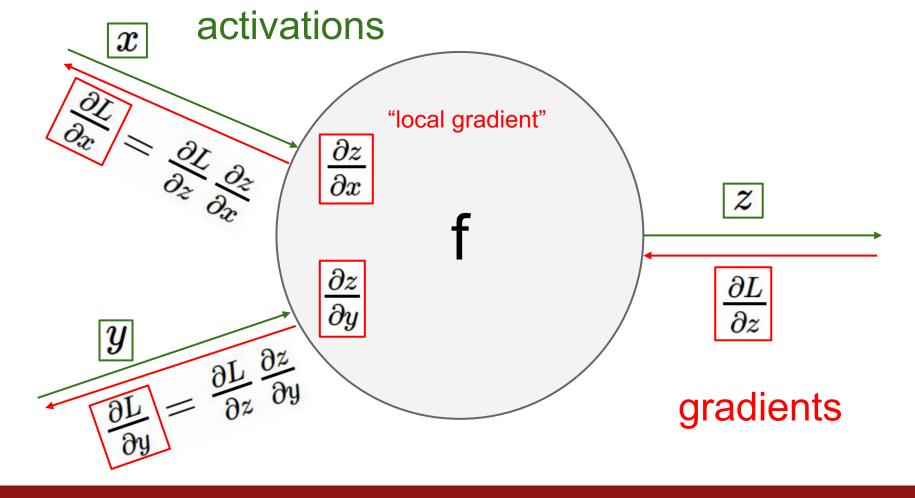
Lecture 6 - 27 Sept. 19, 2024



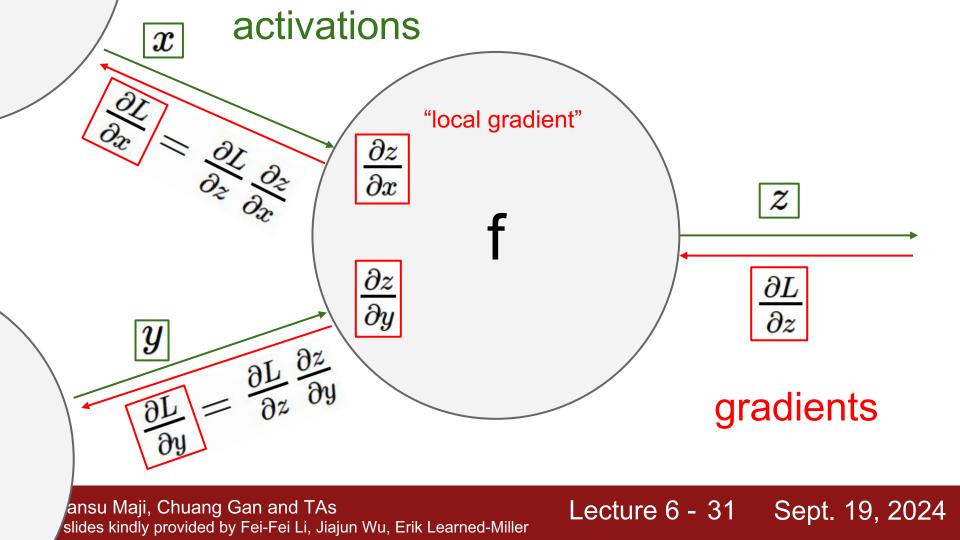
Lecture 6 - 28 Sept. 19, 2024



Lecture 6 - 29 Sept. 19, 2024

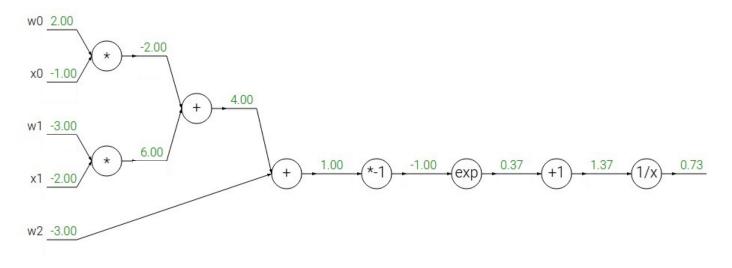


Lecture 6 - 30 Sept. 19, 2024



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

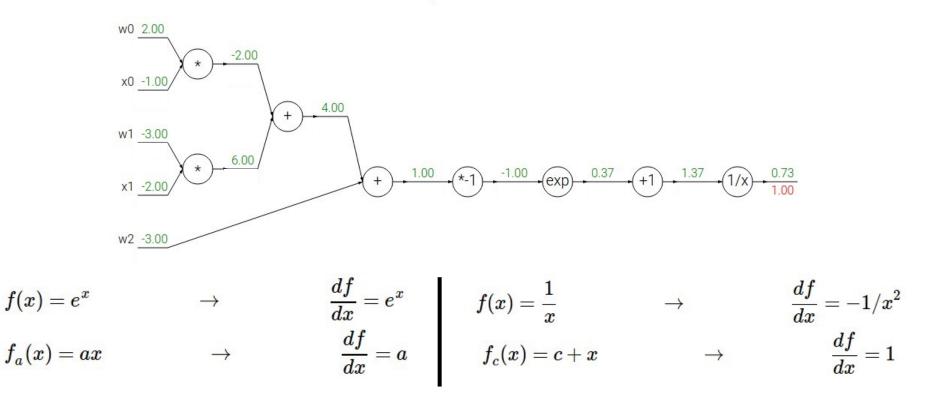
"sigmoid function"



Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 6 - 32 Sept. 19, 2024

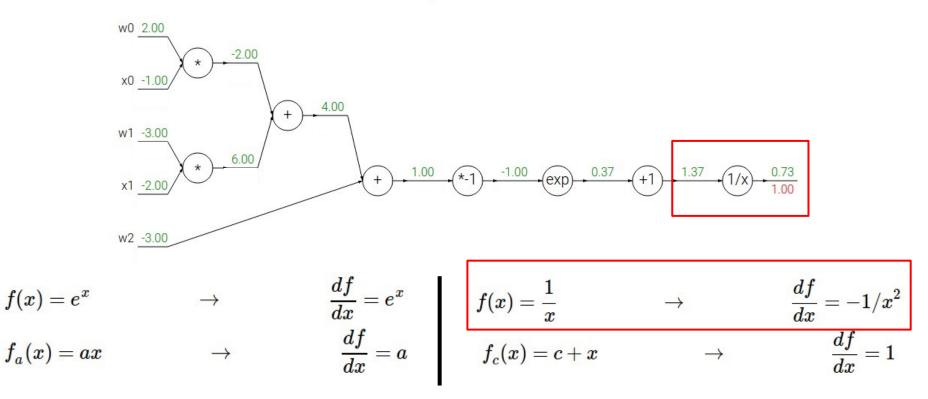
$$f(w,x)=rac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$



Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 6 - 33 Sept. 19, 2024

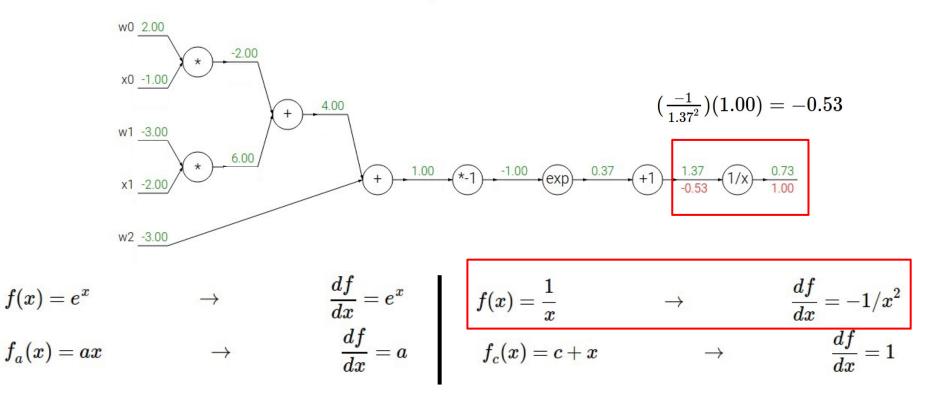
$$f(w,x)=rac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$



Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 6 - 34 Sept. 19, 2024

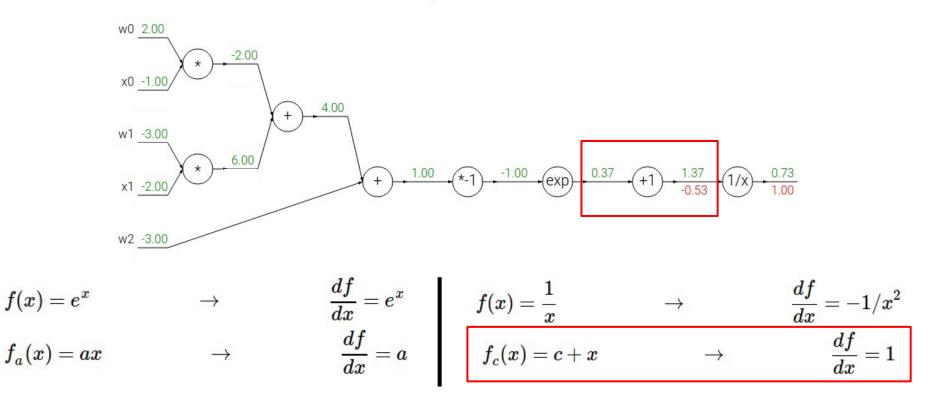
$$f(w,x)=rac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$



Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

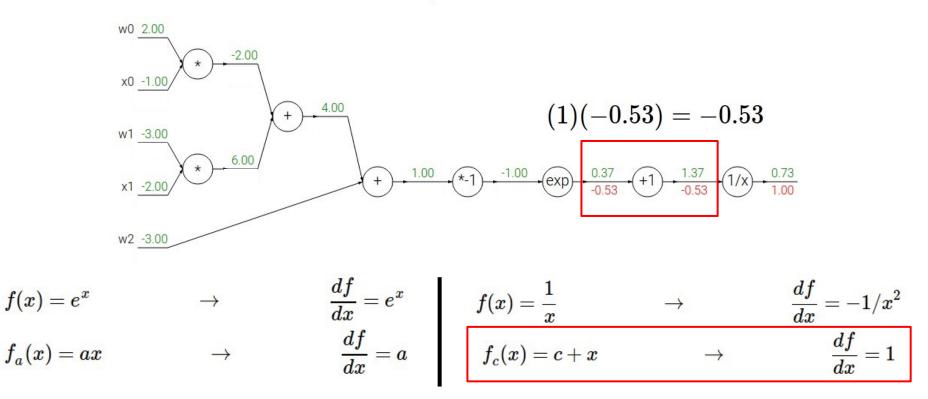
Lecture 6 - 35 Sept. 19, 2024

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



Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller Lecture 6 - 36 Sept. 19, 2024

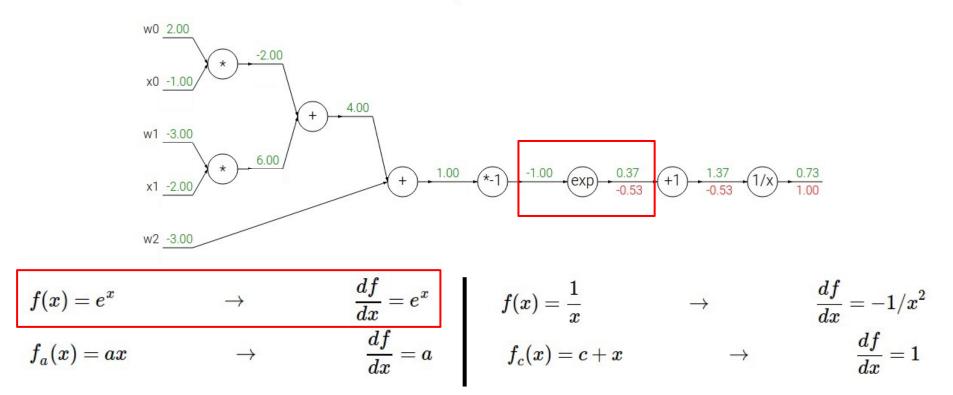
$$f(w,x)=rac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$



Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 6 - 37 Sept. 19, 2024

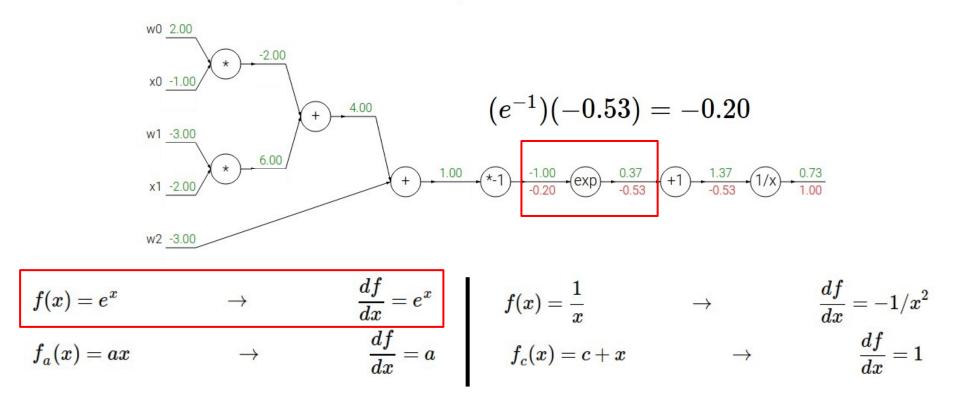
$$f(w,x)=rac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$



Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 6 - 38 Sept. 19, 2024

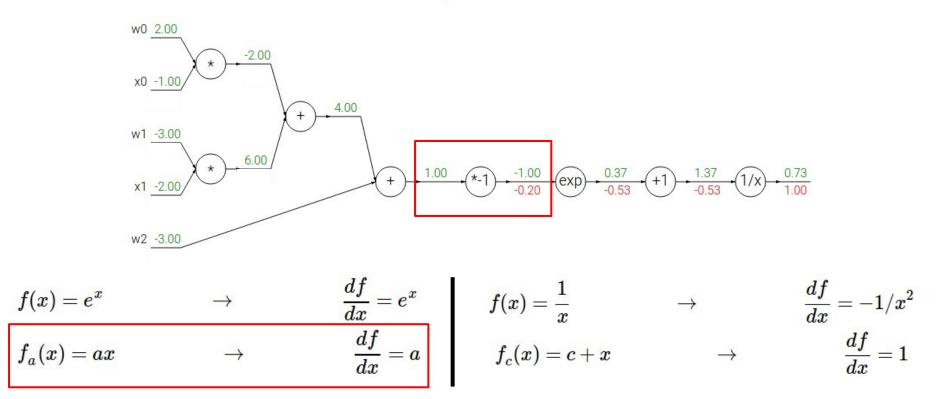
$$f(w,x)=rac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$



Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 6 - 39 Sept. 19, 2024

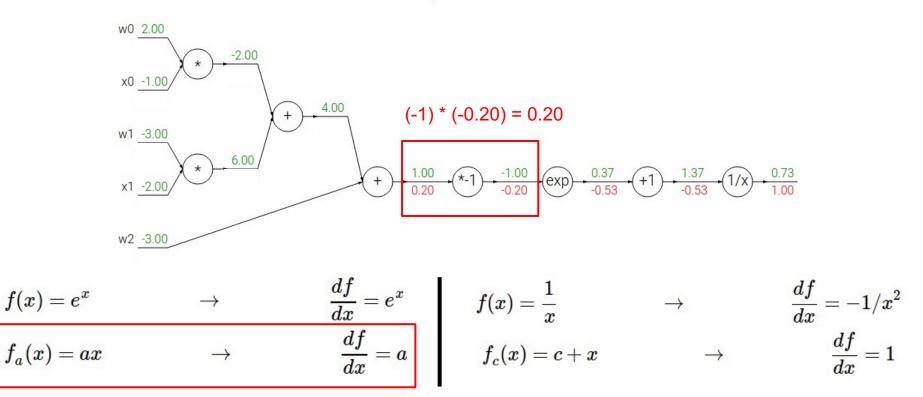
$$f(w,x)=rac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$



Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

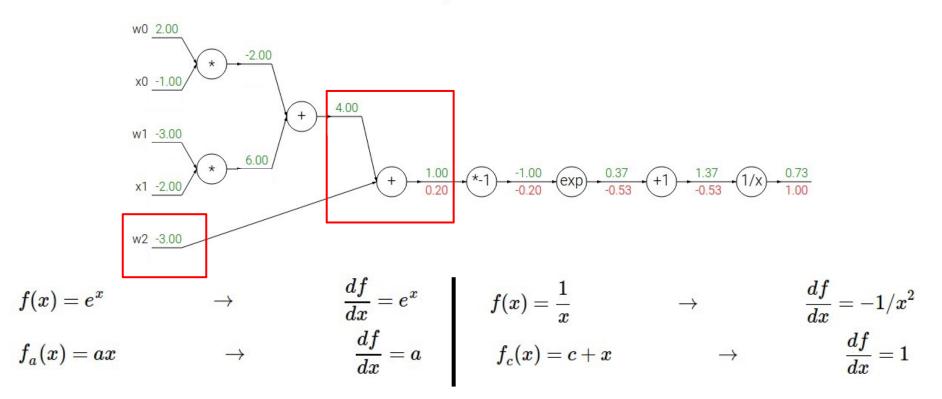
Lecture 6 - 40 Sept. 19, 2024

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



Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller Lecture 6 - 41 Sept. 19, 2024

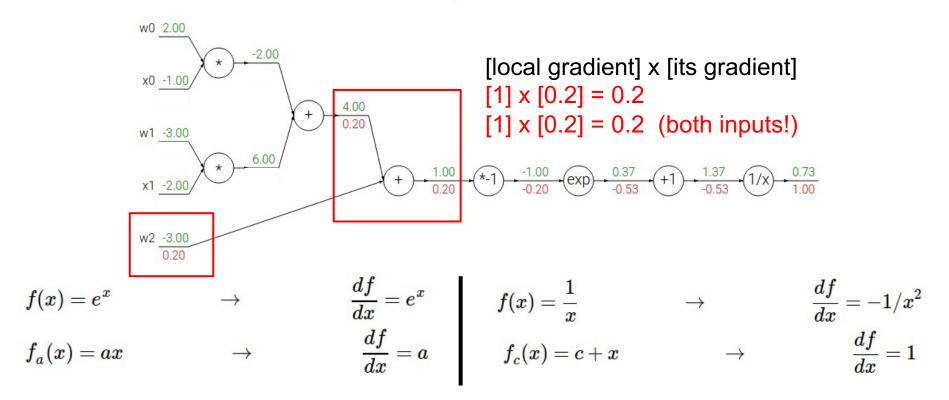
$$f(w,x)=rac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$



Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

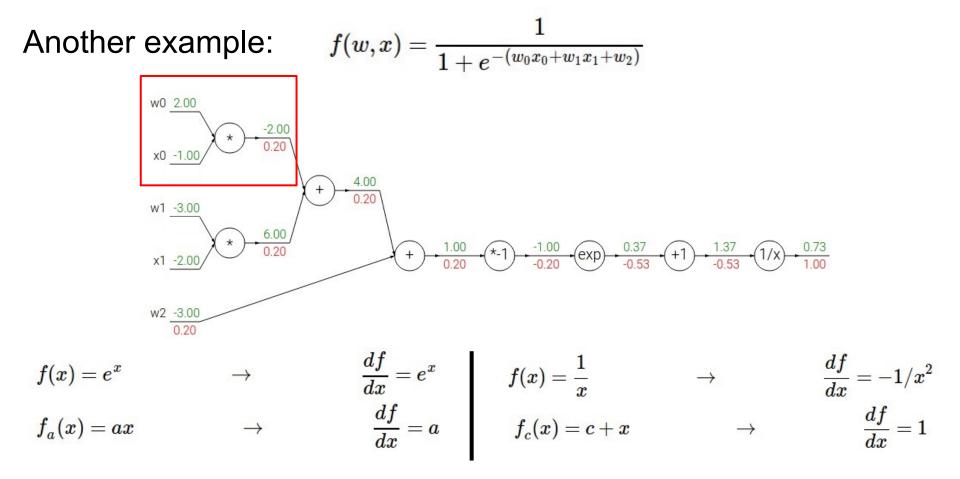
Lecture 6 - 42 Sept. 19, 2024

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

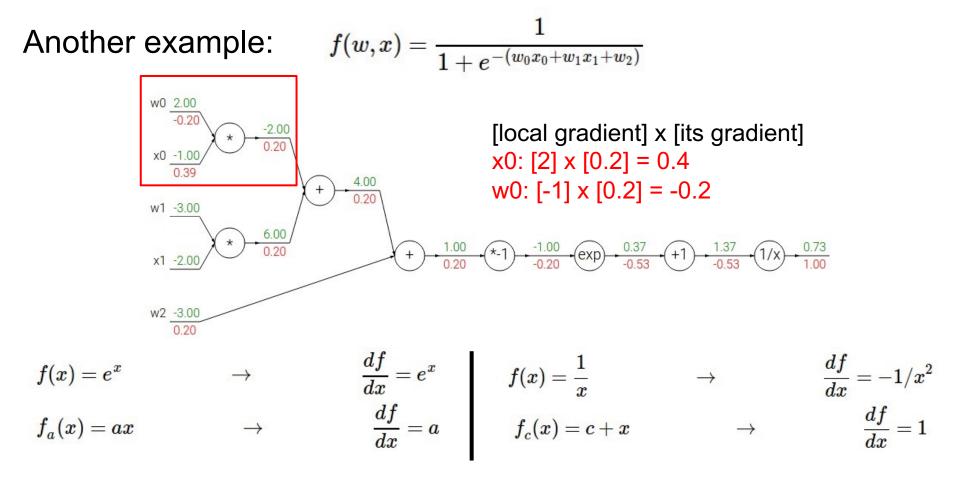


Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 6 - 43 Sept. 19, 2024

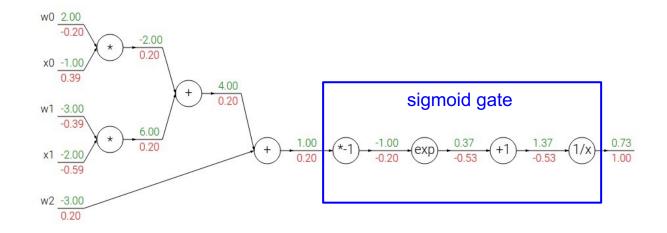


Lecture 6 - 44 Sept. 19, 2024



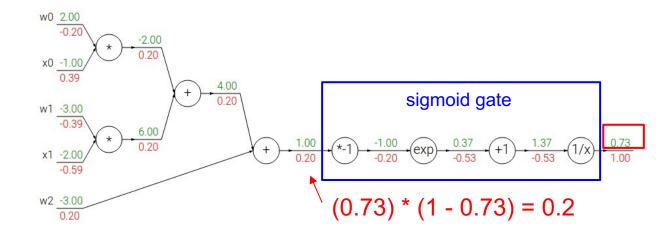
Lecture 6 - 45 Sept. 19, 2024

$$f(w,x) = \frac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}} \qquad \qquad \sigma(x) = \frac{1}{1 + e^{-x}} \quad \text{sigmoid function}$$
$$\frac{d\sigma(x)}{dx} = \frac{e^{-x}}{(1 + e^{-x})^2} = \left(\frac{1 + e^{-x} - 1}{1 + e^{-x}}\right) \left(\frac{1}{1 + e^{-x}}\right) = (1 - \sigma(x))\sigma(x)$$



Lecture 6 - 46 Sept. 19, 2024

$$f(w,x) = \frac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}} \qquad \qquad \sigma(x) = \frac{1}{1 + e^{-x}} \quad \text{sigmoid function}$$
$$\frac{d\sigma(x)}{dx} = \frac{e^{-x}}{(1 + e^{-x})^2} = \left(\frac{1 + e^{-x} - 1}{1 + e^{-x}}\right) \left(\frac{1}{1 + e^{-x}}\right) = (1 - \sigma(x))\sigma(x)$$



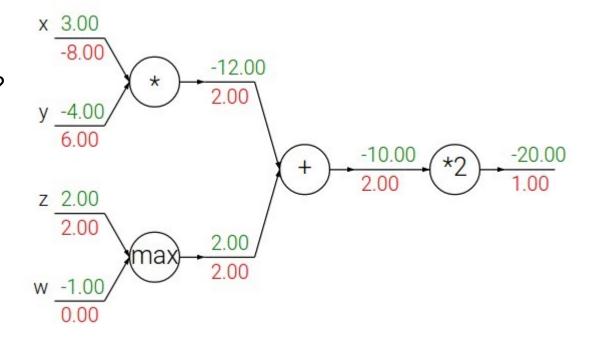
Lecture 6 - 47 Sept. 19, 2024

```
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
```

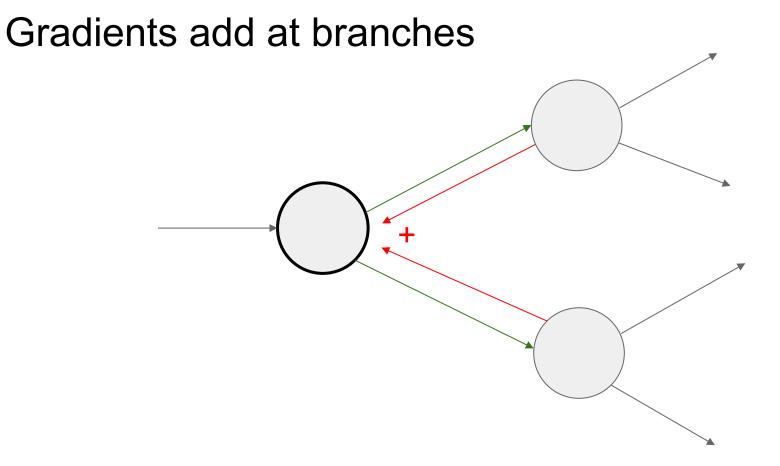
Lecture 6 - 48 Sept. 19, 2024

Patterns in backward flow

add gate: gradient distributor
max gate: gradient router
mul gate: gradient... "switcher"?

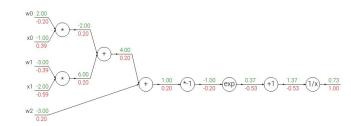


Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller Lecture 6 - 49 Sept. 19, 2024



Lecture 6 - 50 Sept. 19, 2024

Implementation: forward/backward API



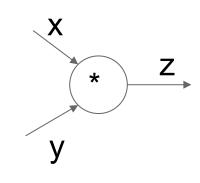
Graph (or Net) object. (Rough pseudo code)

<pre>class ComputationalGraph(object):</pre>
#
<pre>def forward(inputs):</pre>
<pre># 1. [pass inputs to input gates]</pre>
2. forward the computational graph:
<pre>for gate in self.graph.nodes_topologically_sorted():</pre>
gate.forward()
<pre>return loss # the final gate in the graph outputs the loss</pre>
<pre>def backward():</pre>
<pre>for gate in reversed(self.graph.nodes_topologically_sorted()):</pre>
<pre>gate.backward() # little piece of backprop (chain rule applied)</pre>
return inputs_gradients

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 6 - 51 Sept. 19, 2024

Implementation: forward/backward API



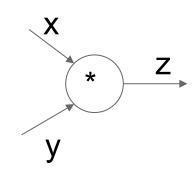
class MultiplyGate(object): def forward(x,y): $z = x^*y$ return z def backward(dz): # dx = ... #todo # dy = ... #todo ∂L return [dx, dy] ∂z ∂x

(x,y,z are scalars)

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 6 - 52 Sept. 19, 2024

Implementation: forward/backward API



<pre>class MultiplyGate(object):</pre>					
def	<pre>forward(x,y):</pre>				
	$z = x^*y$				
	<pre>self.x = x # must keep these around!</pre>				
	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)



Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 6 - 53 Sept. 19, 2024

Example: Torch Layers

O Code 🕐 Insues 27	[] Pull requests (s) III Wiki + Pulse [a	Graphs		
io description or website p	rovided.			
@ 1,039 commits	P 7 branches	⊙ 0 releases	e e	86 contributors
Branch: master - New pa	request New Sie Find Si	HTTPS .	https://github.com/torch	Download
sourrith Merge pul reques	#565 from torch/neveri-563-master		Latest com	mit 23dd17d 15 hours
doc doc	Fix batch mode in MarginRankingCriterion			4 days
in ceneric	Improve error message in SpatialConvolution/MM			a day
in its	THNN: add missing OpenMP include			2 days
in rocks	Add Tuaffi dependency			14 days
Fl. ationore	tell pit to ipnore build output			4 months
Auscheckre	[Torch] Nova test.lua to the top level			0 100
E tanés ani	small fixes for test path			2 months
F) Abs.lue	Add THNN convenion of (ELU, LeakyReLU, LogSigmo	d LosSofther Lor	ber .	7 dava
AbsCriterion.lua	Add THNN conversion of (ELU, LeakyReLU, LogSigmo			7 days
E Addam	fix Add with multi-dim bias	o, coportent, co-		10 months
E AddConstant but	Adding in-place AddConstant and MulConstant			9 months
E) Add Constant Lus	Remove unnecessary mallocs from BCECriterion			3 months
E BCECriteron tax	Nemove unnecessary matrice from BCECriterion Sx batchroom reset			3 months 3 months
E) BatchNormalization.lua	fix batchnorm reset fixing table modules to return correct number of grading			
E) CAddTableJua	foring table modules to return correct number of grading foring table modules to return correct number of grading			6 months 6 months
E CDivitable.loa		ots		
	Add C implementation of SpatialBalchNormalization			7 days
E CMultus	nn.Module preserve type sharing semantics (#187); add			4 months
CMulTable.lua	fixing table modules to return correct number of grading	ots		6 months
CONTRIBUTING.rid	added developing tips			3 months
COPYRIGHT.bt	add copyright file			2 years
CSubTable.lua	fixing table modules to return correct number of grading	uts		6 months
Clamp.kua	Use custom range in HardTanh and mask it as Clamp			3 months
ClassNLLCriterion.lua	Add functional conversion of ClassNLLCriterion			13 days
E Concetilue	fix a bug in conditional expression			a month
ConcatTable Jaa	fixing bug in ConcatTable variable length			4 months
Container.lua	Adding :applyToModules() to nn.Container, which is like	:apply() but		3 months
E Copy.kus	nn.Module preserve type sharing semantics (#187); add	I nn.Module.apply		4 months
Cosine Jua	Fix type() in Cosine			a month
CosineDistance.lua	Do not change state variables in CosineDistance/Cosin	eEmbeddingCriterio	n	2 months
CosineEmbeddingCriterion	Do not change state variables in CosineDistance/Cosin	EmbeddingCriterio	n	2 months
E Criterion kas	nn.Module preserve type sharing semantics (#187); add	rn.Module.apply		4 months
CriterionTableJua	Rename unpack to table unpack for Lue 5.2			8 months
CrossEntropyCriterion.Jus	Check for 'nn Module' and 'nn Otterion' in recursiveTy	pe.		8 months
DepthConcat.kus	adding direct backward to Concat, DepthConcat, Sequ	ential		9 months
DistRLDivCriterion.lua	Use tensor for THNN functions even for single element:	outputs		10 days
DolProductive	Add batch mode in DotProduct + unit test			2 months
Dropout.lus	in-place dropout			4 months
R ELU.ka	Add THNN conversion of (ELU, LeakyReLU, LogSigmo	id LogSoftMax Log	łu	7 davs
ErrorMessages.lua	Give better error messages when trying to use the wron			0 100
Euclidean lus	nn Module preserve type sharing semantics (#187); add	I nn Module apply		4 months
P) Exp.lue	Exp made has priv			9 months
FlattenTable.lup	nn.Module preserve type sharing semantics (#187): add	I no Module apply		4 months
D GradientReserval has	Arid GradientReserval Inver			4 months
P HardShrink.lug	Add functional conversion of HardShrink			10 days
R HardTanhJua	Add functional conversion of HardTanh			10 days
E HingeEmbeddingCriterion Ju				6 months
E) Hingeenbedungoniending F) Identity kaa	Revert to previous identity tas implementation			2 months
E) Index.lua	Simplifying and more efficient nn.Index			2 months
e) index.lue Ri Jacobian.kus				
E) Jacobian lua F) JoinTable lua	Add unit tests for hession lue, fix bugs detected by the t			6 months
	nn Module preserve type sharing semantics (#187); ado			4 months
E L1Cost.lua	Use tensor for THNN functions even for single element	outputs		10 days
E L1HingeEmbeddingCriteria.				9 months
E L1PenaltyJua	fixed L1Penalty constructor arguments			a year
E LeakyRoLU.lua	Add THNN conversion of (ELU, LeakyReLU, LogSigmo	id, LogSoftMax, Los	ku	7 days
E LinearJua	Remove spurious mallocs from nn.Linear			4 months

oSigmoid lue	Add THNN conversion of (ELU, LeekyReLU, LogSigmoid, LogSofMax, Looku	7 days
gSofMax.lua	Add THNN convenion of (ELU, LeekyReLU, LogSigmoid, LogSofMax, Looku	7 days
okupTable Jua	Harmonize LookupTable signature with curn impl	5 days
Uua	Rename unpack to table unpack for Lua 5.2	8 months
SECriterion Jua	Add SizeAverage to criterions in the constructor	2 months
eginCriterion kus	modernized MarginCriterion	a year 4 dana
indune de la composition das	Fis batch mode in MarginRankingCriterion Morge pull request IH64 from vgiro/master	4 days 2 months
en lue	Add support for negative dimension and both batch and non batch input	2 months
nJue	Merge pull request #464 from vgire/master	2 months
itureTable.lua	cancel unused variable and useless expression	29 days
clule.lue	Revert "Don't re-flatten parameters if they are already flattened"	15 hours
Uue	removing the requirement for providing size in nn.Mul	a year
(Constant.)ua	Ignore updateGradinput if self gradinput is nil	3 months
diCriterion lua	asserts in MultCriterion and ParallelCriterion add	2 months
BLabeMarginCriterion.lua	initial revamp of torch7 tree	4 years
MarginCriterionJua	multimargin supports p+2	11 months
rrowJua rrowTable lua	typeAs in Narrow not done in place. NarrowTable	6 months 6 months
rrowrabie lua	Namov late Remove brinn and baddbrinn from Normalize, because they allocate memory,	o morens 20 days
ReLU Jua	Buffers for PReLU cude implementation.	8 months
ddino.ka	fored broken nn.Padding: inout was returned in backprop	5 months
inviseDistance.lue	Merge pull request #532 from xwgengtmaster	29 days
rallel.lua	fix a bug in conditional expression	a month
rallelCriterion.lua	asserts in MultiCriterion and ParallelCriterion add	2 months
ralle/Table.lua	Parallel optimization. ParallelTable inherits Container. unit tests	a year
wer.kas	Use UNIX line endings	7 months
SADVE rid	doc readthedocs	5 months
ReLU.lus	Add randomized leaky rectified linear unit (RReLU)	3 months
LUise	adds in-place ReLU and foxes a potential divide-by-zero in nn.Sqrt	9 months
plicate Jua nhape Jua	Replicate batchNode Added more informative pretty-printing.	8 months
intepe.ua lect.lua	Added more instrume precy-priming. initial revenue of torch7 tree	a year 4 years
lectTable lua	nn. Module preserve type sharing semantics (#187); add nn. Module apply	4 years 4 months
overfiel.lue	foing Sequential remove comer case	6 months
subion	initial revemp of torch7 tree	4 years
sooth, 1Criterion.kas	Add SizeAverage to criterions in the constructor	2 months
Max ka	Fix various unused variables in m	a year
656n Jua	Fix various unused variables in m	a year
APlus kas	fixed a numerical issue in the SofPlus module (it breaks for input g	2 years
f/Shrink.lua	initial revamp of torch7 tree	4 years
#Sign.lus	initial revemp of torch? tree	4 years
arseJacobian kua arseLinear kua	Fix various unused variables in nn Using sparse implementation of zeroGradParameters for SparseUnear	a year a month
arseurear.ua atalAdastiveMasPoolins	Added BoatialAdaptiveMaxPoping	a nonn
atal/weragePooling.lua	Spatial/waragePooling supports padding, cel mode and exclude_pad div	29 days
etalBatchNormalization lua	Add C implementation of SpatialBatchNormalization	7 days
	Make type() truly recursive.	9 months
etalConvolution.lua	Fix type() in SpatialConvolution	3 months
atialConvolutionMM.lua	Fix type() in SpatialConvolution	3 months
atalConvolutionMap.kua	Remove unused and expensive initialization logic from nn.SpatialConvo	8 months
etalCrossMapLRN Aus	cuda consistency	18 days
eta/DivisiveNormalizatio	Spatial(Constructive,Divisive,Subtractive)Normalization work with bat	8 months
etalDropout lua	small fix on error message	6 months
ata/Fractiona/MaxPoolin ata/FullConvolution.lua	Adding Fractional Max Pooling Add adjustment term to SpatialFullConvolution to control the size of	3 months 5 days
		0 days 3 years
ataLPPoolingJua	SpatialAveragePooling divides by kW*kH	10 months
etaMaxPooling.tua	SpatiaMaxPooling supports padding and cell mode	6 months
atiaMaxUrpooling Jua	Add SpatialMaxUnpcoling	26 days
afa/SotMaxJua	Update SoftMax to work in spatial mode	4 months
eta/SubSempling ka	Merge branch 'nn_fast_reset'	3 years
atia/SubtractiveNormaliz	Spatial[Constrastive,Divisive,Bubtractive]Normalization work with bat	8 months
etalUpSamplingNearest.L		7 months
ata/ZeroPadding.lua	Added more informative pretty-printing.	a year
RTable lua	Add support for negative indices in nn.SplitTable	7 months

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 6 - 54 Sept. 19, 2024

Example: Torch Layers

@ 1,039 commits	P 7 branche			O relea	045	(i) 46 con	tributors
Branch: master - New pull	Inquest	New Sie	Find file	HTTPS -	https://githut		Download Z
	#565 from torch/nevert-563-master	(international				Latest commit 23d	
dos	Fix batch mode in MarginRanki						4 dava ac
in ceneric	Improve error message in Spati		101				a day ac
	THNN add mission OpenMP in						2 days ac
accia.	Add 'luaff' dependency	0000					14 days ac
P. ationore	tell pit to ignore build output						4 months ag
R karberke	Torchi Move test jue to the top	level					0 1997 85
E tanés ami	small fixes for test path						2 months ac
F) Abs.lue	Add THNN conversion of (ELU,	LeskyReLU.	LosSignoid L	ozSoftMax.	Looku		7 days as
E AbsCriterion Inc.	Add THNN conversion of (ELU,						7 days ag
E Add lua	fix Add with multi-dim bias						10 months ad
AddConstant.kas	Adding in-place AddConstant at	d MulConsta	nt				9 months an
R RCECillation Int	Remove unnecessary mallocs f						3 months ag
BatchNormalization.lua	fix batchnorm reset						3 months ac
E) CAddTable.lus	foring table modules to return or	mect number	of and inputs				6 months ag
F) CDivTable.lue	fixing table modules to return or						6 months ad
E CMakeLists.txt	Add C implementation of Spetia						7 days ac
E CMulture	nn Module preserve type sharin			Module appl	v.		4 months ap
CMulTable lus	foring table modules to return or						6 months ag
CONTRIBUTING and	added developing tips						3 months ac
E COPYRIGHT M	add copyright file						2 years ac
F) CSubTable Jua	fixing table modules to return or	rect rumber	of and inputs				6 months ag
Clamp.kus	Use custom range in HardTanh	and mask it a	is Clamo				3 months ag
ClassNLLCriterion.lua	Add functional conversion of Clu						13 days ac
Concetiue	fix a bup in conditional expressi	on					a month ag
ConcatTable Jua	foring bug in ConcatTable variab	de length					4 months ag
E Container lua	Adding :applyToModules() to m		hich is like tap	phyl but			3 months ad
Copy.kus	nn.Module preserve type sharin				v.		4 months ap
R) Cosine Jua	Fix type() in Cosine						a month ag
CosineDistance.lua	Do not change state variables in	CosineDista	nce/CosineEn	beddingCrite	arion		2 months ac
CosineEmbeddingCriterion I.							2 months ad
Criterion kas	nn.Module preserve type sharin	g semantics (#187); add nn	Module appl	Y		4 months ag
R CriterionTable Jue	Rename unpack to table unpact						8 months ad
CrossEntropyCriterion Jua	Check for 'nn Module' and 'nn		cursiveType.				8 months ad
DepthConcet.tue	adding direct backward to Con	at, DepthCor	cat, Sequenti	4			9 months ag
DistRLDivCriterion.lua	Use tensor for THNN functions	even for single	element out;	uts			10 days ag
DolProductive	Add batch mode in DotProduct						2 months ad
E Dropout lus	in-place dropout						4 months ag
ELU.ka	Add THNN conversion of (ELU,	LeakyReLU.	LogSigmoid, L	ogSoftMax, I	Looku		7 days ag
ErrorMossages.lua	Give better error messages whe	in trying to us	e the wrong ki	nd of Tensor			a year ag
Euclidean.tus	nn Module preserve type sharin	g semantics (#187); add nn	Module appl	v		4 months ap
Explus	Exp made lue only						9 months ag
FlattenTable.lup	nn.Module preserve type sharin	g semantics (#187): add rn	Module appl	ý.		4 months ag
CadectReversal lus	Add GradientReversal layer						4 months as
E) HardShrink.lua	Add functional convension of Ha	rdShrink					10 days ap
HardTanh.lua	Add functional conversion of Ha	rdTanh					10 days a
E) HingeEmbeddingCriterion kus	rewrote HingeEmbeddingCriteri	on to support	batch mode				6 months as
E Identity.kus	Revert to previous Identity lua in	plementation					2 months ag
E Index.lue	Simplifying and more efficient n						2 months ag
E Jacobian Jua	Add unit tests for hession lue, fi	x bugs detect	ed by the tests				6 months a
JoinTable.tua	nn Module preserve type sharin				¥.		4 months as
E L1Cost.lua	Use tensor for THNN functions						10 days ag
E L1HingeEmbeddingCriteria							9 months ay
E L1PenabyJua	fixed L1Penalty constructor arg	arrents					a year a
R LeakyRoLU.toa	Add THNN conversion of (ELU.		LooSigmoid L	.ogSoftMax, I	Looku		7 dava ar
R Lineer Jua	Remove spurious mallocs from		- any sold by				4 months ac

LogSigmoid lug	Add THNN conversion of (ELU, LeekyReLU, LogSigmoid, LogSofMax, Looku	7 day
LogSofMax kas	Add THNN conversion of (ELU, LeekyReLU, LogSigmoid, LogSofMax, Looku	7 day
LookupTable Jua	Harmonize LookupTable signature with curn impl	5 day
Milua	Rename unpack to table unpack for Lua 5.2	8 month
MSECriterion lua	Add SizeAverage to criterions in the constructor	2 month
MarginCriterion Jua	modernized MarginCriterion	a yea
MarginRankingCriterion Aus	Fix batch mode in MarginRankingCriterion	4 day
Maxiua	Merge pull request #464 from vgire/master	2 month
Mean Jua Min Jua	Add support for negative dimension and both batch and non batch input	2 month
	Merge pull request #464 from vgire/master	2 month
MatureTable Jua Nodule Jua	cancel unused variable and useless expression Revert "Don't re-flaties currameters if they are already flattered"	29 day
Module.lus Multue	Revert 'Don't re-flatten parameters if they are already flattened' removing the requirement for providing size in nn.Mul	
Nullionatant.lua	removing the requirement for providing size in init.Mul Ignore updateGradinput If self.gradinput is nil	a yes
MultiCriterion.lua	aprore updatecursomput in ser greatingut is nil asserts in MultiCriterion and ParallelCriterion add	2 month
MultiLabeMarginCriterion.lua	initial revenue of torch? tree	4 year
NutlMarginCriterionJua	multimargin supports p=2	11 month
Narrow Jua	topeAs in Narrow not done in place.	6 month
NarrowTable lua	Narrow Table	6 month
Normalize Jua	Remove brum and beddbrum from Normalize, because they allocate memory,	20 day
PRoLUJua	Buffers for PReLU cude implementation.	8 month
Padding.ka	fixed broken nn.Padding: input was returned in backprop	5 month
PairwiseDistance.lua	Merge pull request #532 from xwpeng/master	29 day
Parallel ka	fix a bug in conditional expression	a mont
ParallelCriterion.lua	asserts in MultiCriterion and ParallelCriterion add	2 month
Paralle/Table.lua	Parallel optimization. ParallelTable inherits Container. unit tests	a yes
PowerJus	Use UNX line endings	7 month
READ/VE.md	doc wadthedocs	5 month
RReLUlus	Add randomized leaky rectified linear unit (RReLU)	3 month
ReLUlus	adds in-place ReLU and foxes a potential divide-by-zero in nn.Sqrt	9 month
Replicate Jua	Replicate batchMode	8 month
Reshape.lus	Added more informative pretty-printing.	a yes
Select kas	initial revemp of torch7 tree	4 year
SelectTable.lua	nn.Module preserve type sharing semantics (#187): add nn.Module apply	4 month
Sequential lua	fixing Sequential remove corner case	6 month
SigmoidJua	initial revemp of terch7 tree	4 year
SmoothL1Criterion.kus	Add SizeAverage to criterions in the constructor	2 month
SofMax lua	Fix various unused variables in m	a yea
SoftMin.lua	Fix various unused variables in m	a yes
SofPlus kas	fixed a numerical issue in the SofPlus module (it breaks for input g	2 year
SoftShrink.lua	initial revenp of torch? tree	4 year
SoftSign.lue	initial reversp of torch? tree	4 year
Sparse Jacobian Ika	Fix various unused variables in nn	a yea
SparseLinear Aua	Using sparse implementation of zeroGradParameters for SparseLinear	a mont
SpatialAdaptiveMastPooling		a yea
Spatial/weragePooling.lua	SpatialAveragePooling supports padding, ceil mode and exclude_pad div Add C implementation of SpatialilatchNormalization	29 day
		7 day 9 month
SpatialContrastiveNormaliz SpatialConvolution.lua	Make type() kuly secursive. Fix type() in SpatialConvolution	9 month 3 month
SpatialConvolution.Ua	Fit type() in SpatialConvolution	3 month
Spatial Convolution Map Jua	Pik type() in Spatial-Convolution Remove unused and expensive initialization logic from m.SpatialConvo	8 month
SpatialCrossMapLRN Jus	cuda consistency	18 day
Spatia/DivisiveNormalizatio	Spatial(Constructive,Divisive,Subtractive)Normalization work with bat	8 month
Speta/Dropout.lue	small fix on error message	6 month
SpatialFractionalMaxPoolin	Adding Fractional Max Pooling	3 month
SpatialFullConvolution Jua	Add adjustment term to SpatialFullConvolution to control the size of	5 day
SpetialFullConvolutionWep lue		3 year
SpatialLPPcoling.tua	SpatialAveragePooling divides by KW1KH	10 month
Spatial Max Pooling Iua	SpatialMaxPooling supports padding and cell mode	6 month
Spatia/MaxUrpcoling.lua	Add SpatialManUnpooling	26 day
SpatialSoftMaxJua	Update SoftMax to work in spatial mode	4 month
SpetalSubSempling ke	Merge branch 'nn_fast_reset'	3 year
SpatialSubtractiveNormaliz	Spatial(Constrastive,Divisive,Subtractive(Normalization work with bat	8 month
SpatialUpSamplingNearest.L.	Use UNIX line endings	7 month
SpatialZeroPadding lua	Added more informative pretty-printing.	a yea
SpiRTable lua	Add support for negative indices in nn. SplitTable	7 month



Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

Lecture 6 - 55 Sept. 19, 2024

```
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')
  end
end
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)
  end
 return self.output
end
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)
   end
   return self.gradInput
  end
```

Example: Torch MulConstant

Lecture 6 - 56

Sept. 19, 2024

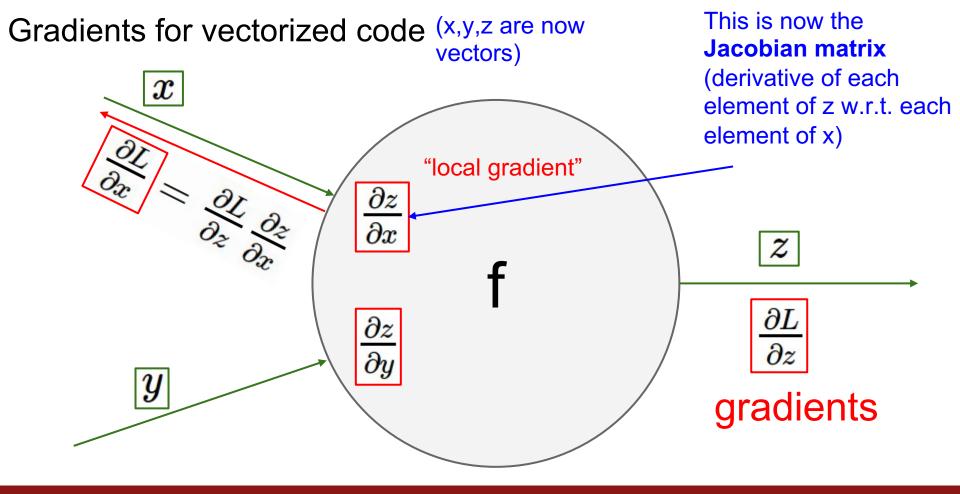
$$f(X) = aX$$

initialization

_forward()

backward()

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller



Lecture 6 - 57 Sept. 19, 2024

[slides] [backprop notes] [Efficient BackProp] (optional) related: [1], [2], [3] (optional)

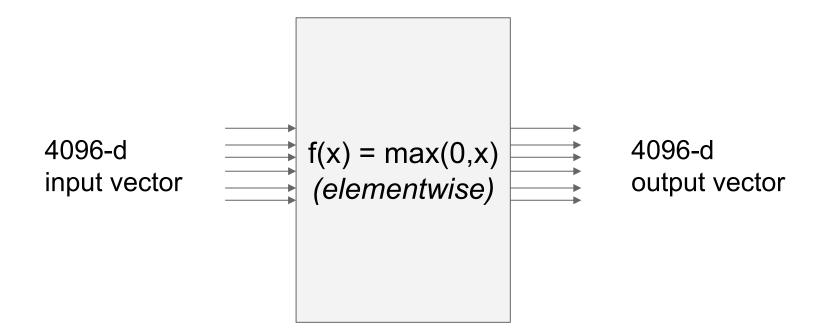
[slides]

handout 1: Vector, Matrix, and Tensor Derivatives handout 2: Derivatives, Backpropagation, and Vectorization Deep Learning [Nature] (optional)

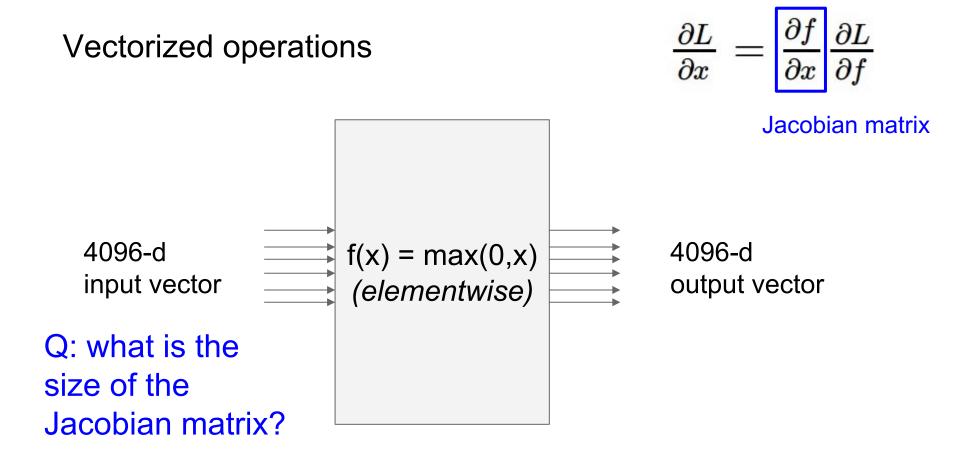
[slides] tips/tricks: [1], [2] (optional)

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller Lecture 6 - 58 Sept. 19, 2024

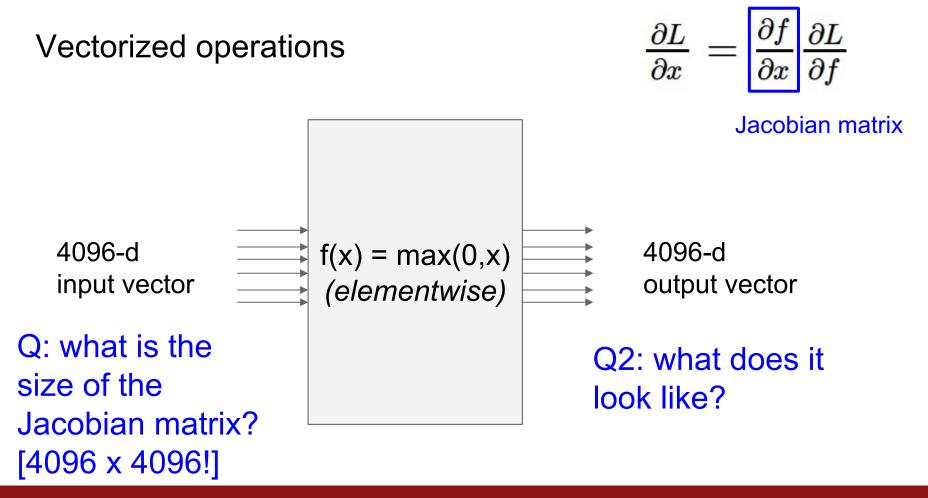
Vectorized operations



Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller Lecture 6 - 59 Sept. 19, 2024

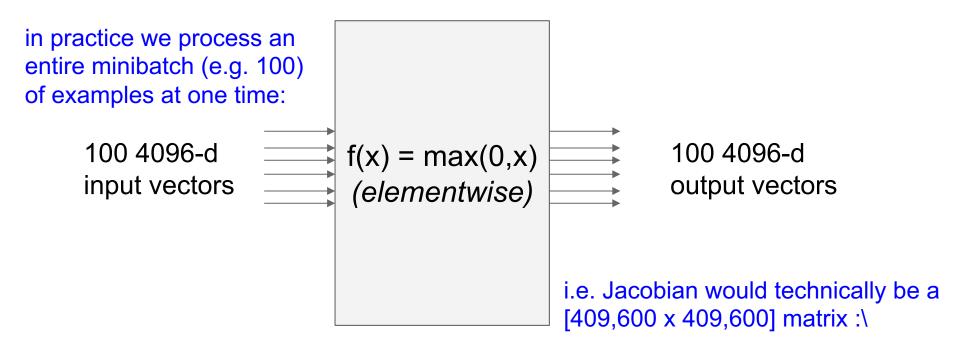


Lecture 6 - 60 Sept. 19, 2024



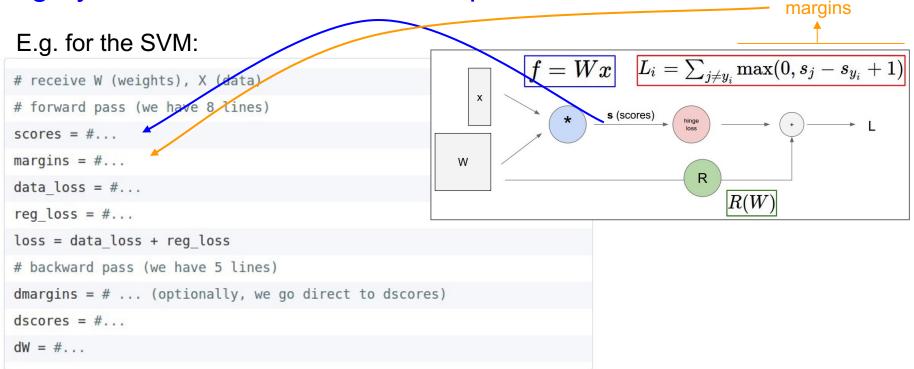
Lecture 6 - 61 Sept. 19, 2024

Vectorized operations



Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller Lecture 6 - 62 Sept. 19, 2024

Assignment: Writing SVM/Softmax Stage your forward/backward computation!



Lecture 6 - 63

Sept. 19, 2024

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller

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.

Lecture 6 - 64 Sept. 19, 2024

- 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.

Subhransu Maji, Chuang Gan and TAs Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller