Announcements

Reminder: self-grading forms for ps1 and ps2 due 10/5 at midnight (Boston)

Assignment Project Exam Help

- ps3 out today, due 10/8 (1 week)

https://powcoder.com

- Midterm practice questions out next week

Add WeChat powcoder

Today: Outline

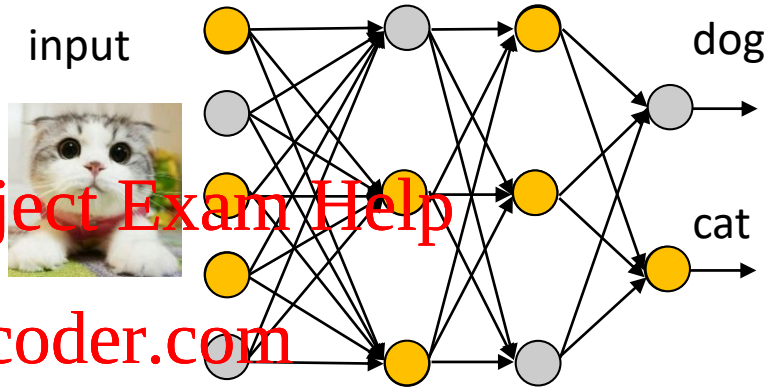
- **Neural networks cont'd:** learning via gradient descent; chain rule review; gradient computation using the backpropropagation algorithm

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Neural Networks II

Learning

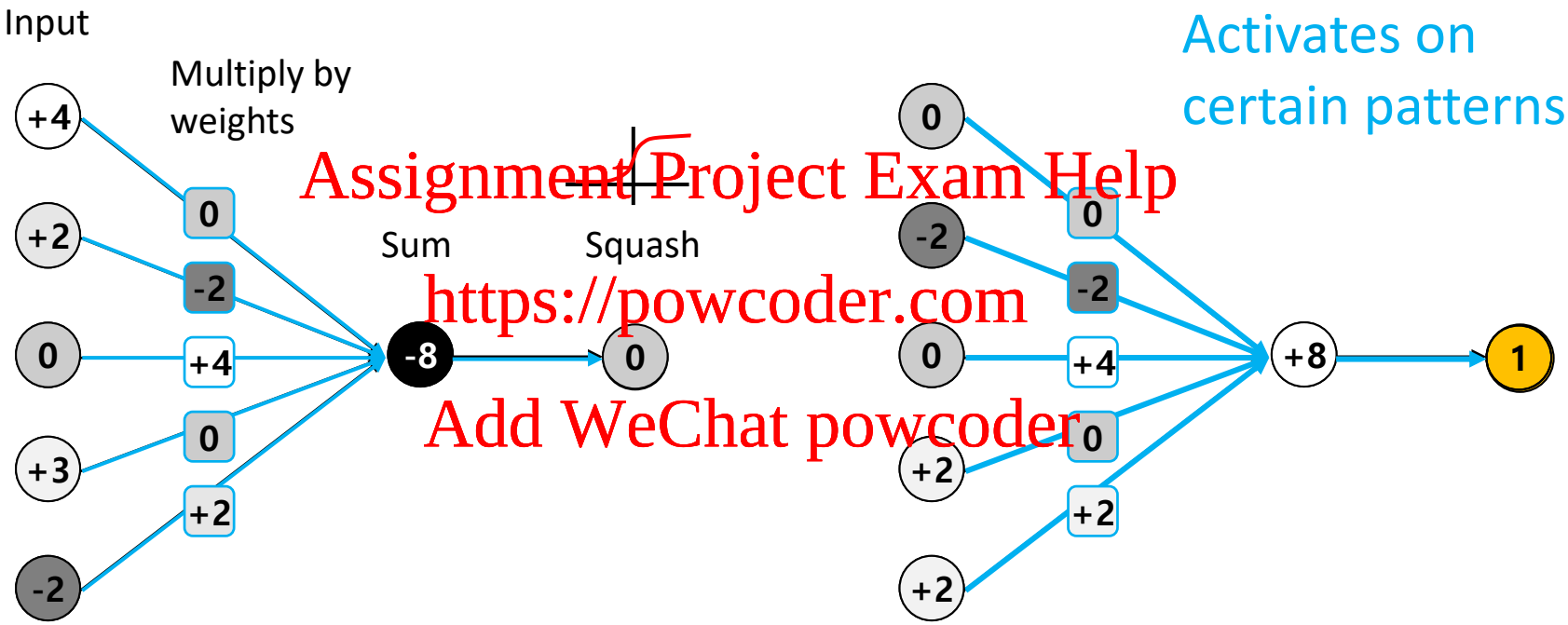
Artificial Neural Network



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

- **Artificial neural networks:** consist of many inter-connected neurons organized in layers
- **Neurons**: each neuron receives inputs from neurons in previous layer, passes its output to next layer
- **Activation**: neuron's output between 1 (excited) and 0 (not excited)

Artificial Neuron: Activation

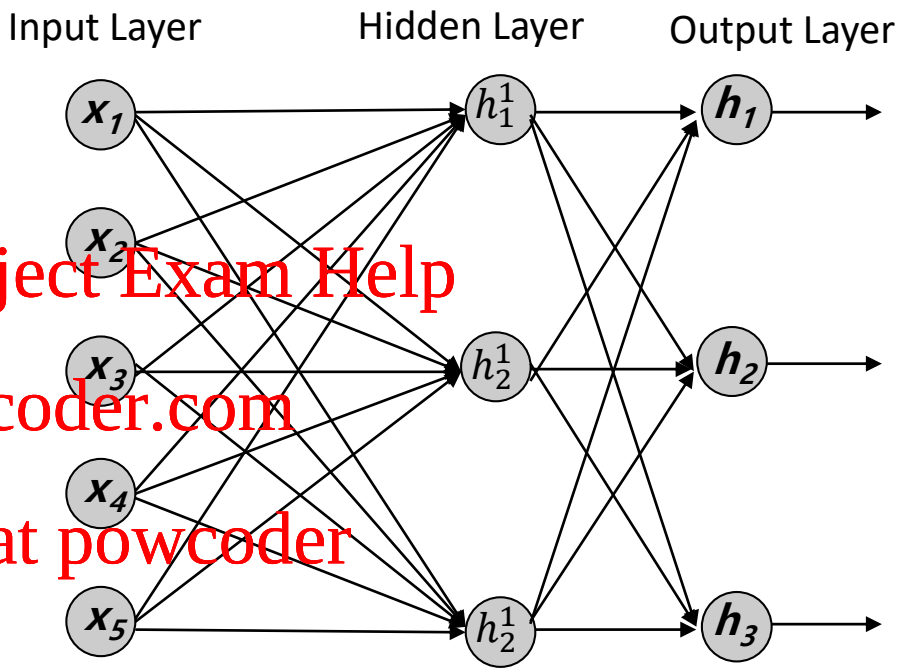


Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Artificial Neural Network: notation

input $x = \begin{bmatrix} x_1 \\ \dots \\ x_5 \end{bmatrix}$

hidden layer activations



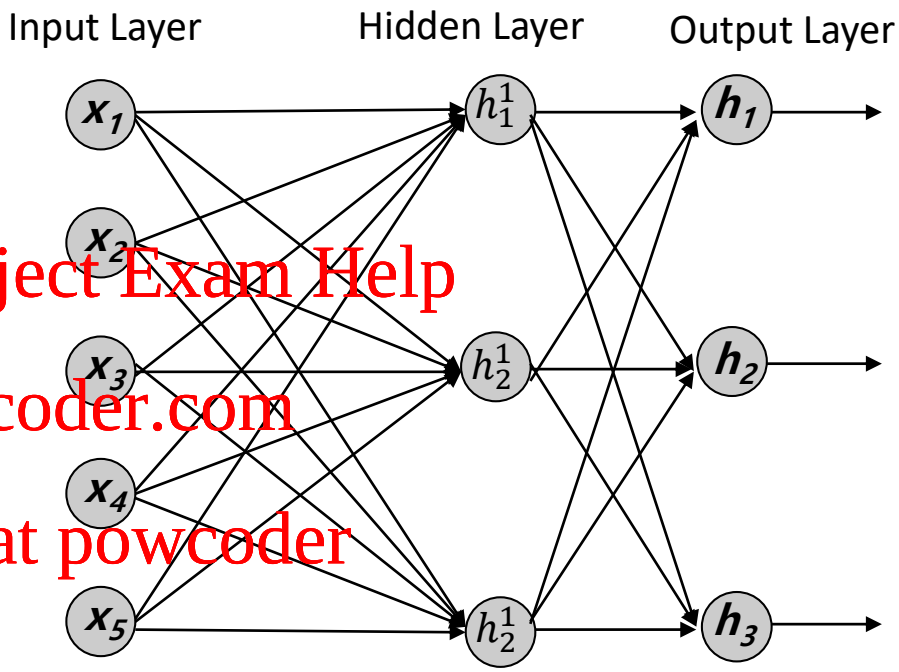
Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Artificial Neural Network: notation

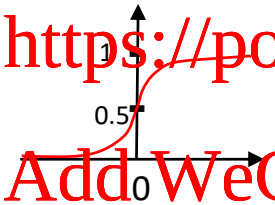
input $x = \begin{bmatrix} x_1 \\ \dots \\ x_5 \end{bmatrix}$



hidden layer activations

$$h^i = g(\Theta^{(i)} x)$$

$$g(z) = \frac{1}{1 + \exp(-z)}$$



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

output

$$h_\Theta(\mathrm{x}) = g(\Theta^{(2)} h^i)$$

weights $\Theta^{(1)} = \begin{pmatrix} \theta_{11} & \cdots & \theta_{15} \\ \vdots & \ddots & \vdots \\ \theta_{31} & \cdots & \theta_{35} \end{pmatrix}$ $\Theta^{(2)} = \begin{pmatrix} \theta_{11} & \cdots & \theta_{13} \\ \vdots & \ddots & \vdots \\ \theta_{31} & \cdots & \theta_{33} \end{pmatrix}$

Cost function

Neural network: $h_\Theta(x) \in \mathbb{R}^K \quad (h_\Theta(x))_i = i^{th}$ output

Assignment Project Exam Help

training error

$$J(\Theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{k=1}^{K} y_k^{(i)} \log(h_\Theta(x^{(i)}))_k + (1 - y_k^{(i)}) \log(1 - (h_\Theta(x^{(i)}))_k)\right]$$

https://powcoder.com

Add WeChat powcoder

$$+\frac{\lambda}{2m}\sum_{l=1}^{L-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}(\Theta_{ji}^{(l)})^2$$

regularization

Gradient computation

$$J(\Theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{k=1}^{K} y_k^{(i)} \log h_\theta(x^{(i)})_k + (1 - y_k^{(i)}) \log(1 - h_\theta(x^{(i)})_k)\right]$$

$$+\frac{\lambda}{2m}\sum_{l=1}^{L-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}} (\Theta_{ji}^{(l)})^2$$

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

$$\min_{\Theta} J(\Theta)$$

Need code to compute:

- $J(\Theta)$
- $\frac{\partial}{\partial \Theta_{ij}^{(l)}} J(\Theta)$

Use "Backpropagation algorithm"

- Efficient way to compute $\frac{\partial}{\partial \Theta_{ij}^{(l)}} J(\Theta)$
- Computes gradient incrementally by "propagating" backwards through the network

Neural Networks II

backpropagation

Chain Rule

- Need to compute gradient of

$$\log(h_\Theta(\mathrm{x})) = \log(g(\Theta^{(2)} g(\Theta^{(1)} x))) \quad \text{w.r.t}\Theta$$

Assignment Project Exam Help

- How can we compute the gradient of several chained functions?

https://powcoder.com

$$f(\theta) = f_1(f_2(\theta)) \qquad f'(\theta) = f_1'(f_2(\theta)) * f_2'(\theta)$$

Add WeChat powcoder

$$f(\theta) = f_1(f_2(f_3(\theta))) \qquad f'(\theta) =$$

- What about functions of multiple variables?

$$f(\theta_1, \theta_2) = f_1(f_2(\theta_1, \theta_2)) \qquad \frac{\partial f}{\partial \theta_1} = \qquad \frac{\partial f}{\partial \theta_2} =$$

https://www.khanacademy.org/math/multivariable-calculus/multivariable-derivatives/multivariable-chain-rule/v/multivariable-chain-rule

Backpropagation: Efficient Chain Rule

- Partial gradient computation via chain rule:

$$\frac{\partial f}{\partial \theta_1} = \frac{\partial f_1}{\partial f_2}\left(f_2\left(f_3(\theta)\right)\right) * \frac{\partial f_2}{\partial f_3}\left(f_3(\theta)\right) * \frac{\partial f_3}{\partial \theta_1}(\theta)$$

$$\frac{\partial f}{\partial \theta_2} = \frac{\partial f_1}{\partial f_2}\left(f_2\left(f_3(\theta)\right)\right) * \frac{\partial f_2}{\partial f_3}\left(f_3(\theta)\right) * \frac{\partial f_3}{\partial \theta_2}(\theta)$$

$$\frac{\partial f}{\partial \theta_3} = \frac{\partial f_1}{\partial f_2}\left(f_2\left(f_3(\theta)\right)\right) * \frac{\partial f_2}{\partial f_3}\left(f_3(\theta)\right) * \frac{\partial f_3}{\partial \theta_3}(\theta)$$

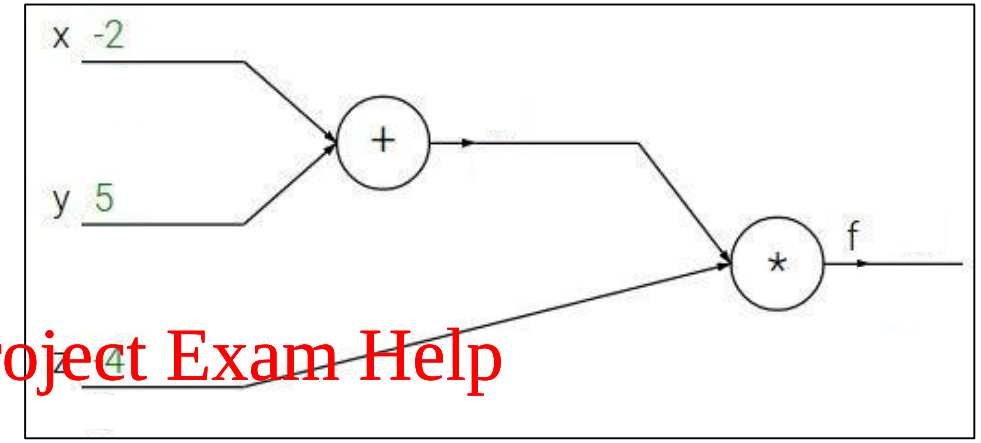
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

- need to re-evaluate functions many times
- Very inefficient! E.g. 100,000-dim parameters

Chain Rule with a Computational Graph

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

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



Assignment Project Exam Help

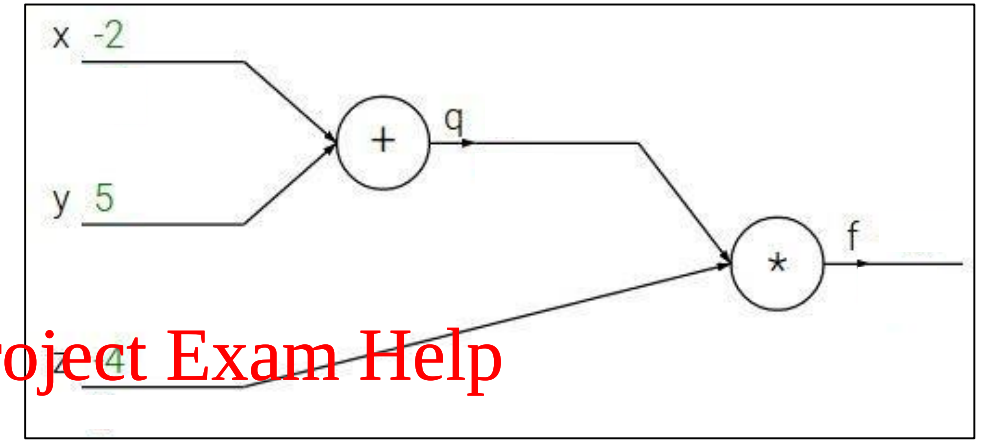
https://powcoder.com

Add WeChat powcoder

Chain Rule with a Computational Graph

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

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



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

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

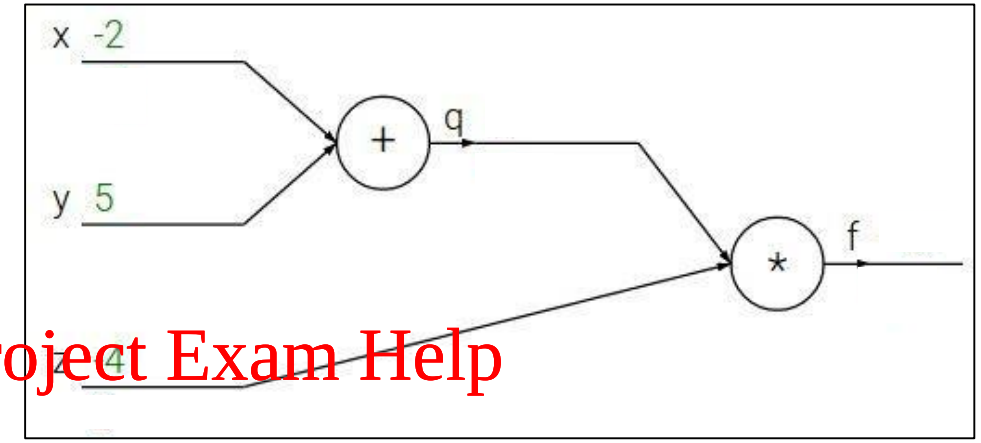
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

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

Computation Graph: Forward

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

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



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

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

Forward pass: compute values

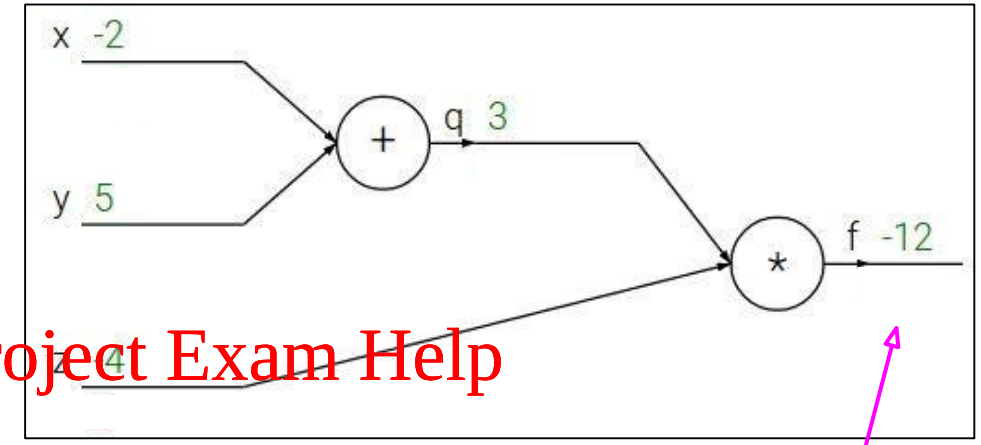
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

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

Computation Graph: Backward

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

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



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

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

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

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

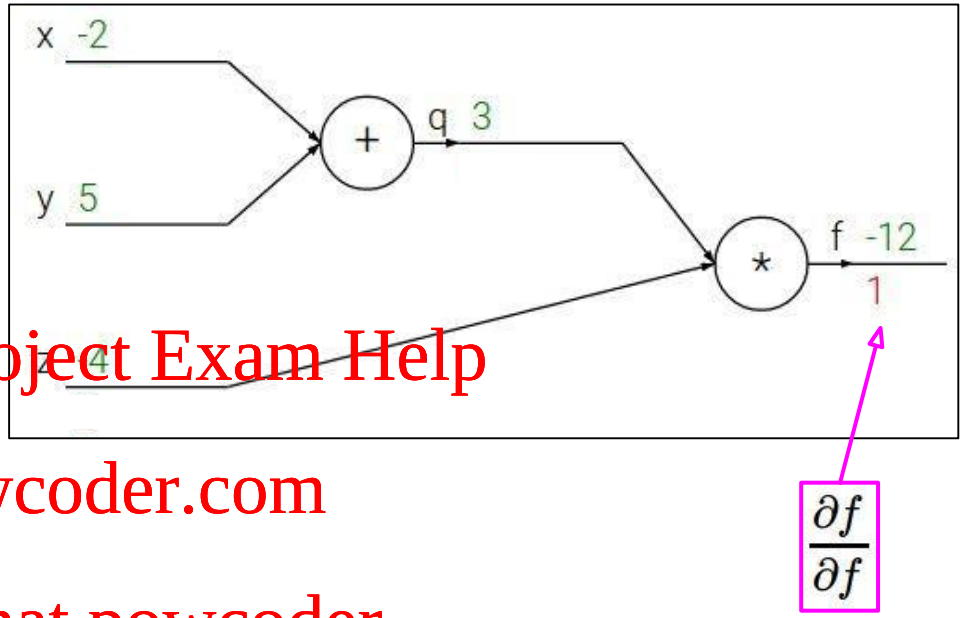
compute gradients

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

Computation Graph: Backward

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

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



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

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

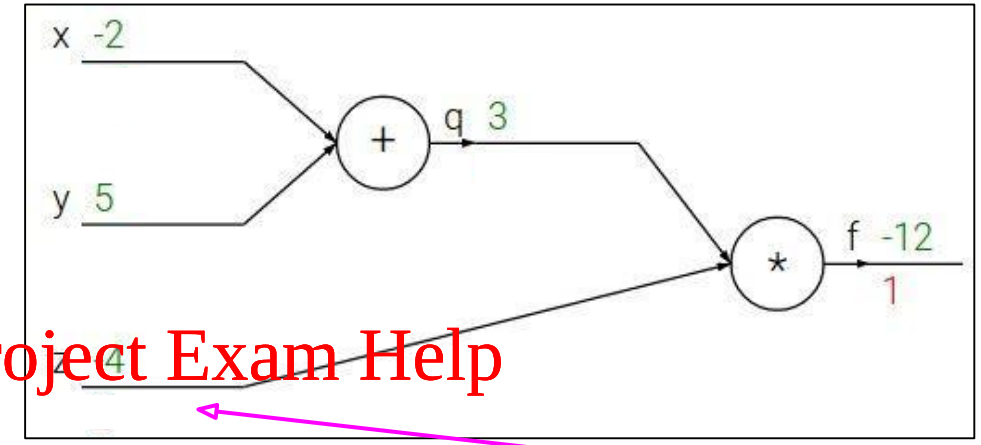
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

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

Computation Graph: Backward

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

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



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

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

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

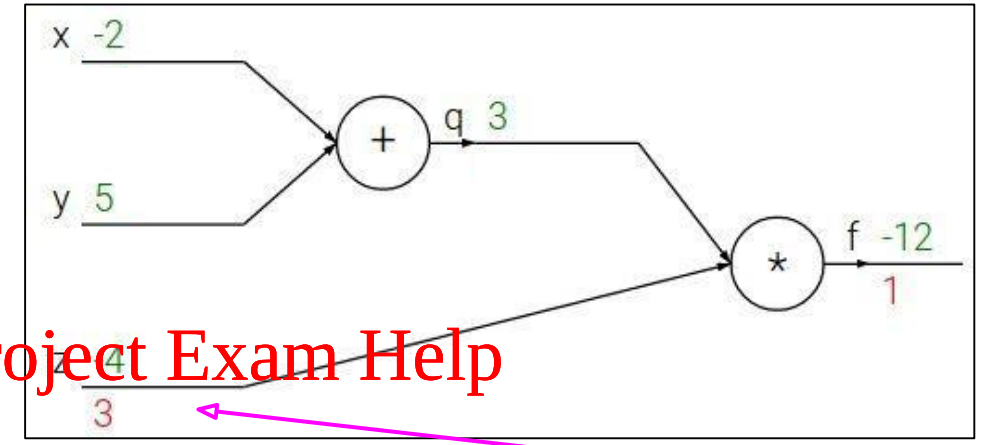
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

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

Computation Graph: Backward

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

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



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

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

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

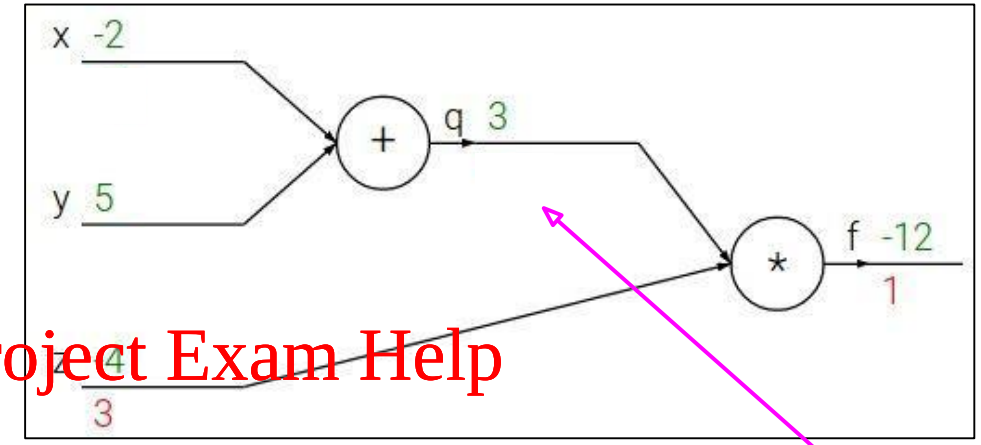
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

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

Computation Graph: Backward

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

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



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

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

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

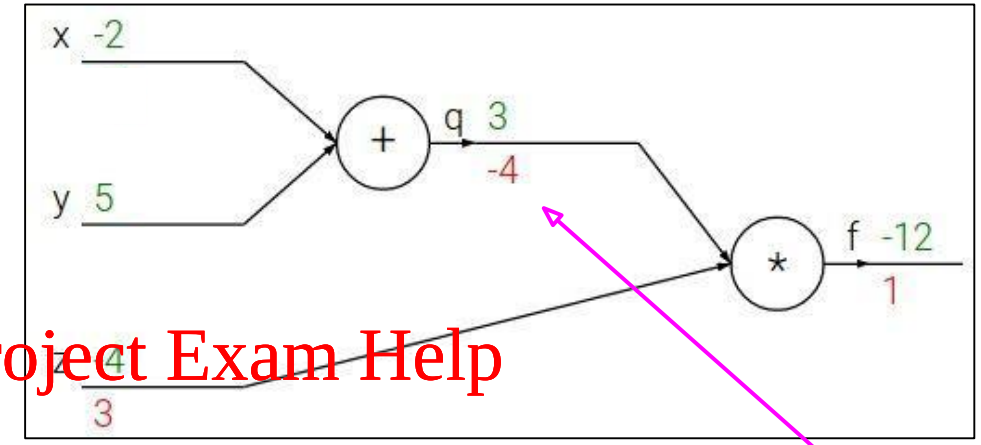
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

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

Computation Graph: Backward

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

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



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

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

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

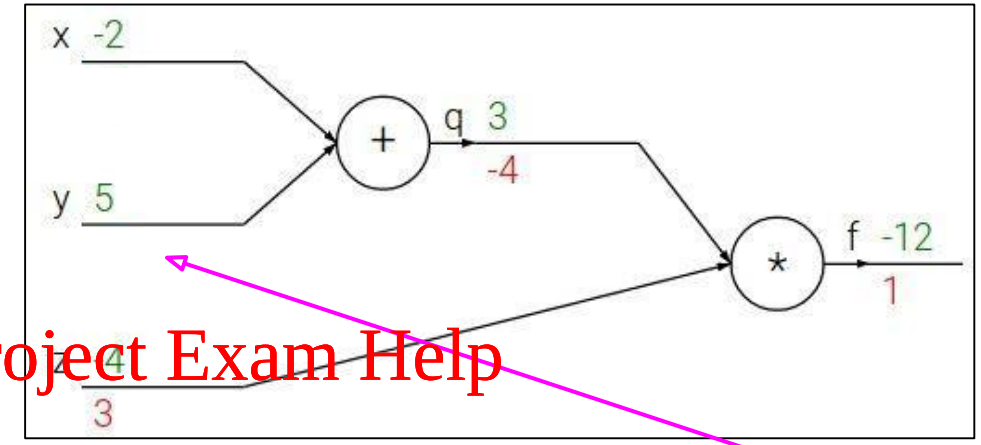
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

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

Computation Graph: Backward

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

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



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

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

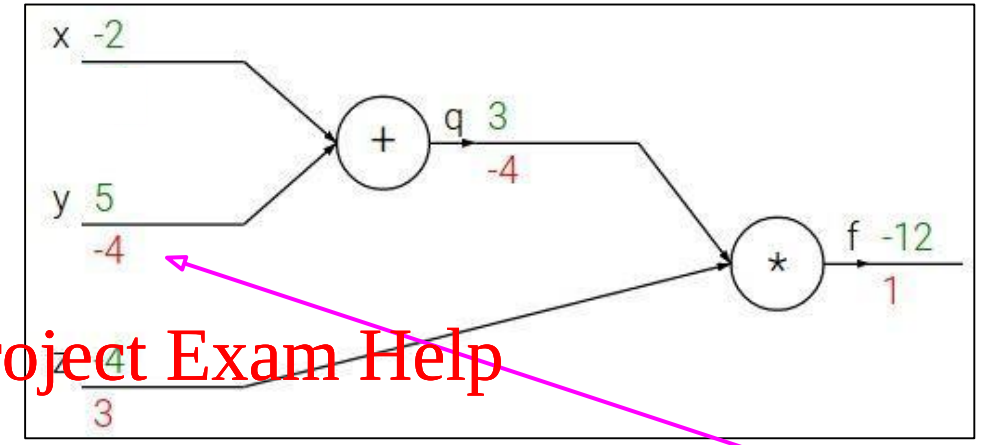
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

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

Computation Graph: Backward

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

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



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

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

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Chain rule:

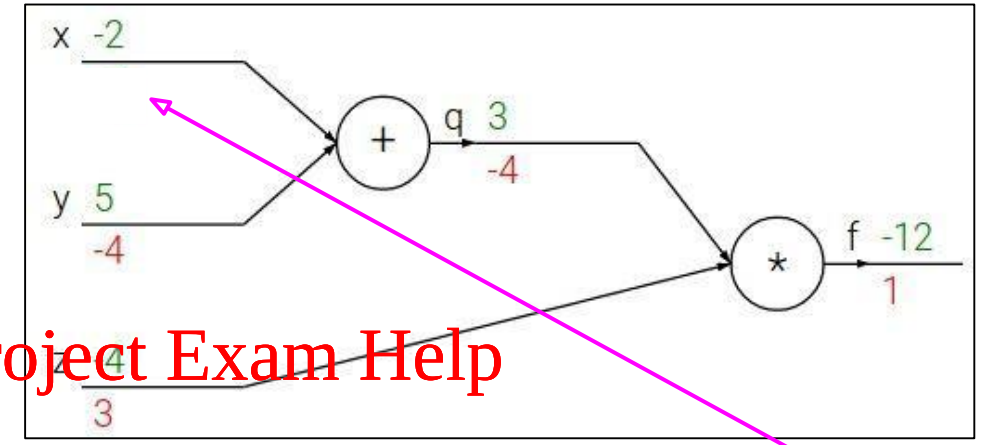
$$\frac{\partial f}{\partial y} = \frac{\partial f}{\partial q}\frac{\partial q}{\partial y}$$

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

Computation Graph: Backward

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

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



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

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

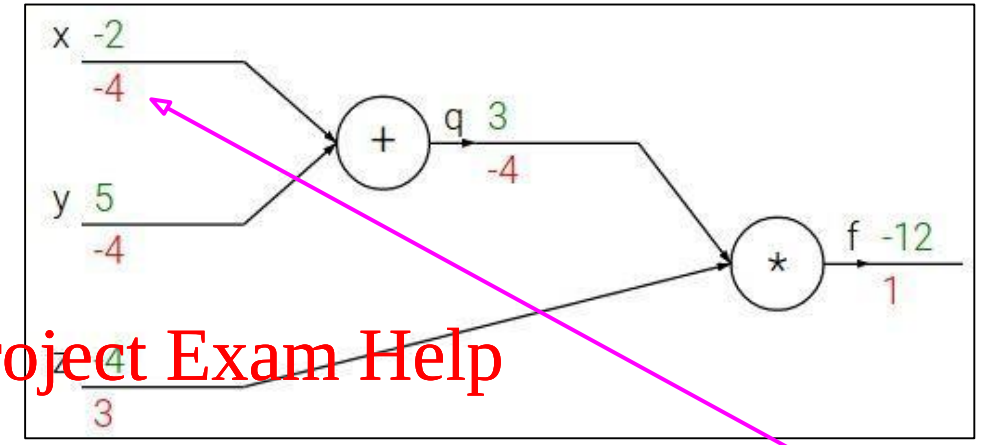
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

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

Computation Graph: Backward

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

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



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

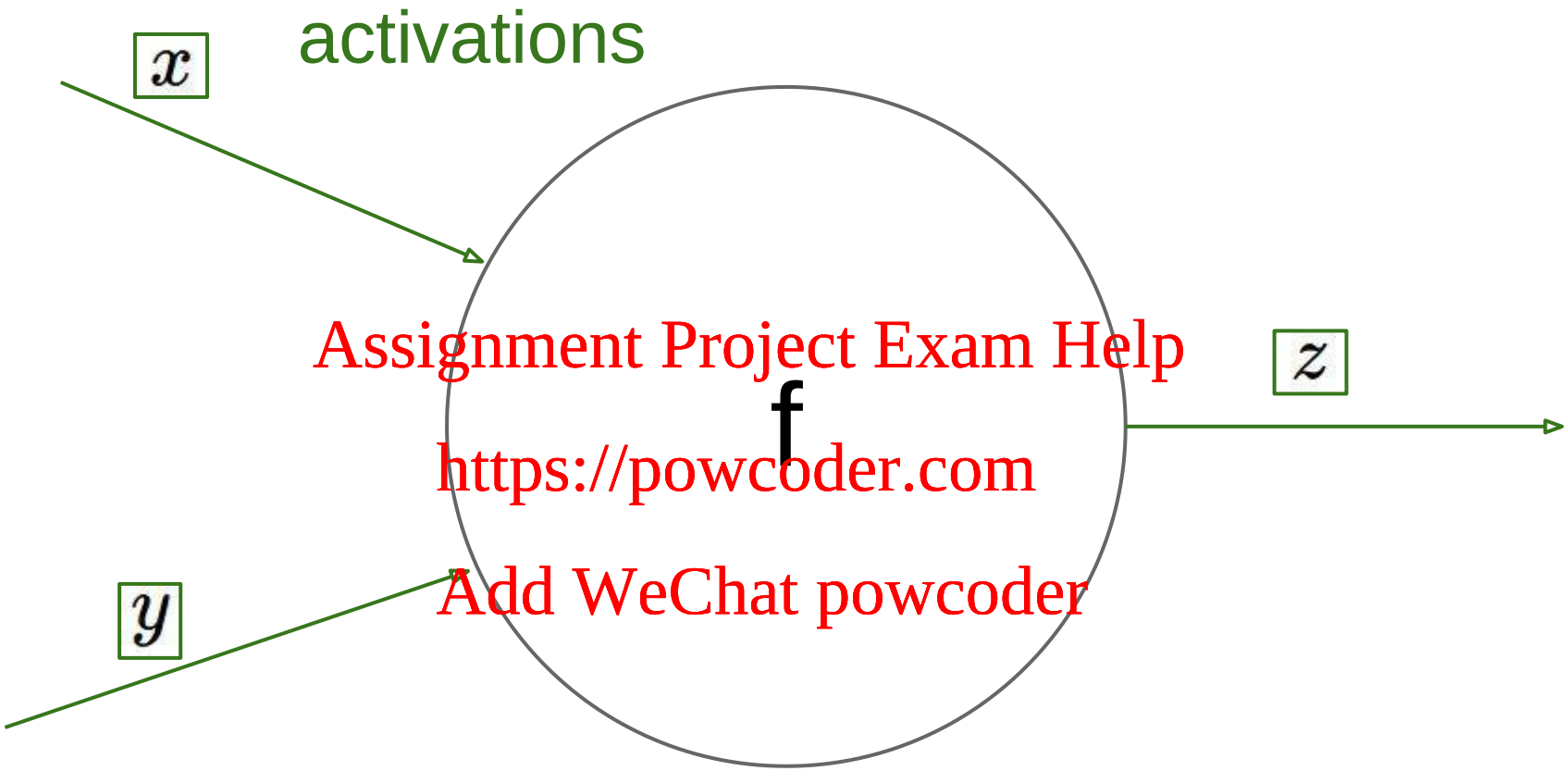
$$f = qz \qquad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

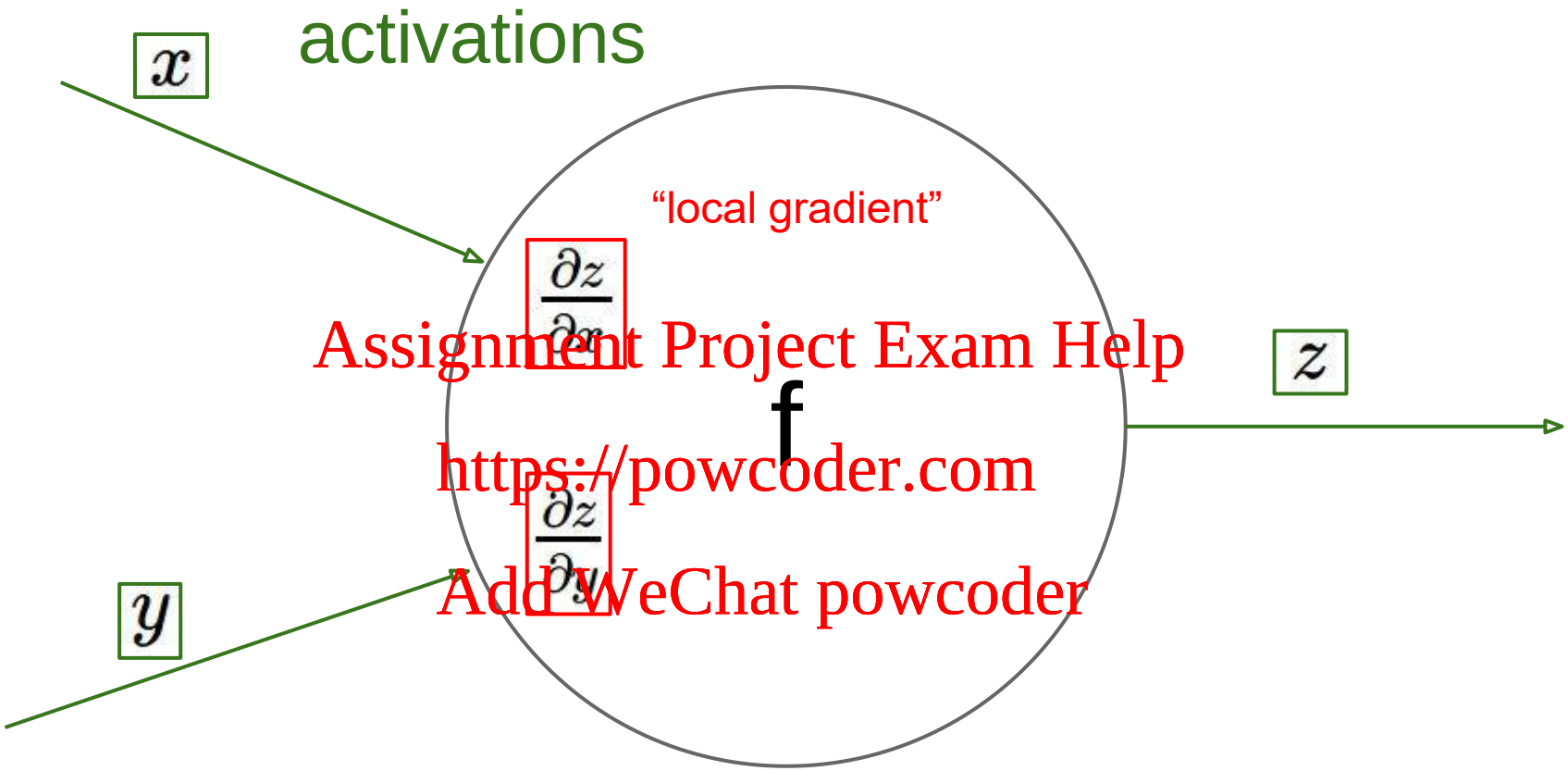
Chain rule:

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

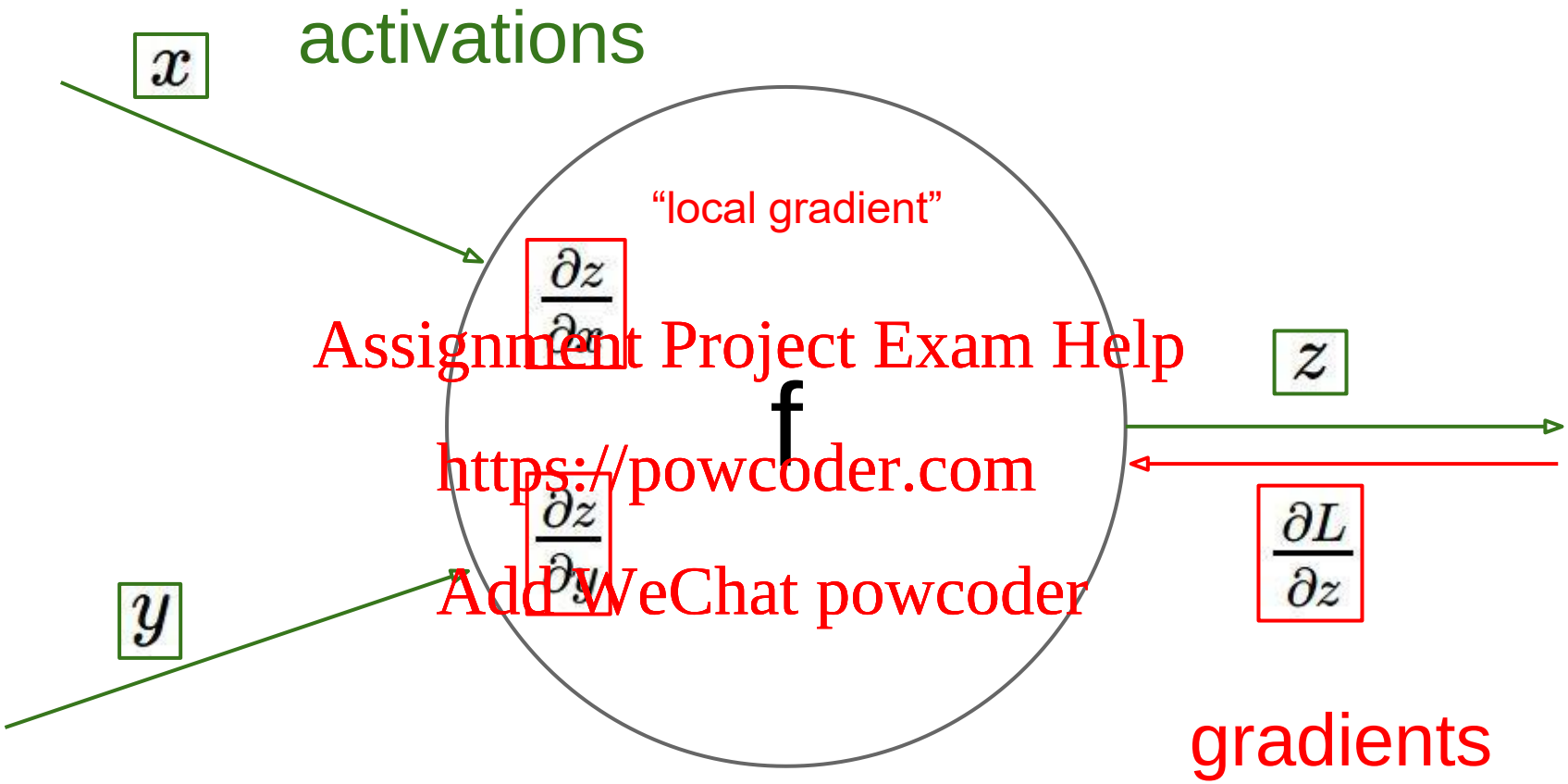
Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



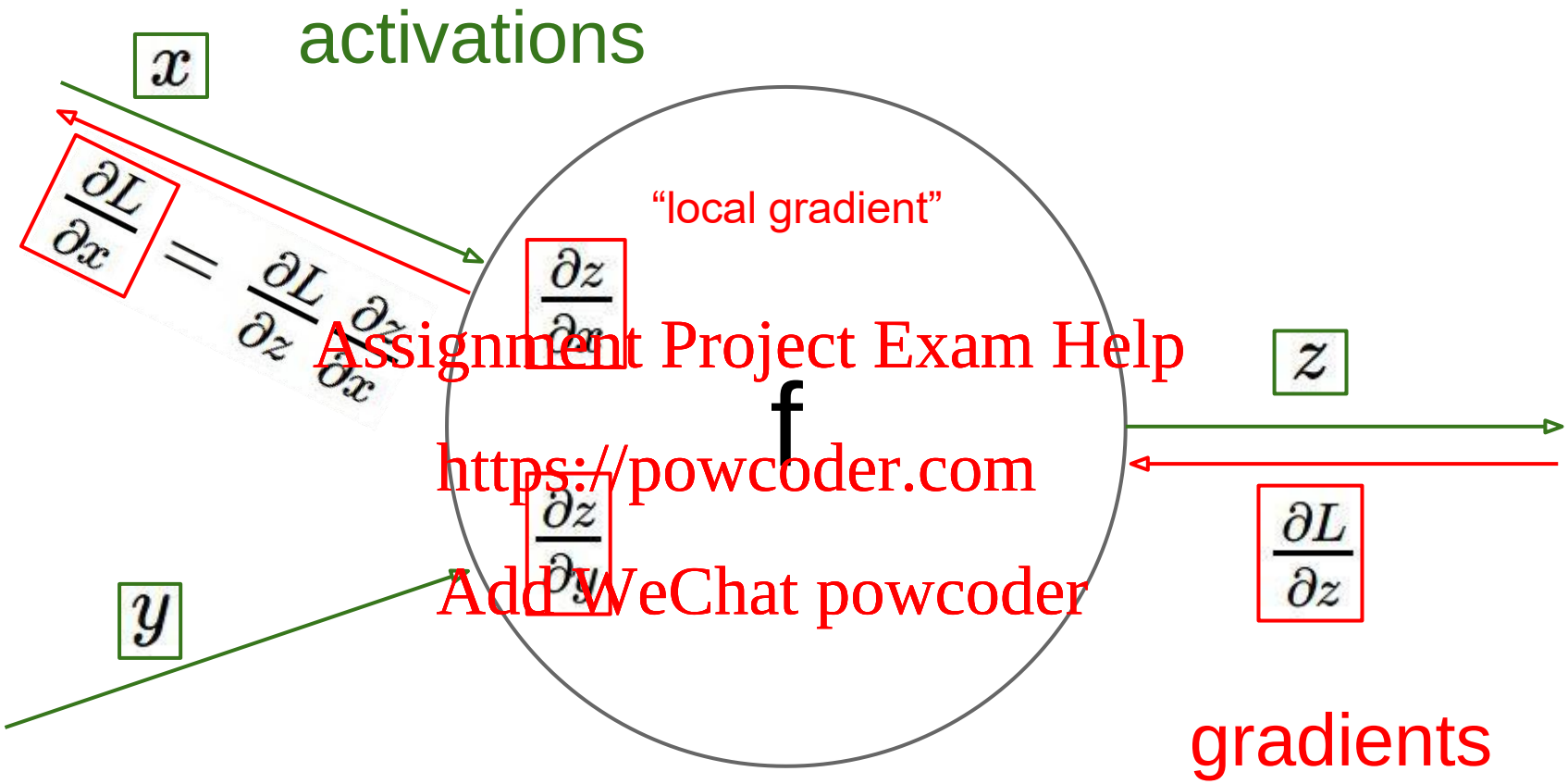
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder



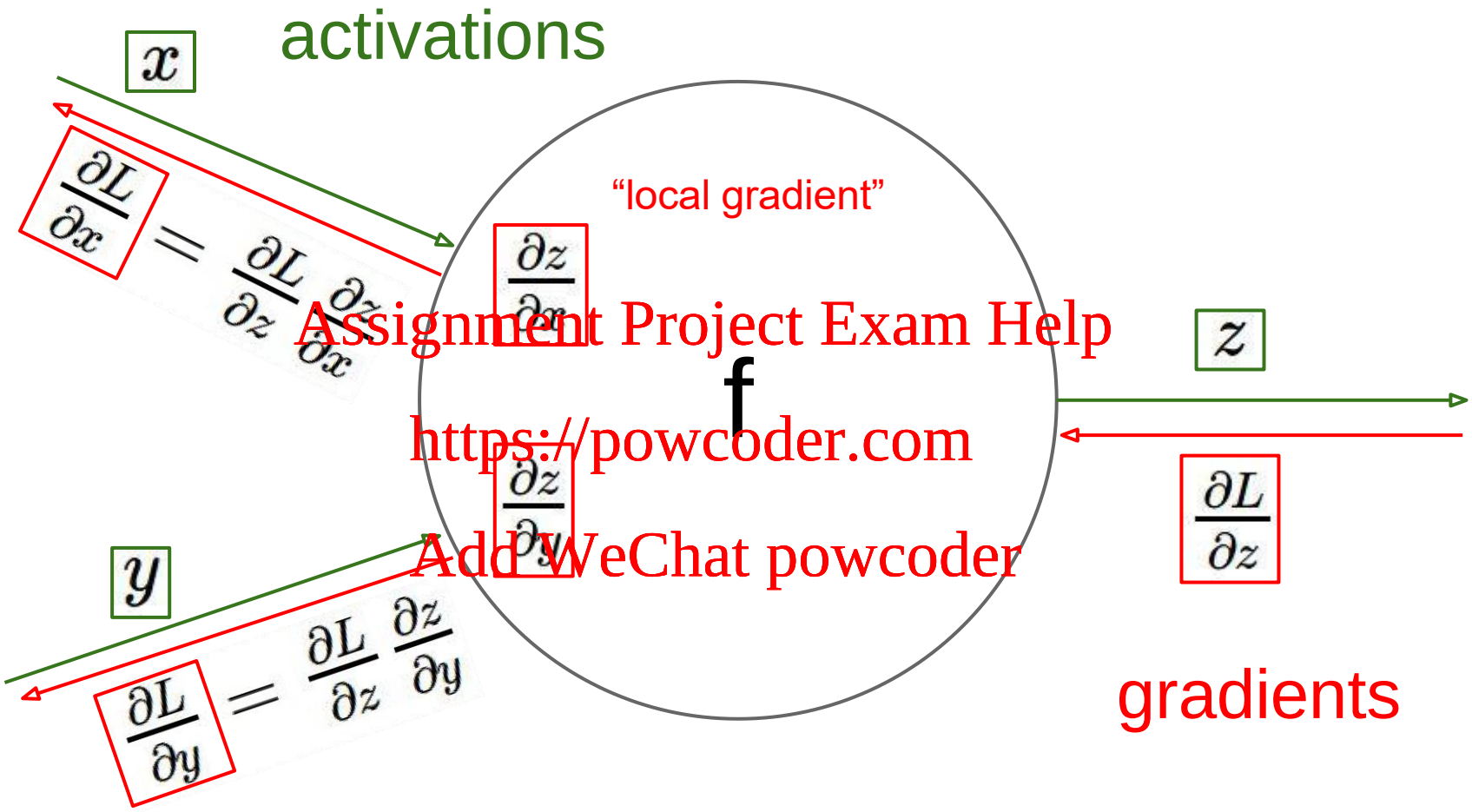
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder



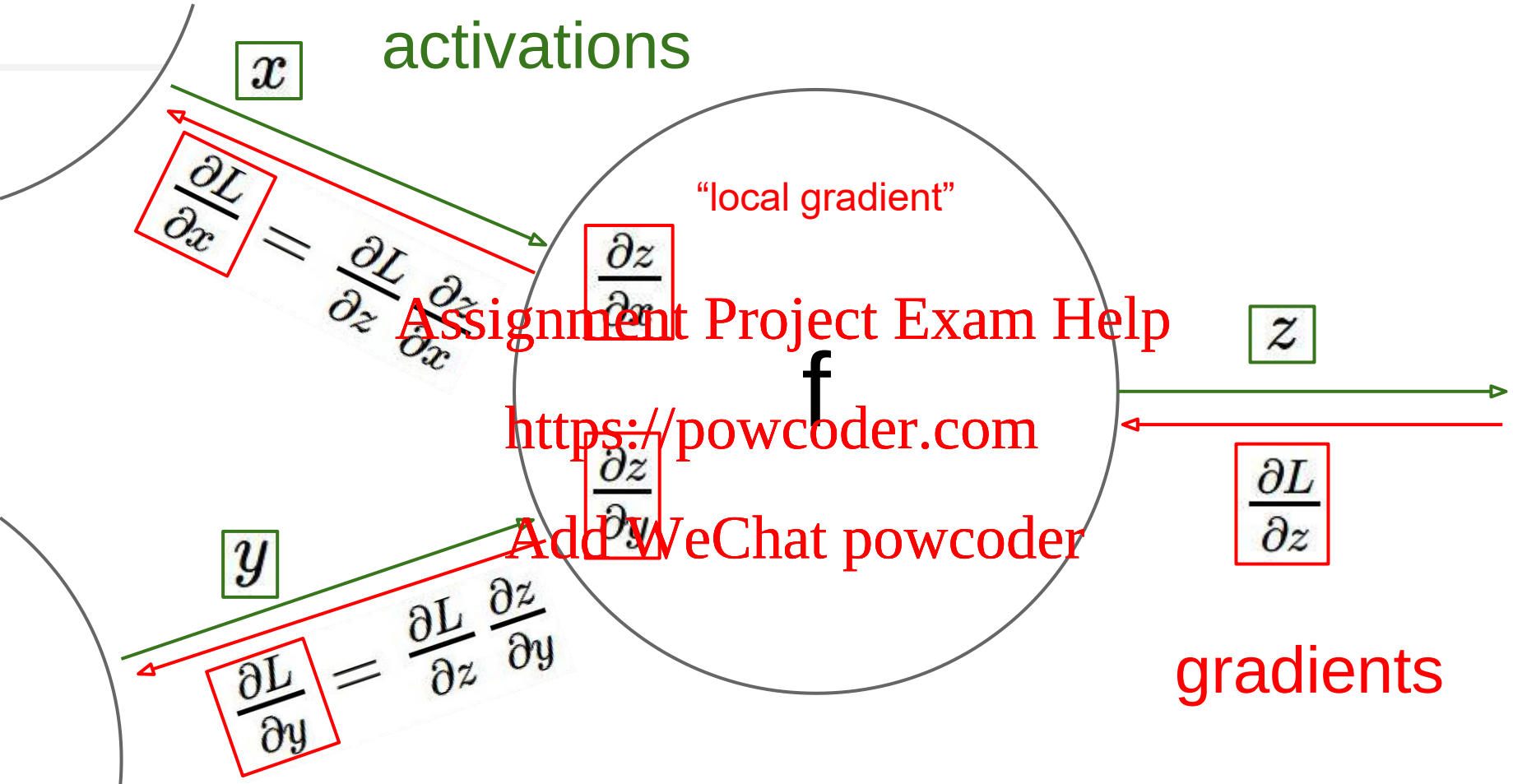
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder



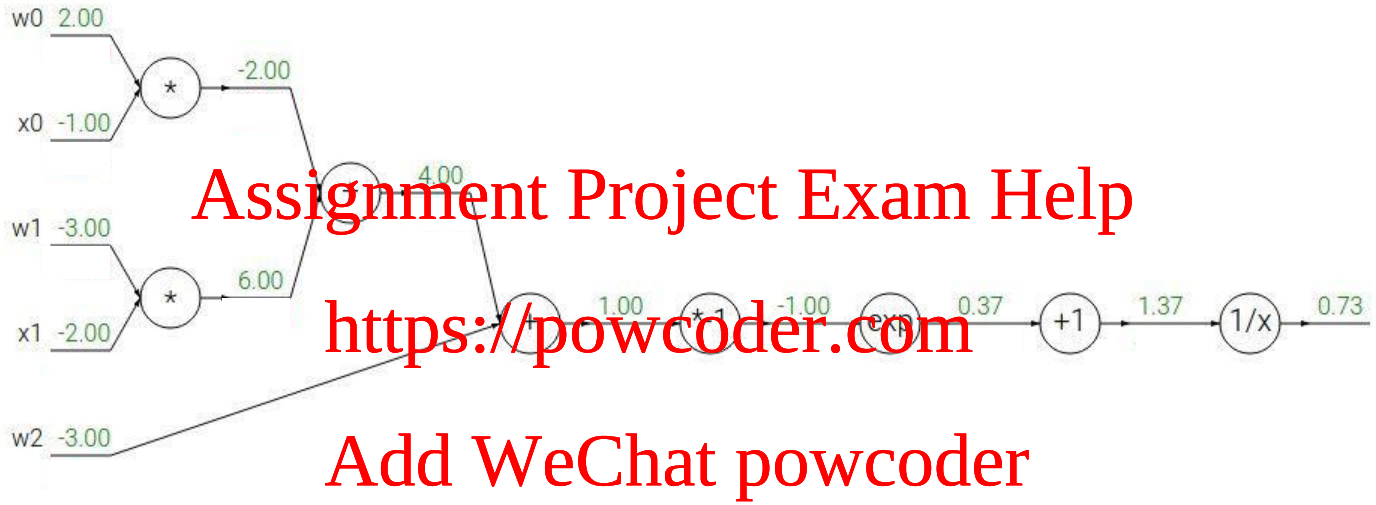
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

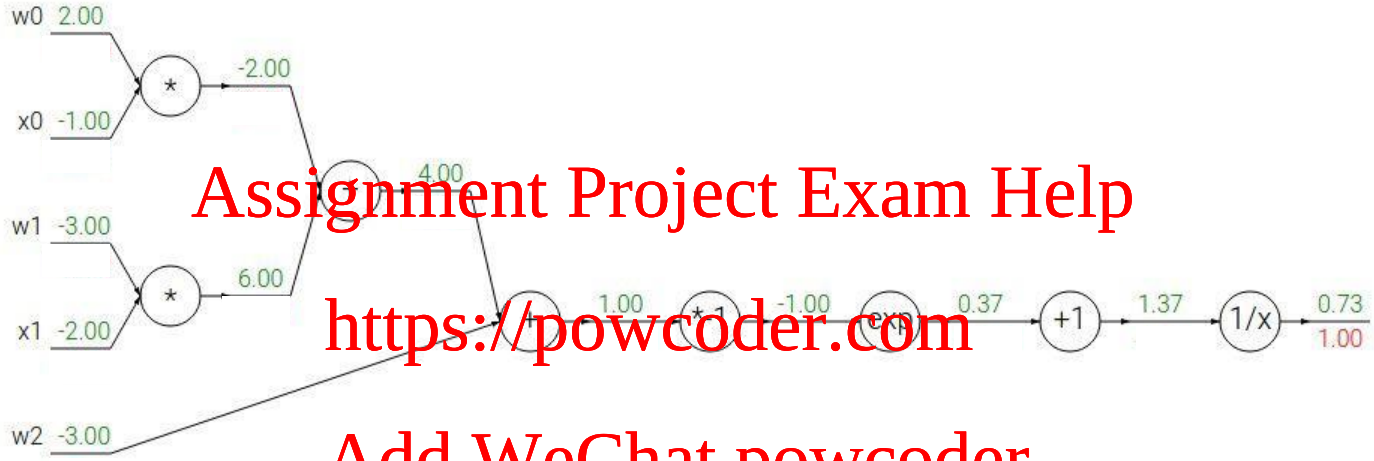
Another example:

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



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Another example: $$f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$



Assignment Project Exam Help

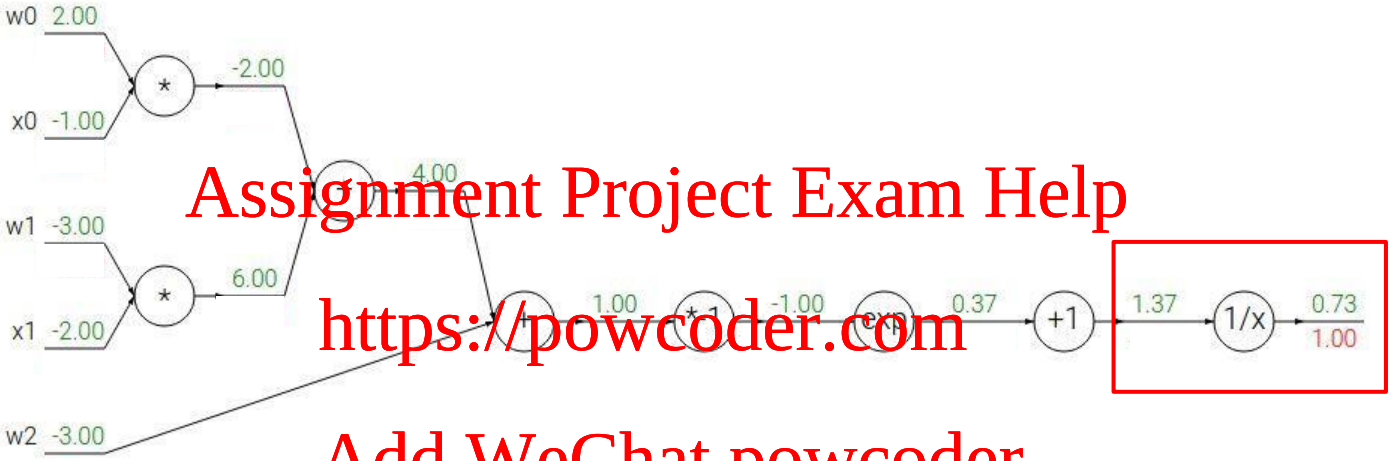
https://powcoder.com

Add WeChat powcoder

$$f(x) = e^x \rightarrow \frac{df}{dx} = e^x \qquad f(x) = \frac{1}{x} \rightarrow \frac{df}{dx} = -1/x^2$$

$$f_a(x) = ax \rightarrow \frac{df}{dx} = a \qquad f_c(x) = c + x \rightarrow \frac{df}{dx} = 1$$

Another example: $$f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$



Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

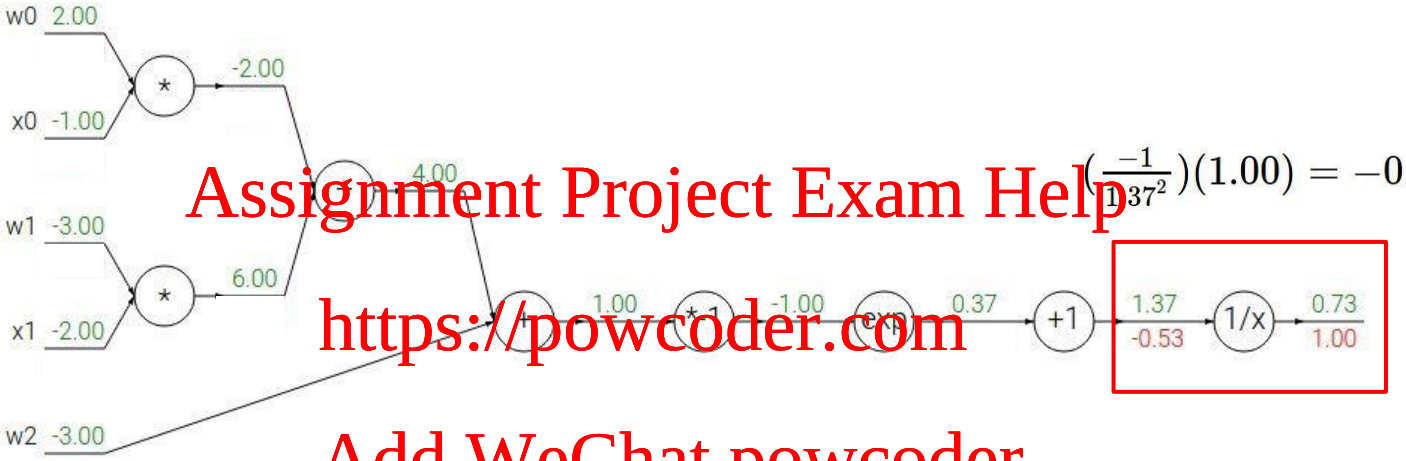
$f(x) = e^x \rightarrow \frac{df}{dx} = e^x$

$f_a(x) = ax \rightarrow \frac{df}{dx} = a$

$f(x) = \frac{1}{x} \rightarrow \frac{df}{dx} = -1/x^2$

$f_c(x) = c + x \rightarrow \frac{df}{dx} = 1$

Another example: $$f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$



$$(\frac{-1}{1.37^2})(1.00) = -0.53$$

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

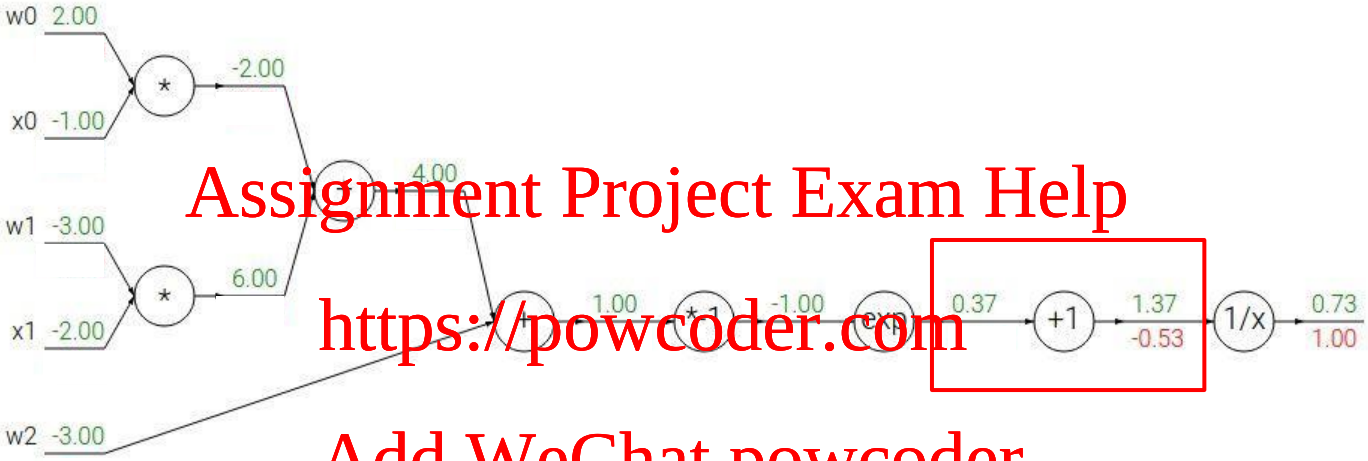
$f(x) = e^x \rightarrow \frac{df}{dx} = e^x$

$f_a(x) = ax \rightarrow \frac{df}{dx} = a$

$f(x) = \frac{1}{x} \rightarrow \frac{df}{dx} = -1/x^2$

$f_c(x) = c + x \rightarrow \frac{df}{dx} = 1$

Another example: $$f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

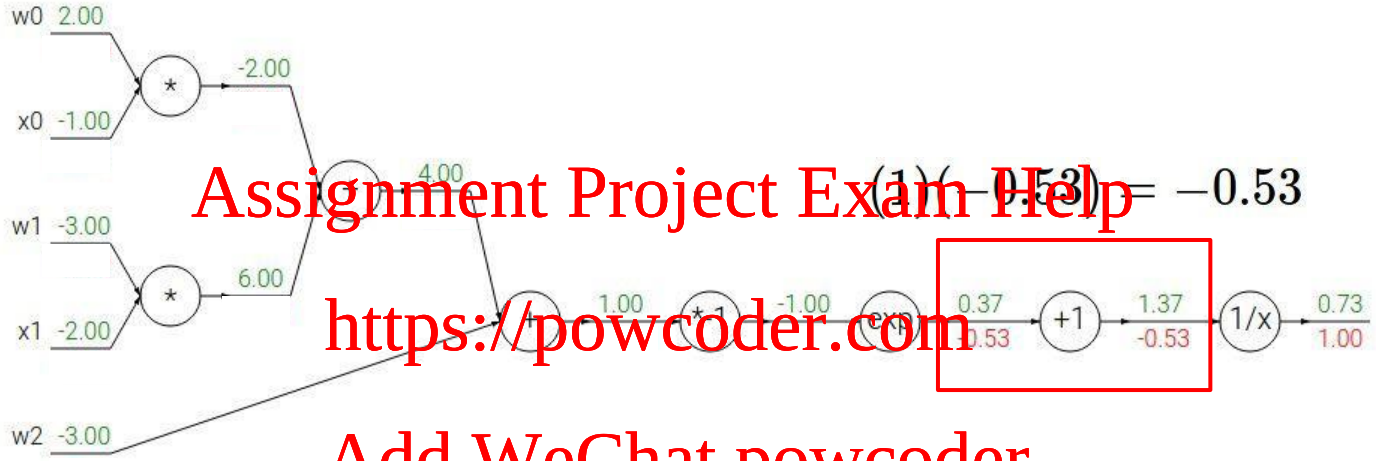
$f(x) = e^x \rightarrow \frac{df}{dx} = e^x$

$f_a(x) = ax \rightarrow \frac{df}{dx} = a$

$f(x) = \frac{1}{x} \rightarrow \frac{df}{dx} = -1/x^2$

$f_c(x) = c + x \rightarrow \frac{df}{dx} = 1$

Another example: $$f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$



Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

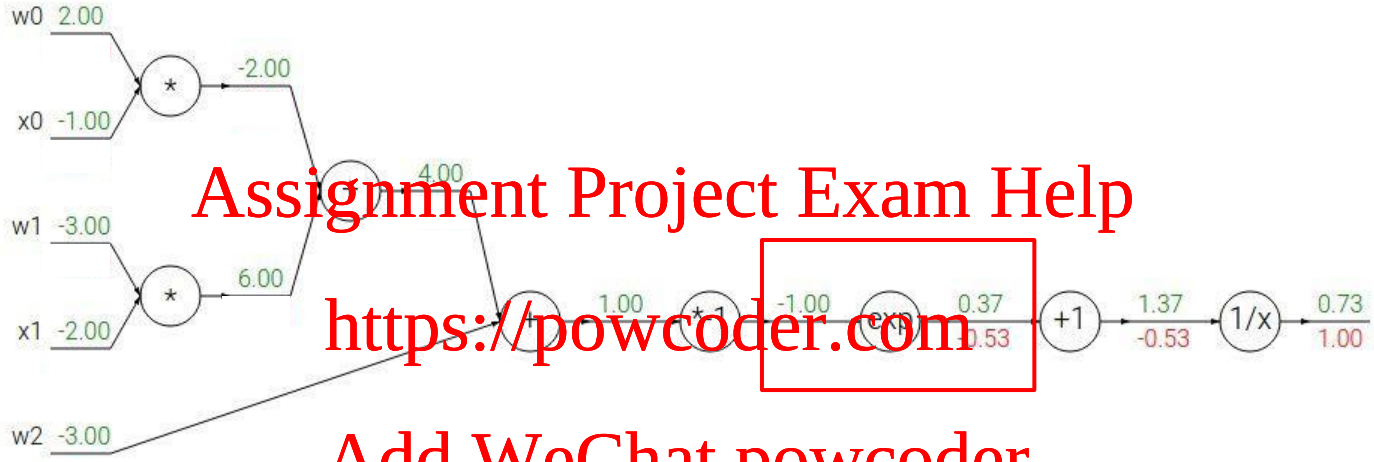
$$f(x) = e^x \rightarrow \frac{df}{dx} = e^x$$

$$f_a(x) = ax \rightarrow \frac{df}{dx} = a$$

$$f(x) = \frac{1}{x} \rightarrow \frac{df}{dx} = -1/x^2$$

$$f_c(x) = c + x \rightarrow \frac{df}{dx} = 1$$

Another example: $$f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

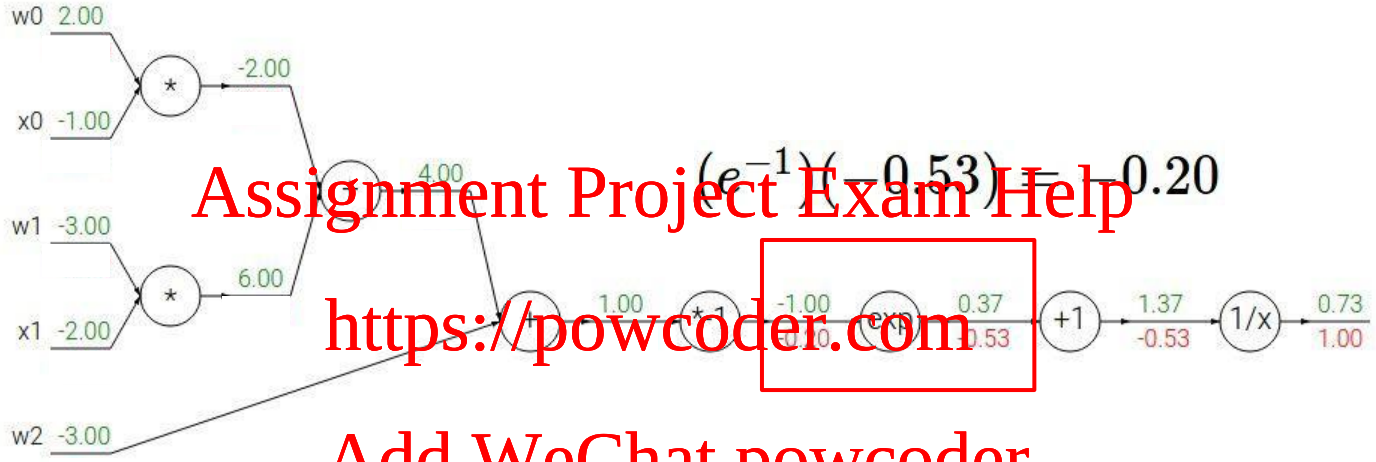
$f(x) = e^x \rightarrow \frac{df}{dx} = e^x$

$f_a(x) = ax \rightarrow \frac{df}{dx} = a$

$f(x) = \frac{1}{x} \rightarrow \frac{df}{dx} = -1/x^2$

$f_c(x) = c + x \rightarrow \frac{df}{dx} = 1$

Another example: $$f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$



$$(e^{-1})(-0.53) = -0.20$$

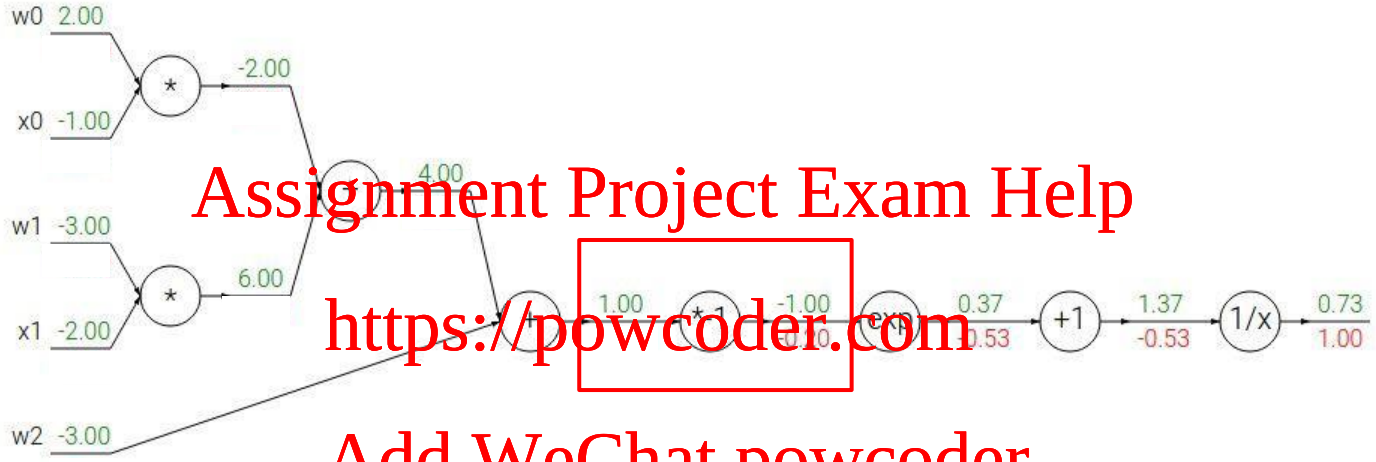
Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

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

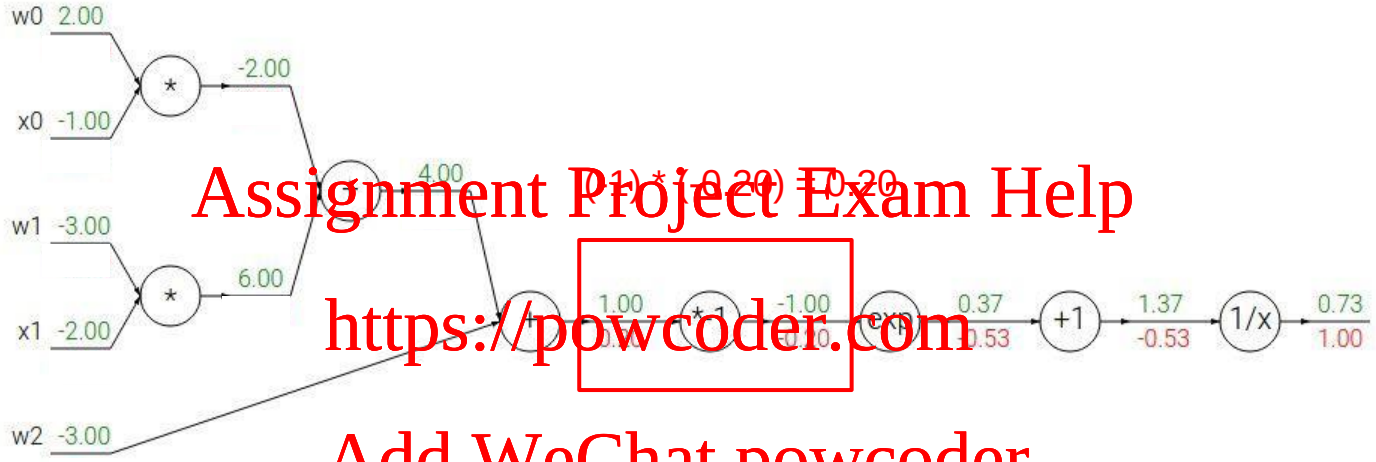
Another example: $$f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

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

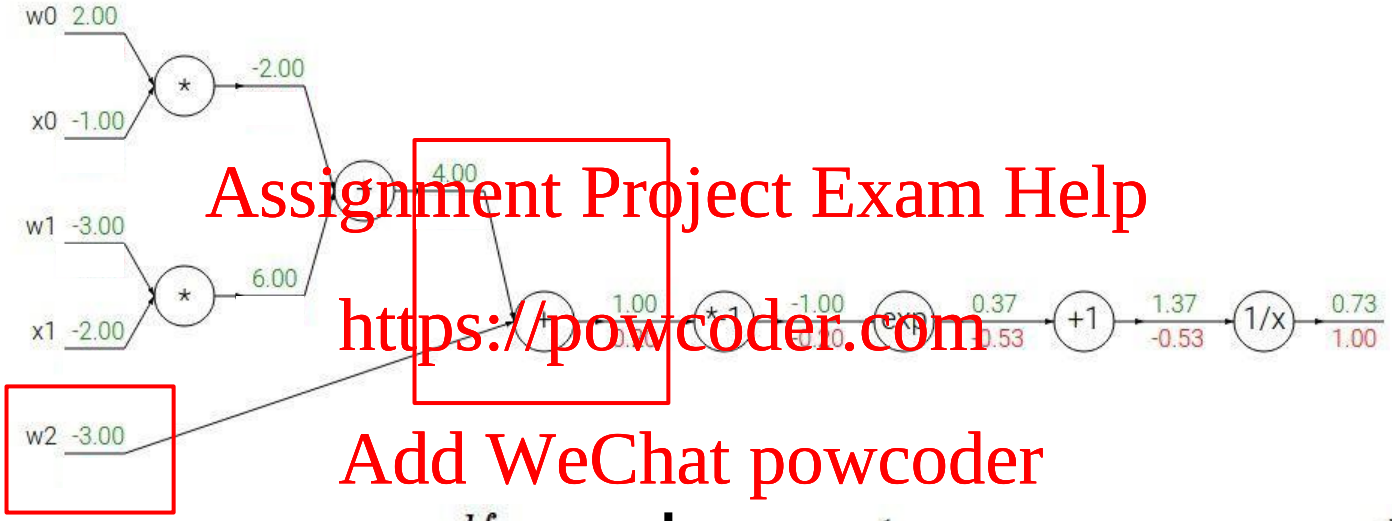
Another example: $$f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

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

Another example: $$f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$

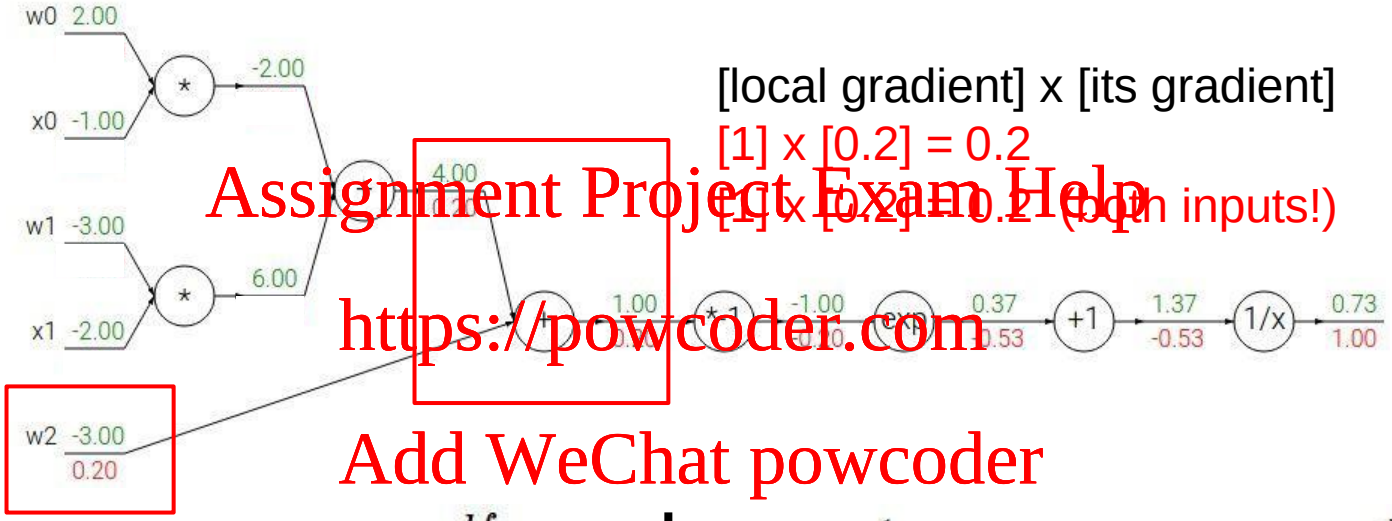


Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

$$f(x) = e^x \rightarrow \frac{df}{dx} = e^x \qquad f(x) = \frac{1}{x} \rightarrow \frac{df}{dx} = -1/x^2$$

$$f_a(x) = ax \rightarrow \frac{df}{dx} = a \qquad f_c(x) = c + x \rightarrow \frac{df}{dx} = 1$$

Another example: $$f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$



[local gradient] x [its gradient]

[1] x [0.2] = 0.2

[1] x [0.2] = 0.2 (both inputs!)

Assignment Project Exam Help

https://powcoder.com

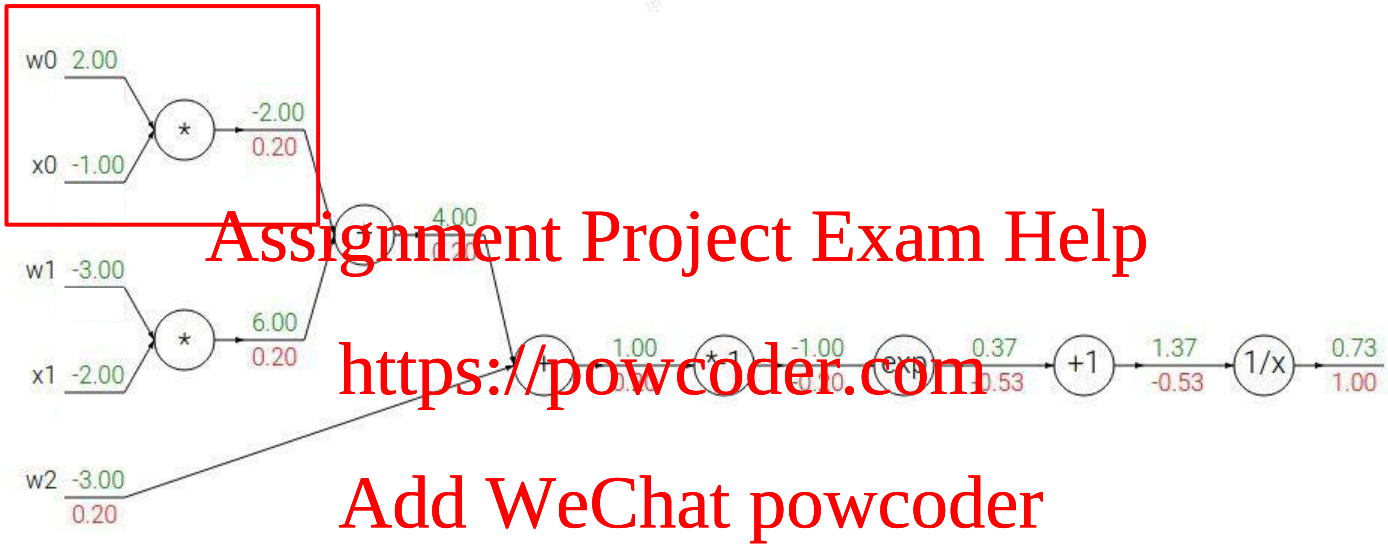
Add WeChat powcoder

$$f(x) = e^x \rightarrow \frac{df}{dx} = e^x \qquad f(x) = \frac{1}{x} \rightarrow \frac{df}{dx} = -1/x^2$$

$$f_a(x) = ax \rightarrow \frac{df}{dx} = a \qquad f_c(x) = c + x \rightarrow \frac{df}{dx} = 1$$

Another example:

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



Assignment Project Exam Help

https://powcoder.com

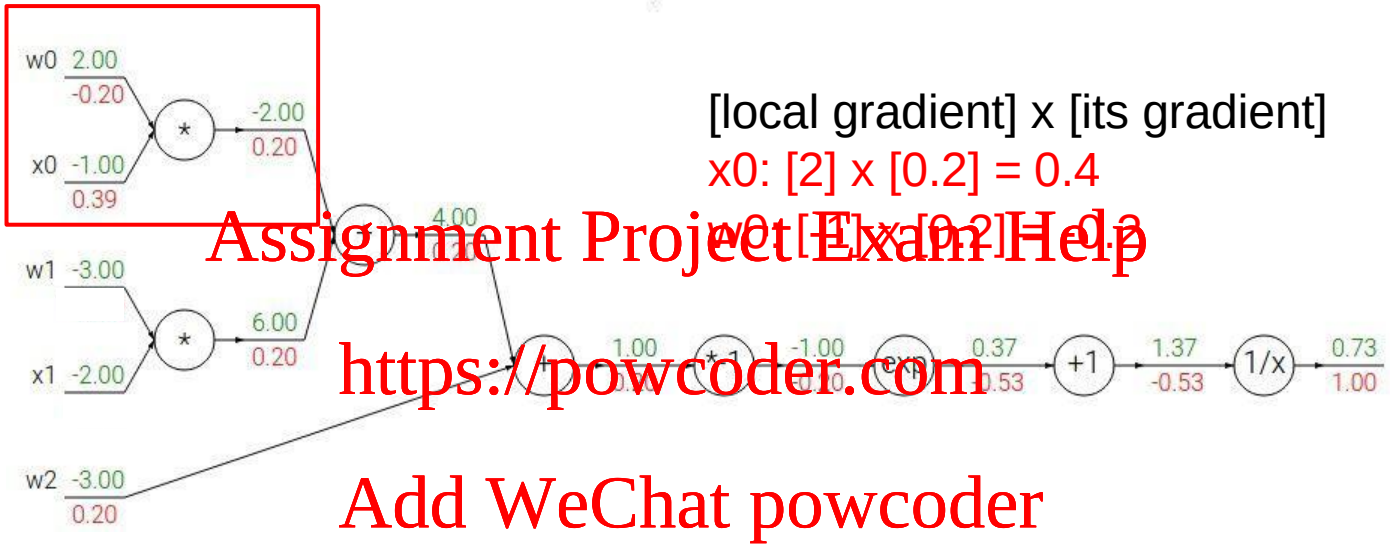
Add WeChat powcoder

$$f(x) = e^x \rightarrow \frac{df}{dx} = e^x \qquad f(x) = \frac{1}{x} \rightarrow \frac{df}{dx} = -1/x^2$$

$$f_a(x) = ax \rightarrow \frac{df}{dx} = a \qquad f_c(x) = c + x \rightarrow \frac{df}{dx} = 1$$

Another example:

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



[local gradient] x [its gradient]

x0: [2] x [0.2] = 0.4

w0: [-1] x [0.2] = -0.2

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

$$f(x) = e^x \rightarrow \frac{df}{dx} = e^x \qquad f(x) = \frac{1}{x} \rightarrow \frac{df}{dx} = -1/x^2$$

$$f_a(x) = ax \rightarrow \frac{df}{dx} = a \qquad f_c(x) = c + x \rightarrow \frac{df}{dx} = 1$$

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

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

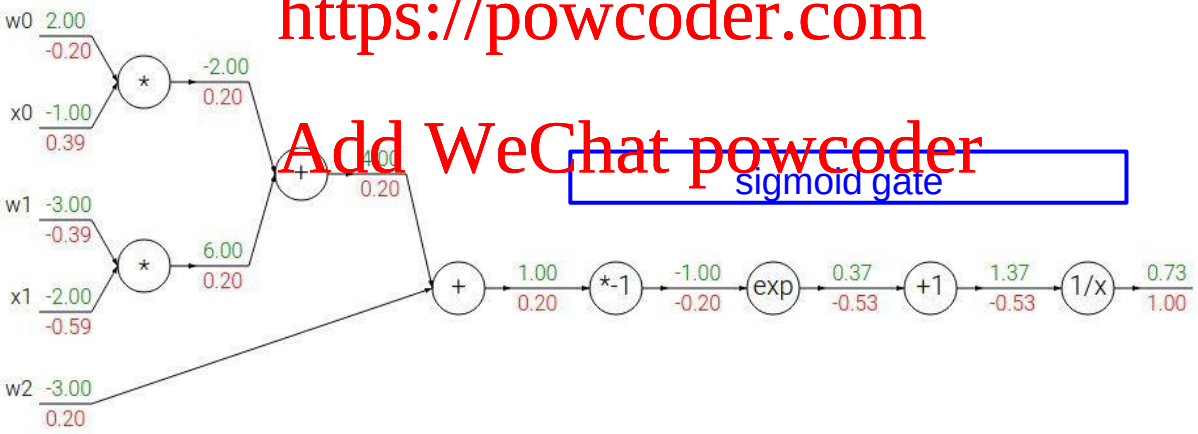
sigmoid function

$$\frac{d\sigma(x)}{dx} = \frac{e^{-x}}{(1+e^{-x})^2} = \left(\frac{1+e^{-x}-1}{1+e^{-x}}\right)\left(\frac{1}{1+e^{-x}}\right) = (1-\sigma(x))\,\sigma(x)$$

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder



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

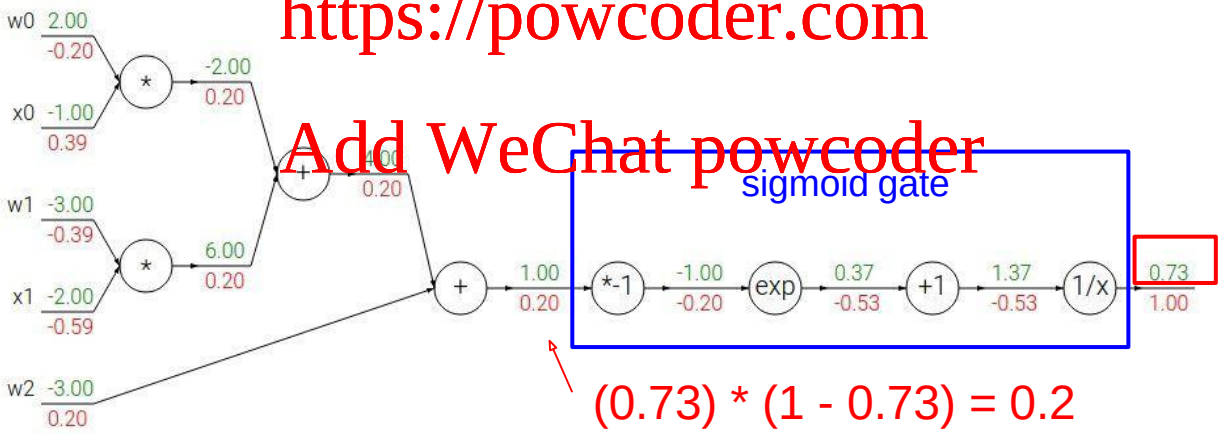
$$\sigma(x) = \frac{1}{1+e^{-x}}$$ sigmoid function

$$\frac{d\sigma(x)}{dx} = \frac{e^{-x}}{(1+e^{-x})^2} = \left(\frac{1+e^{-x}-1}{1+e^{-x}}\right)\left(\frac{1}{1+e^{-x}}\right) = (1-\sigma(x))\,\sigma(x)$$

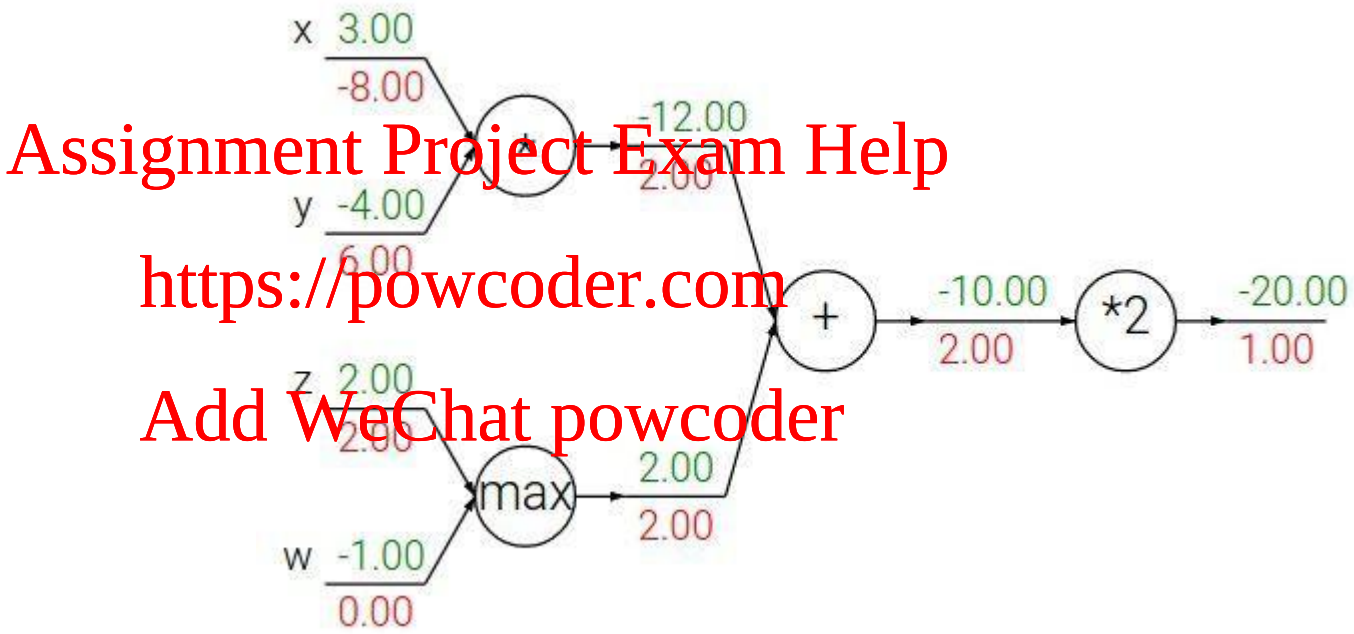
Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder



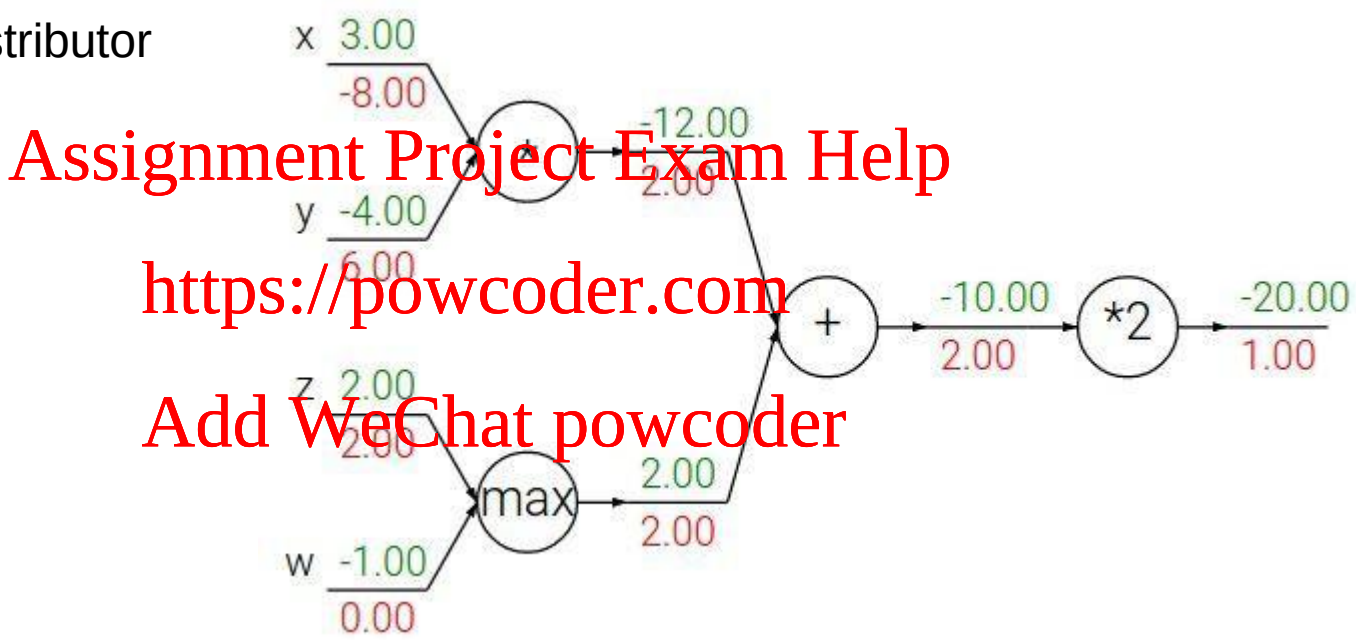
Patterns in backward flow



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Patterns in backward flow

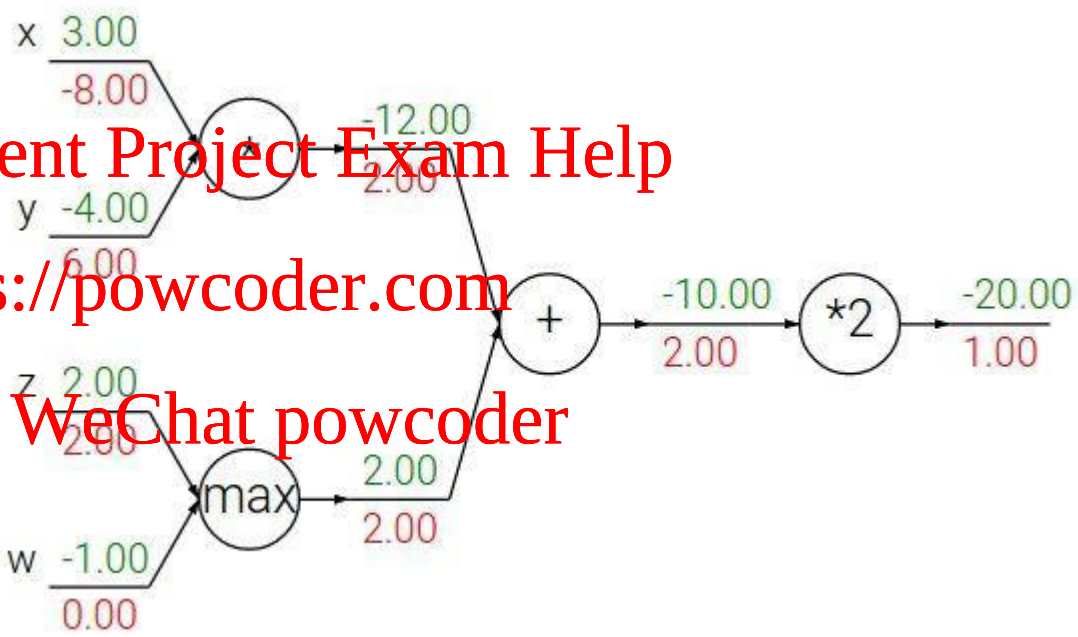
add gate: gradient distributor



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Patterns in backward flow

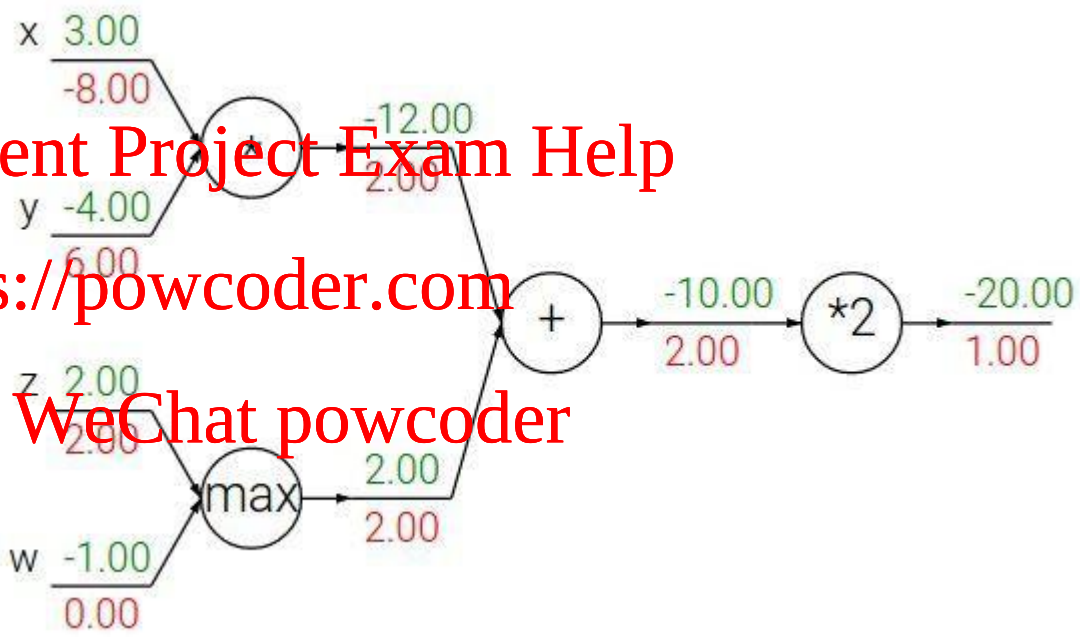
add gate: gradient distributor
max gate: gradient router



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

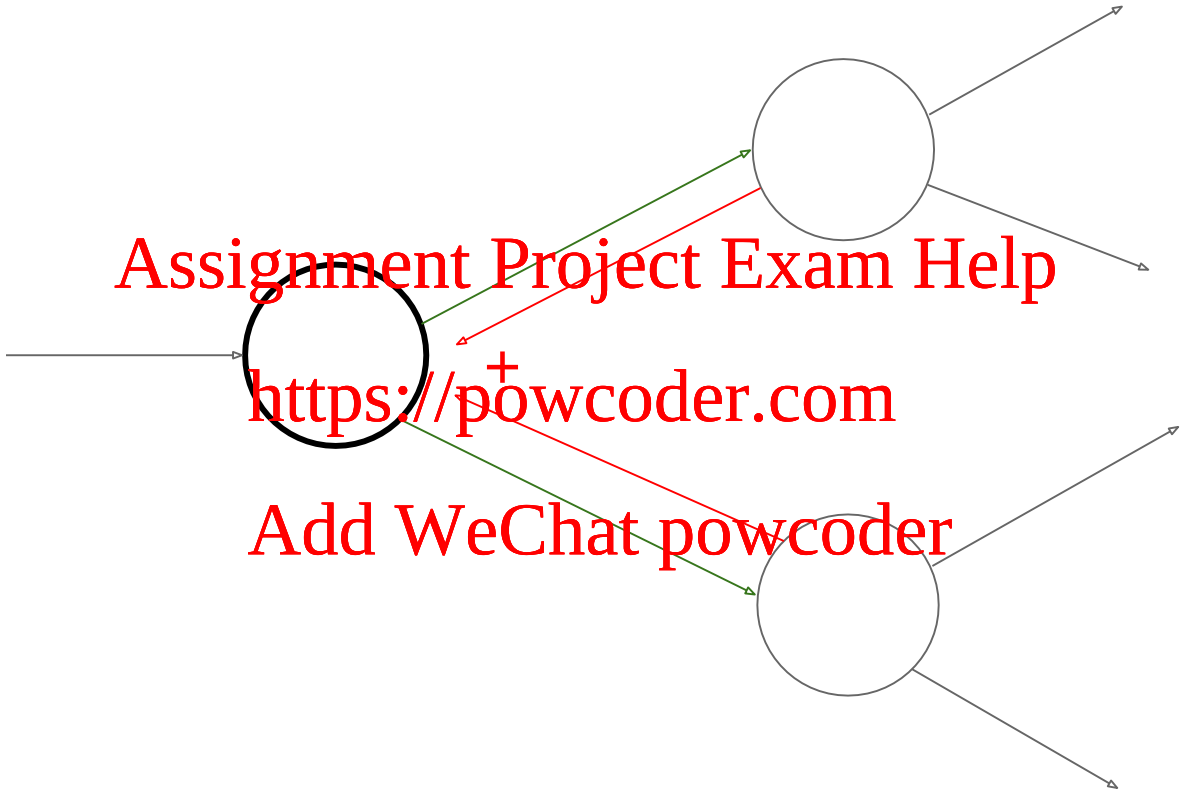
Patterns in backward flow

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



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Gradients add at branches



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Neural Networks II

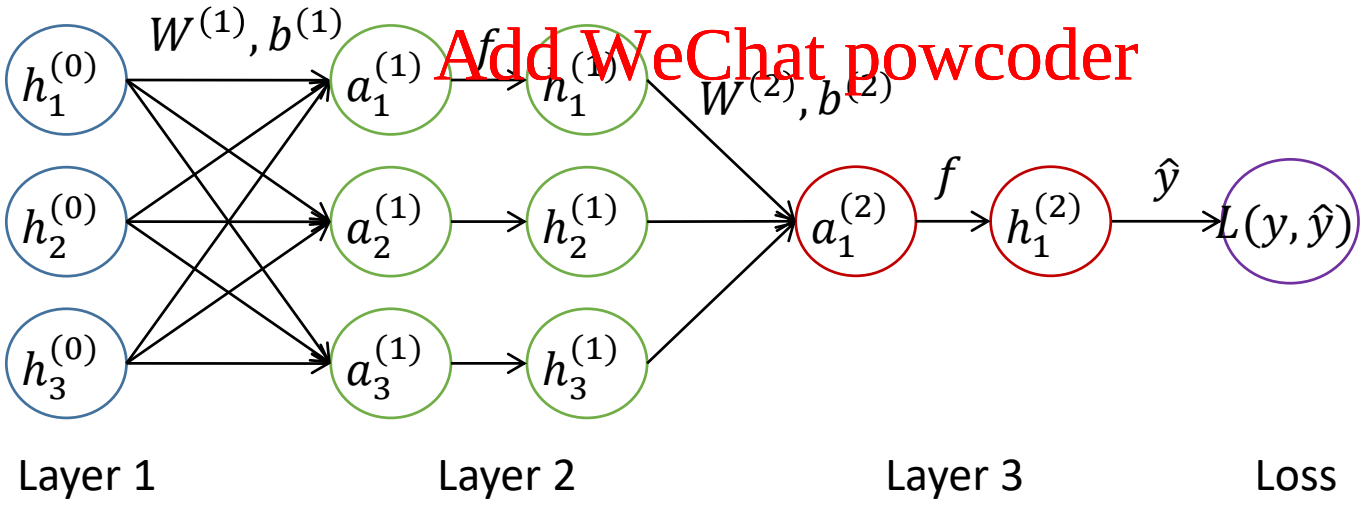
Vectorized Backpropagation

Forward Pass

Require: Network depth, l
Require: $\boldsymbol{W}^{(i)}, i \in \{1, \ldots, l\}$, the weight matrices of the model
Require: $\boldsymbol{b}^{(i)}, i \in \{1, \ldots, l\}$, the bias parameters of the model
Require: $\boldsymbol{x}$, the input to process
Require: $\boldsymbol{y}$, the target output

$\boldsymbol{h}^{(0)} = \boldsymbol{x}$
for $k = 1, \ldots, l$ **do**
  $\boldsymbol{a}^{(k)} = \boldsymbol{b}^{(k)} + \boldsymbol{W}^{(k)} \boldsymbol{h}^{(k-1)}$
  $\boldsymbol{h}^{(k)} = f(\boldsymbol{a}^{(k)})$
end for
$\hat{\boldsymbol{y}} = \boldsymbol{h}^{(l)}$
$J = L(\hat{\boldsymbol{y}}, \boldsymbol{y}) + \lambda \Omega(\theta)$

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder



Backward Pass

After the forward computation, compute the gradient on the output layer:

$\boldsymbol{g} \leftarrow \nabla_{\hat{\boldsymbol{y}}} J = \nabla_{\hat{\boldsymbol{y}}} L(\hat{\boldsymbol{y}}, \boldsymbol{y})$

for $k = l, l-1, \ldots, 1$ **do**

Convert the gradient on the layer's output into a gradient into the pre-nonlinearity activation (element-wise multiplication if f is element-wise):

$\boldsymbol{g} \leftarrow \nabla_{\boldsymbol{a}^{(k)}} J = \boldsymbol{g} \odot f'(\boldsymbol{a}^{(k)})$

Compute gradients on weights and biases (including the regularization term, where needed):

$\nabla_{\boldsymbol{b}^{(k)}} J = \boldsymbol{g} + \lambda \nabla_{\boldsymbol{b}^{(k)}} \Omega(\theta)$

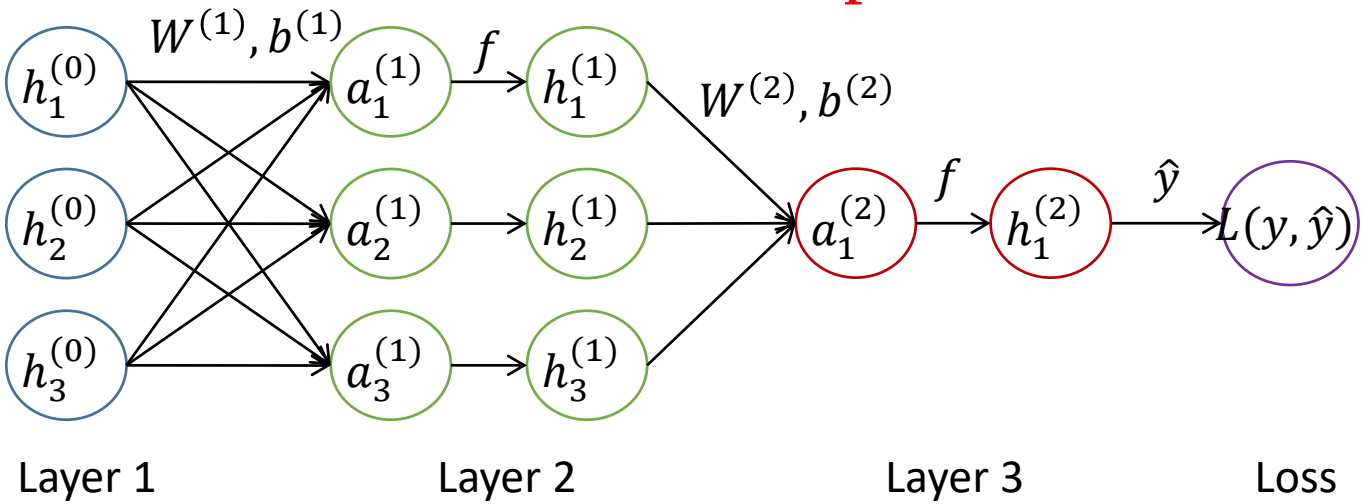
$\nabla_{\boldsymbol{W}^{(k)}} J = \boldsymbol{g}\, \boldsymbol{h}^{(k-1)\top} + \lambda \nabla_{\boldsymbol{W}^{(k)}} \Omega(\theta)$

Propagate the gradients w.r.t. the next lower-level hidden layer's activations:

$\boldsymbol{g} \leftarrow \nabla_{\boldsymbol{h}^{(k-1)}} J = \boldsymbol{W}^{(k)\top} \boldsymbol{g}$

end for

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder



Backpropagation example with vectorized gradients:

https://web.stanford.edu/class/cs224n/readings/gradient-notes.pdf

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Next Class

Neural Networks III: Convolutional Nets:

Convolutional networks.

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Reading: Bishop Ch 5.5