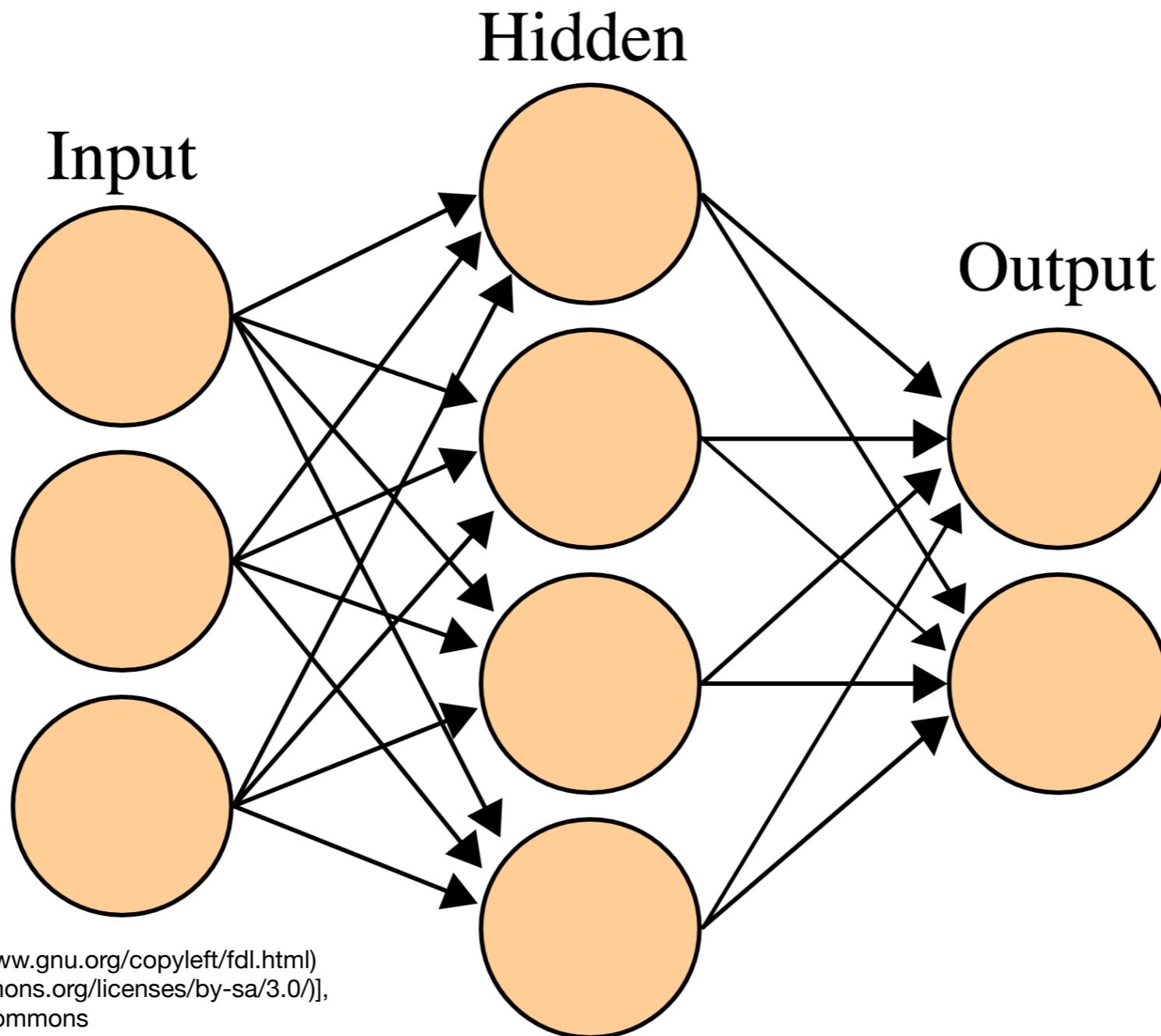
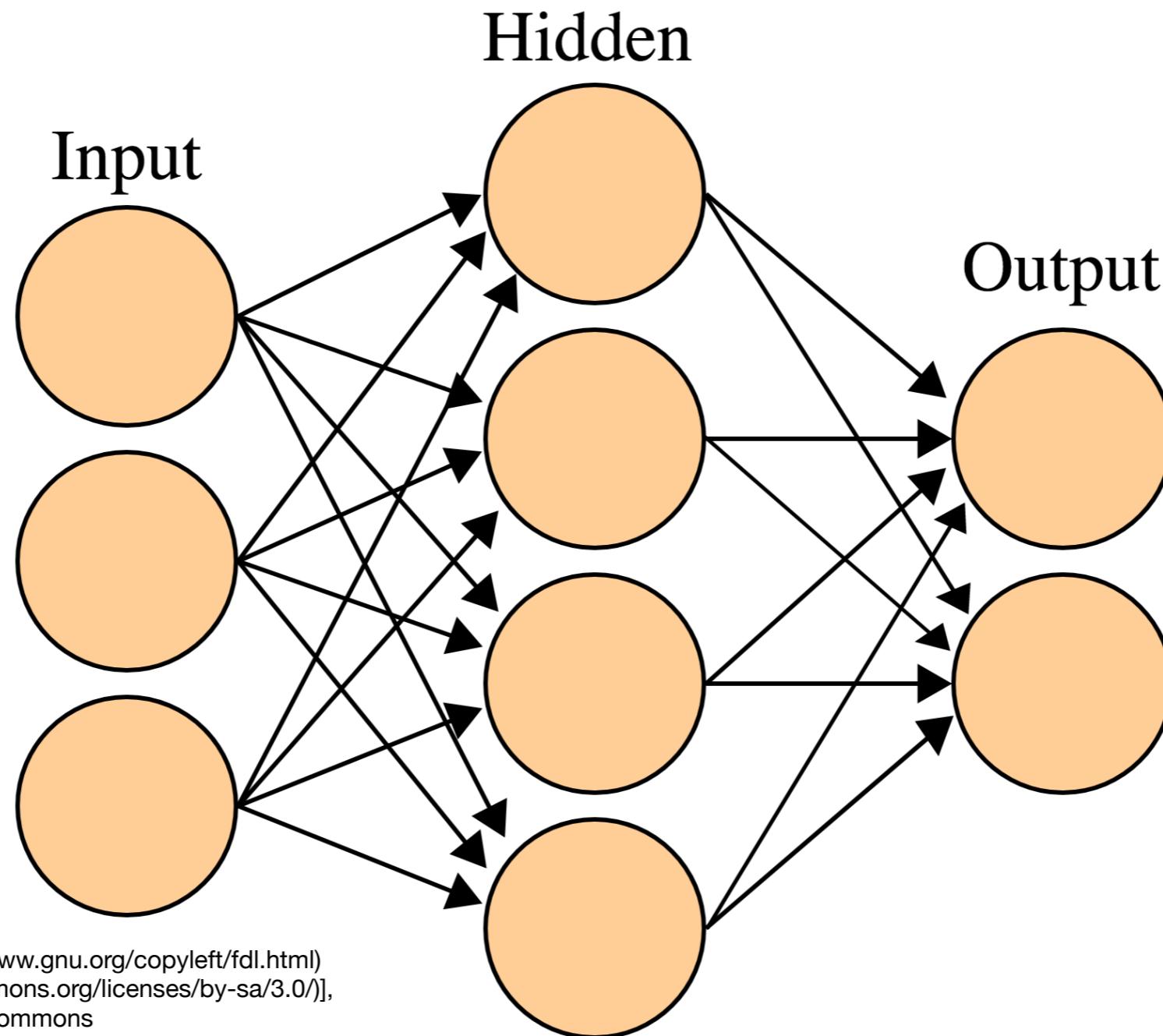


# **Neural Network Training**



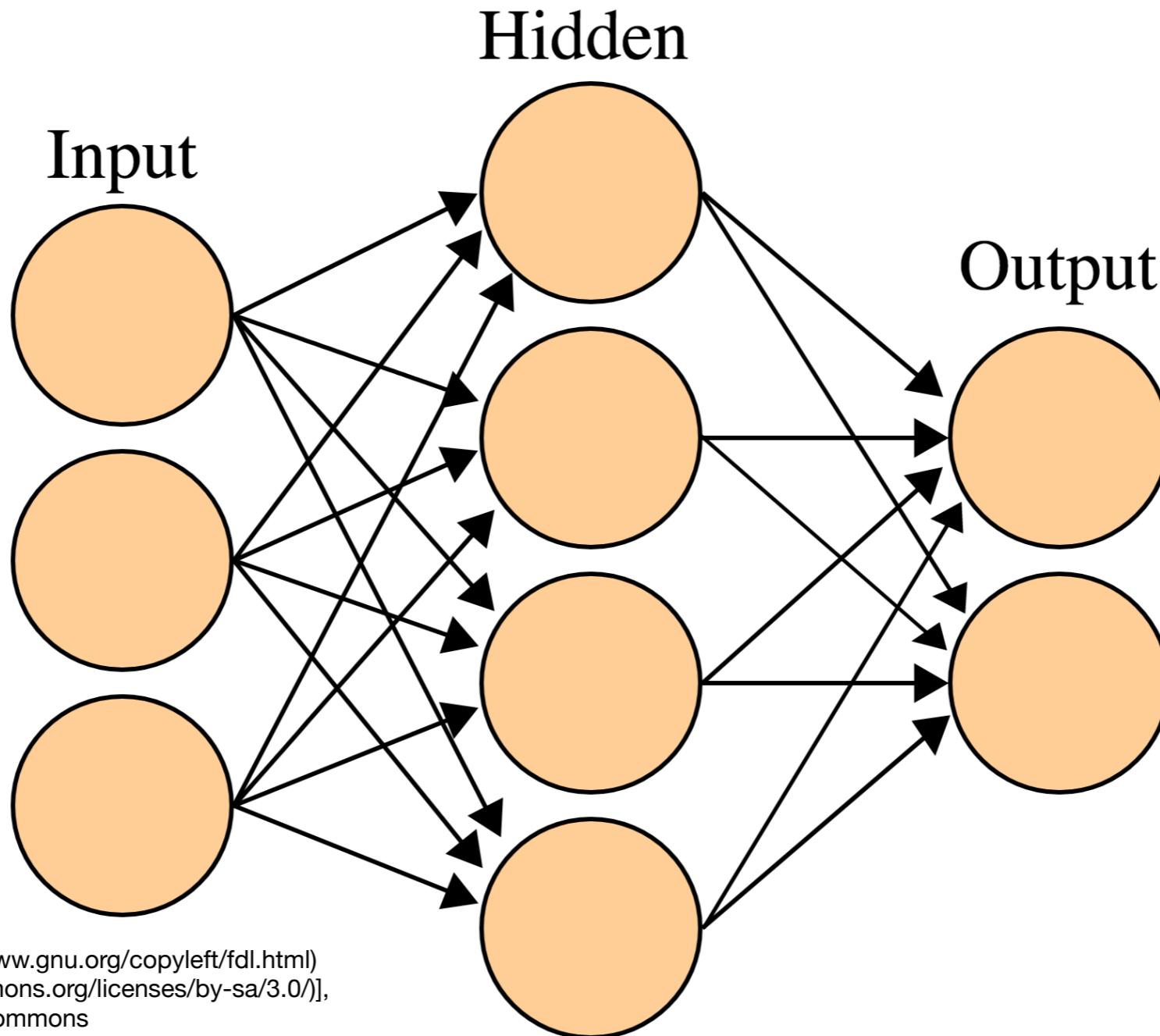
By en:User:Cburnett [GFDL (<http://www.gnu.org/copyleft/fdl.html>)  
or CC-BY-SA-3.0 (<http://creativecommons.org/licenses/by-sa/3.0/>)],  
via Wikimedia Commons

$X$

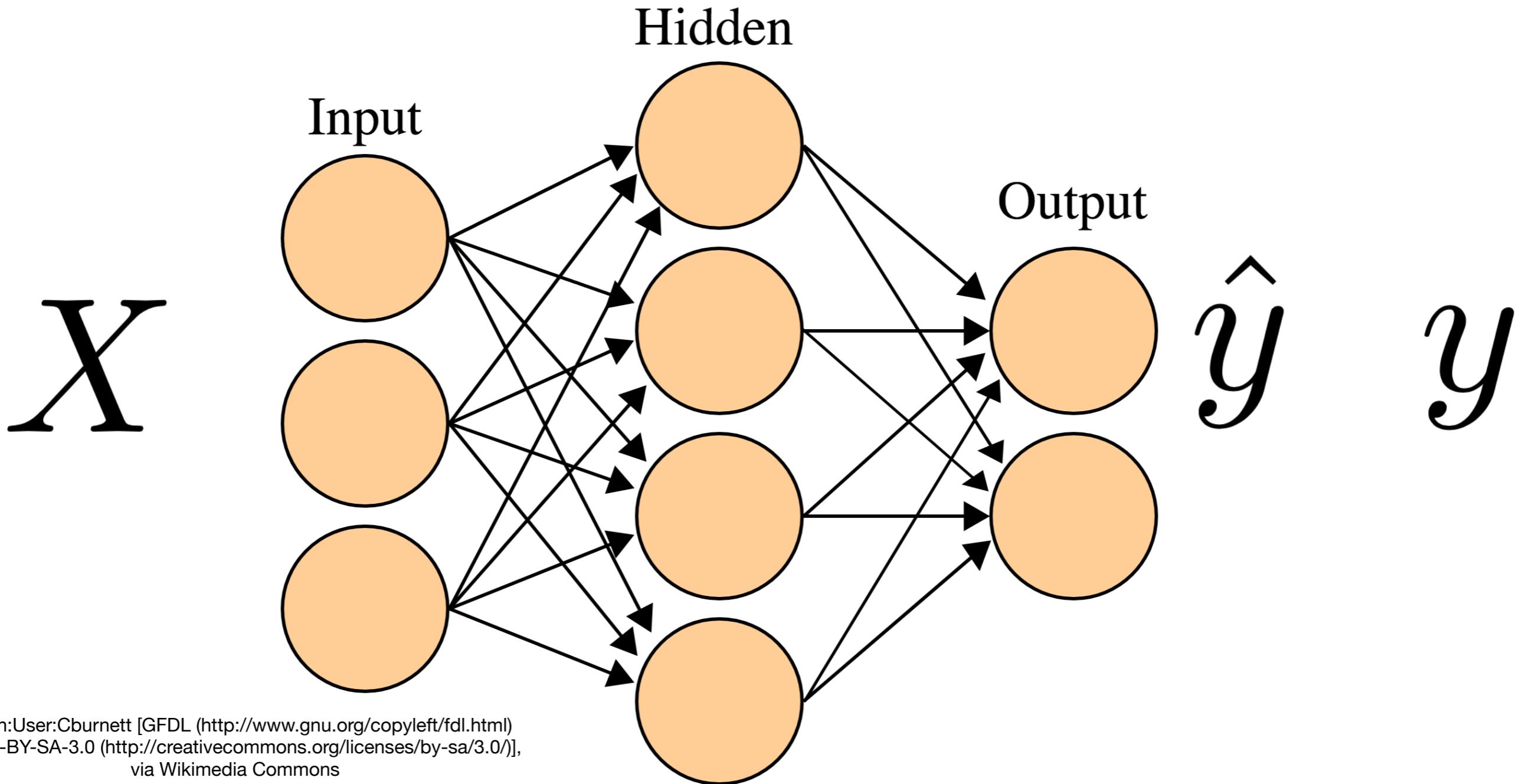


By en:User:Cburnett [GFDL (<http://www.gnu.org/copyleft/fdl.html>)  
or CC-BY-SA-3.0 (<http://creativecommons.org/licenses/by-sa/3.0/>)],  
via Wikimedia Commons

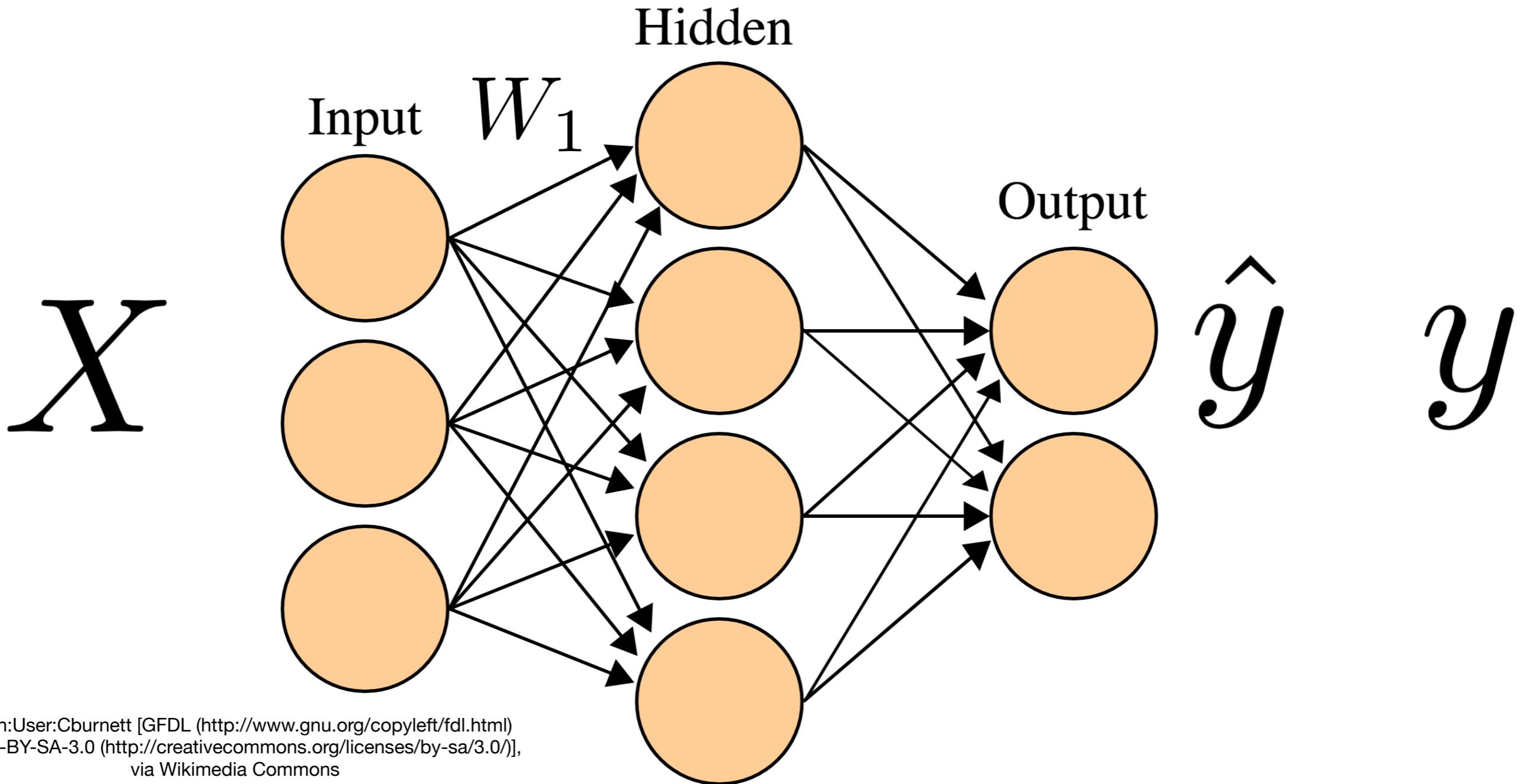
$X$



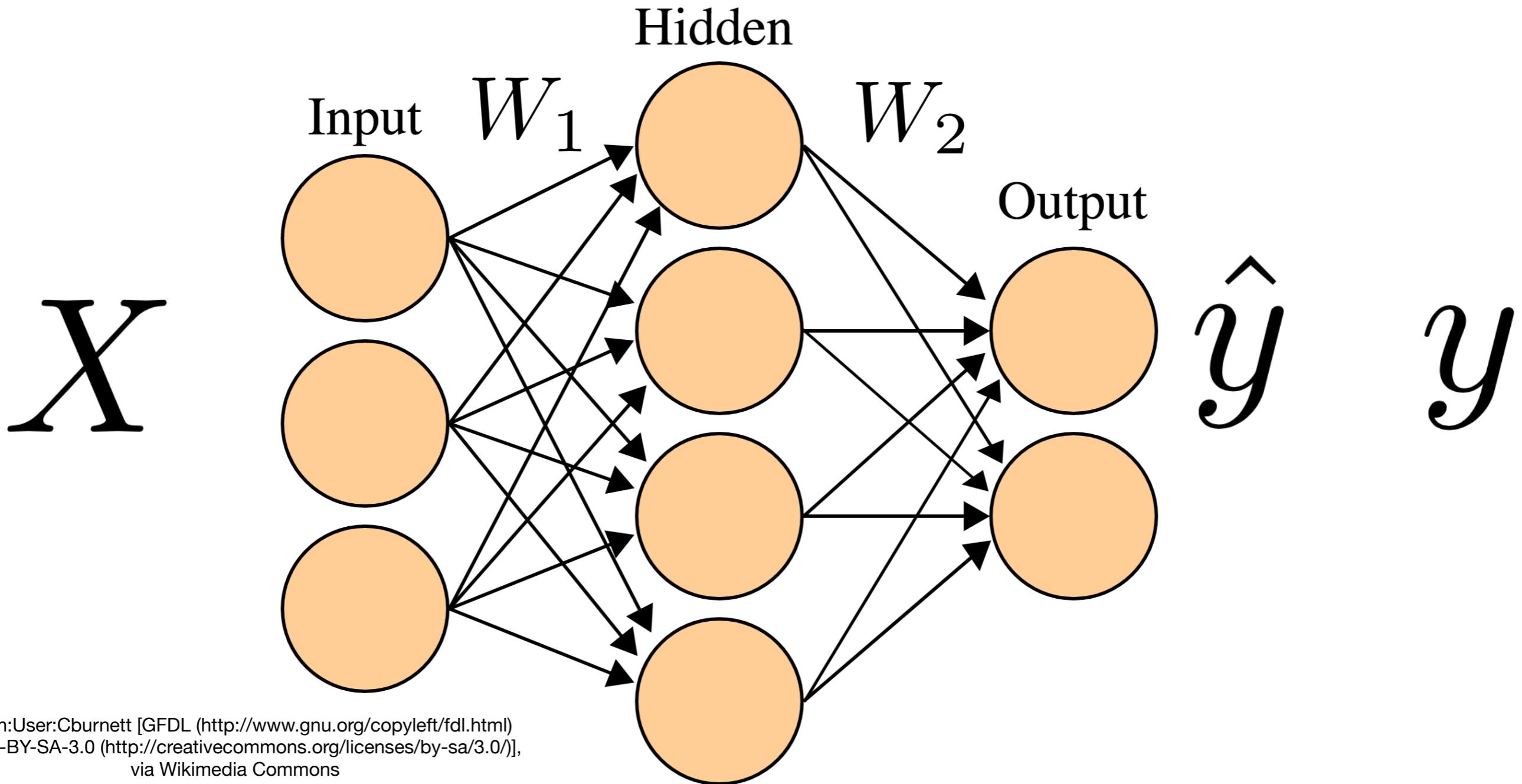
By en:User:Cburnett [GFDL (<http://www.gnu.org/copyleft/fdl.html>)  
or CC-BY-SA-3.0 (<http://creativecommons.org/licenses/by-sa/3.0/>)],  
via Wikimedia Commons

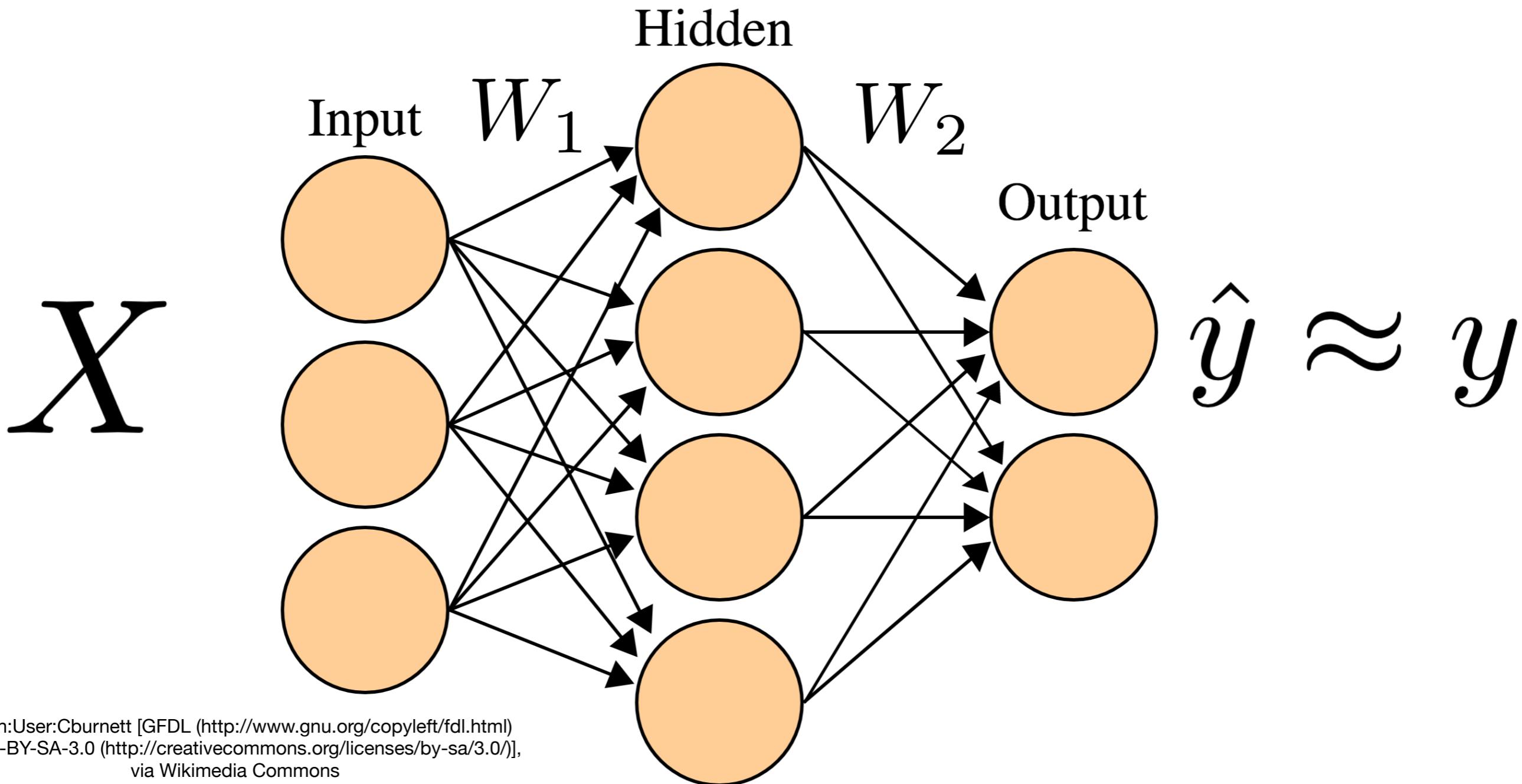


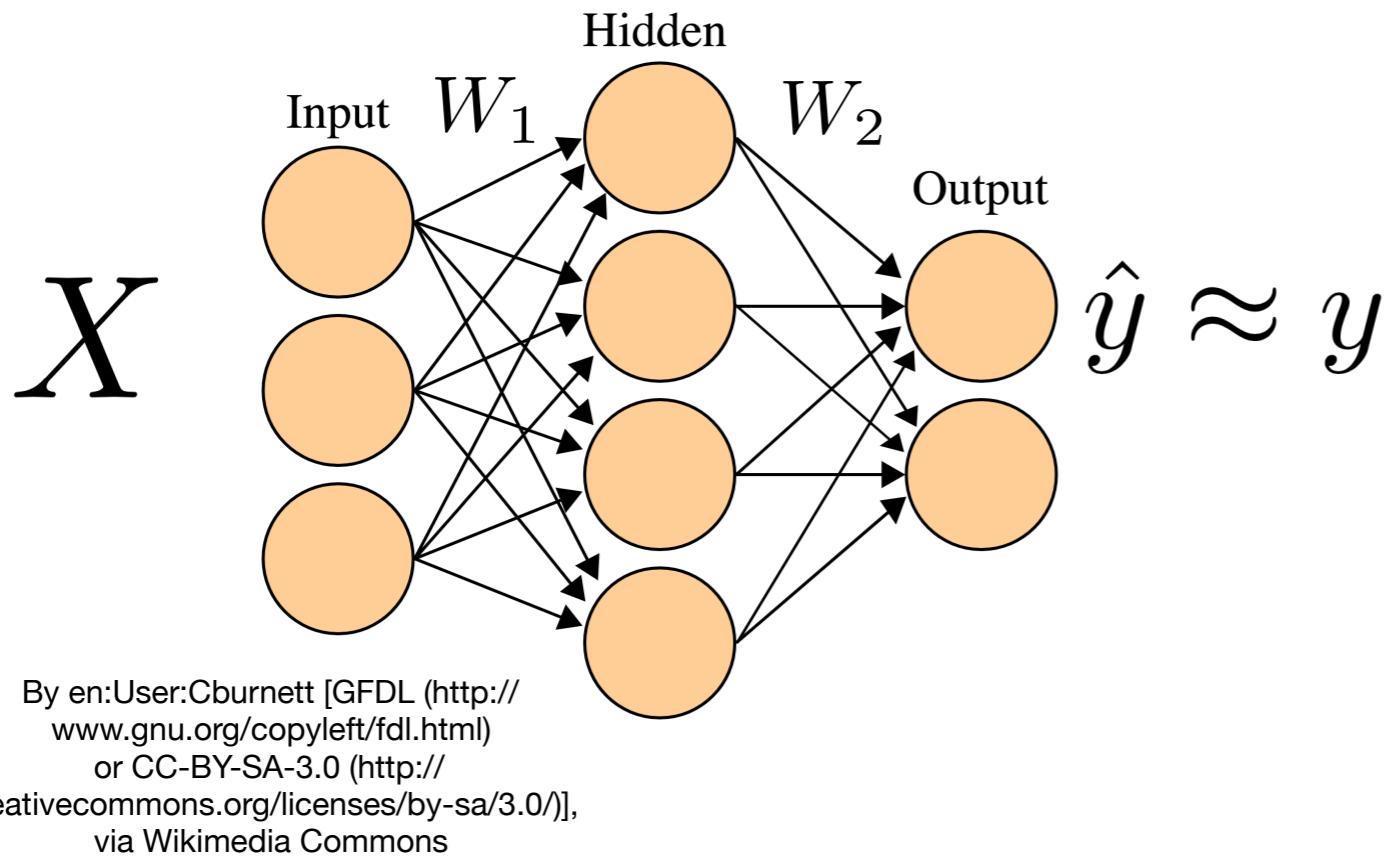
By en:User:Cburnett [GFDL (<http://www.gnu.org/copyleft/fdl.html>)  
or CC-BY-SA-3.0 (<http://creativecommons.org/licenses/by-sa/3.0/>)],  
via Wikimedia Commons

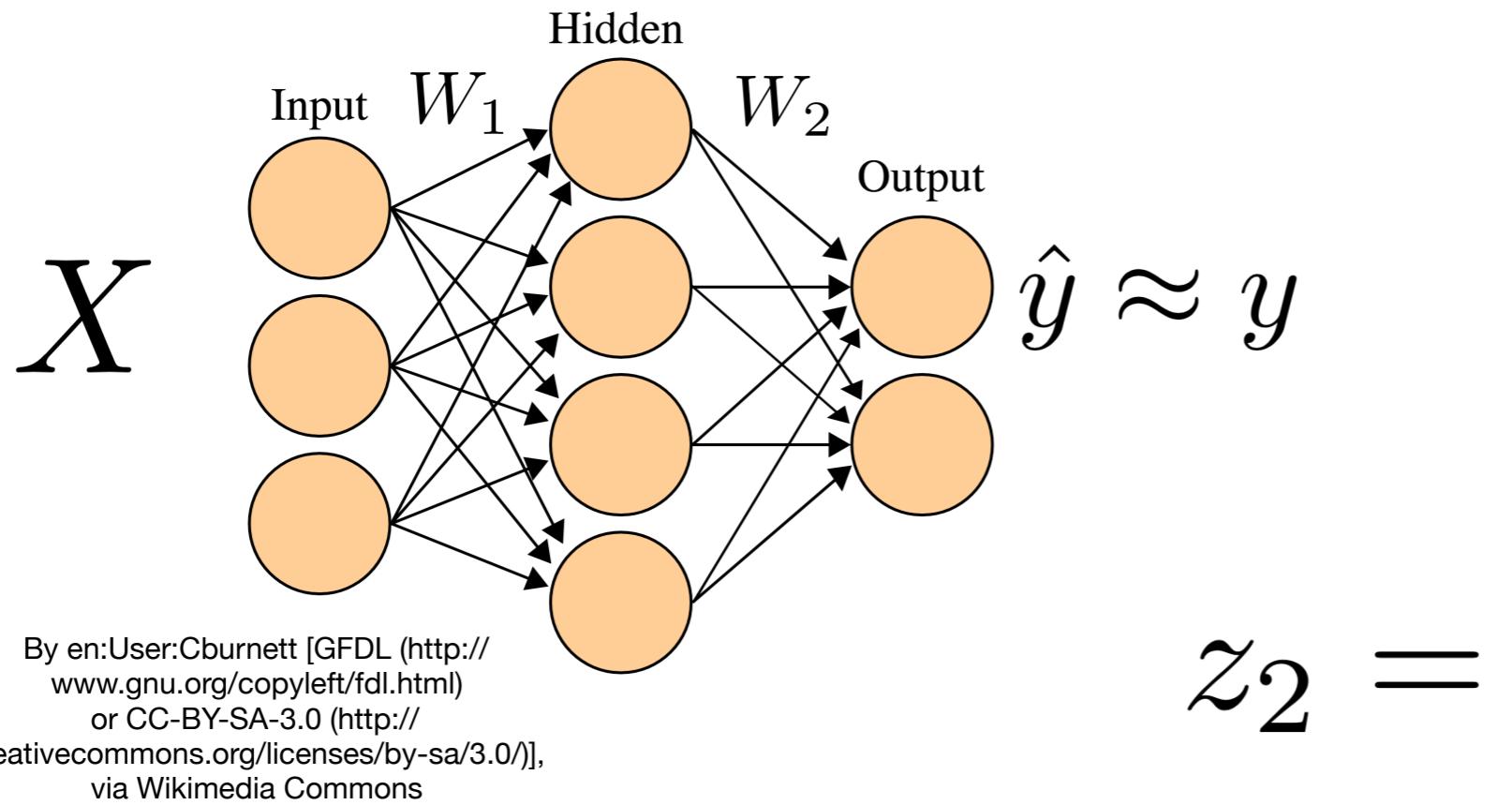


By en:User:Cburnett [GFDL (<http://www.gnu.org/copyleft/fdl.html>)  
or CC-BY-SA-3.0 (<http://creativecommons.org/licenses/by-sa/3.0/>)],  
via Wikimedia Commons

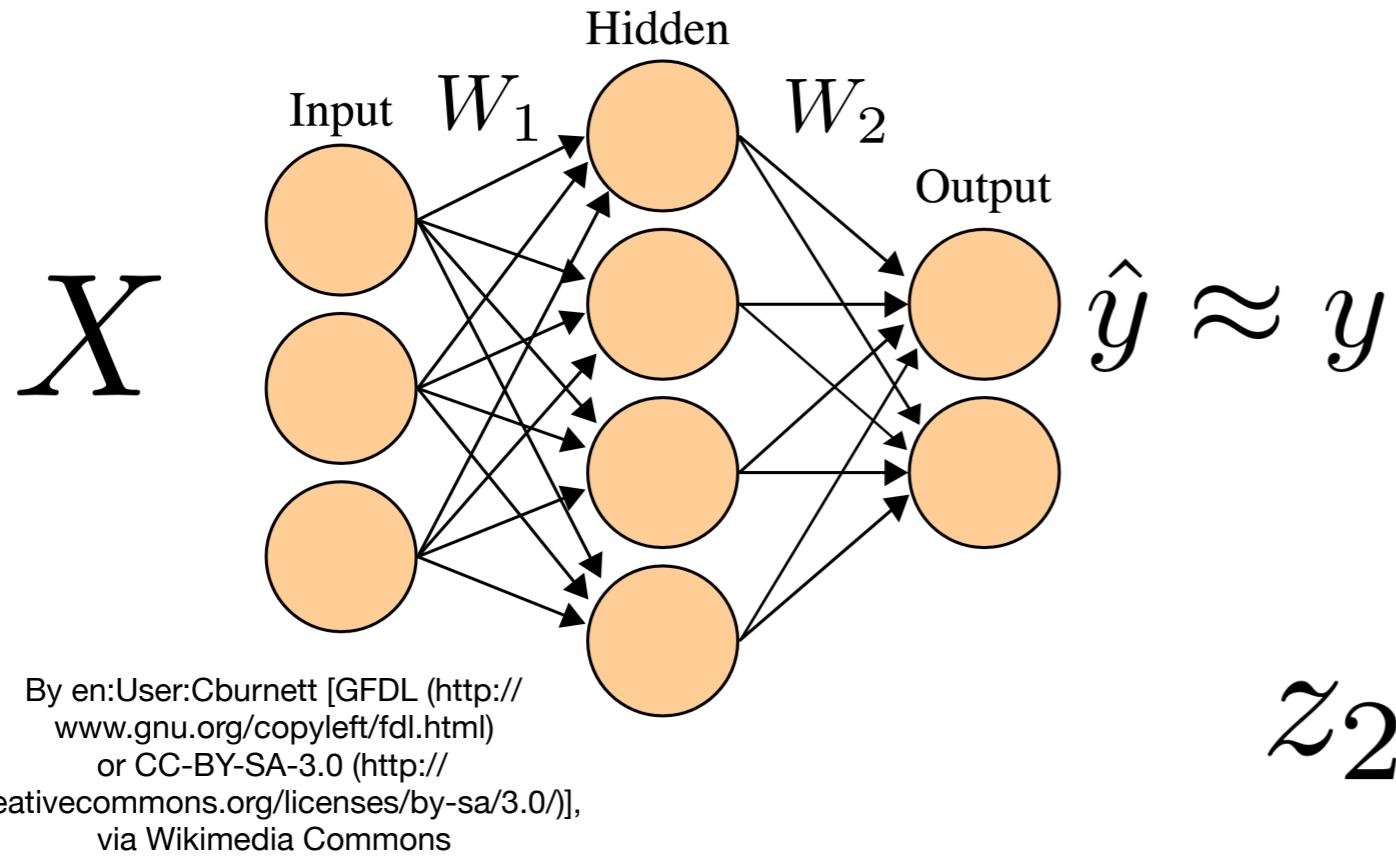






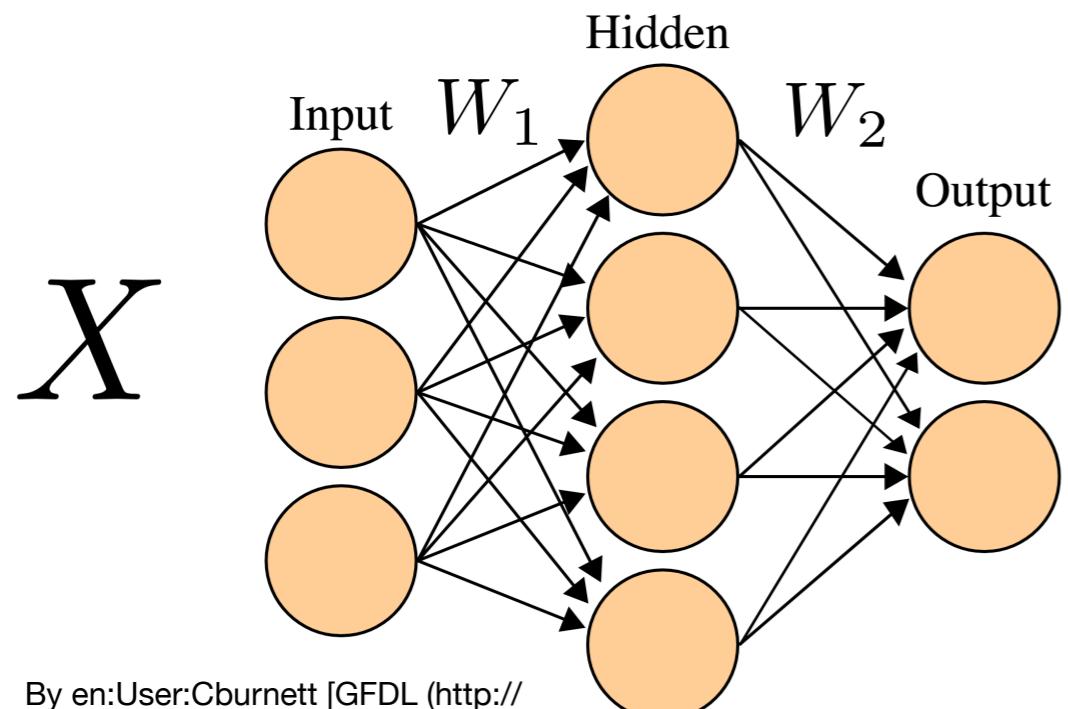


$$z_2 = XW_1$$



$$z_2 = XW_1$$

$$a_2 = f(z_2)$$



By en:User:Cburnett [GFDL (<http://www.gnu.org/copyleft/fdl.html>)  
or CC-BY-SA-3.0 (<http://creativecommons.org/licenses/by-sa/3.0/>)],  
via Wikimedia Commons

$$\hat{y} \approx y$$

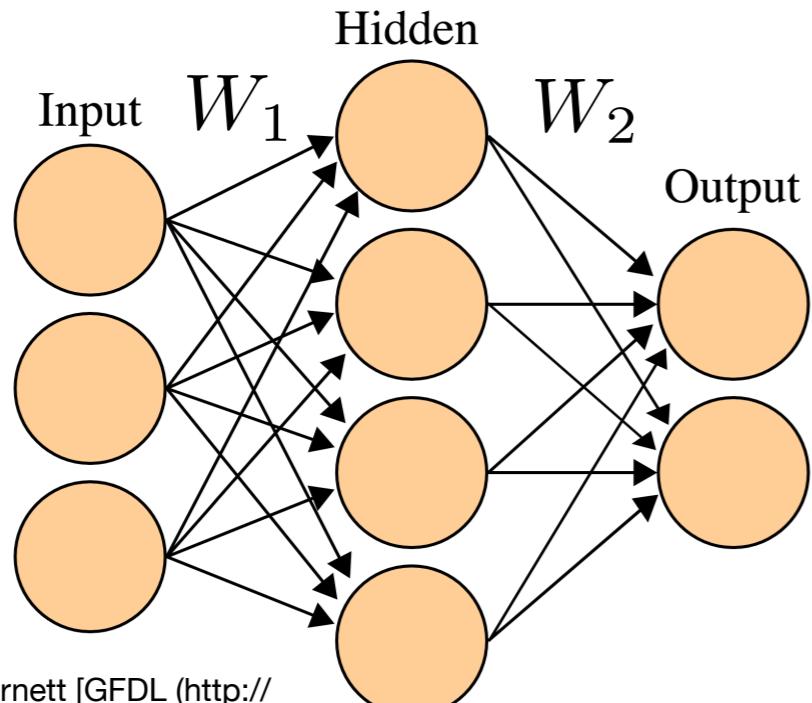
$$f(x) = \text{sigmoid}(x)$$

$$z_2 = XW_1$$

$$a_2 = f(z_2)$$

$$z_3 = a_2W_2$$

$X$



By en:User:Cburnett [GFDL (<http://www.gnu.org/copyleft/fdl.html>)  
or CC-BY-SA-3.0 (<http://creativecommons.org/licenses/by-sa/3.0/>)],  
via Wikimedia Commons

$$\hat{y} \approx y$$

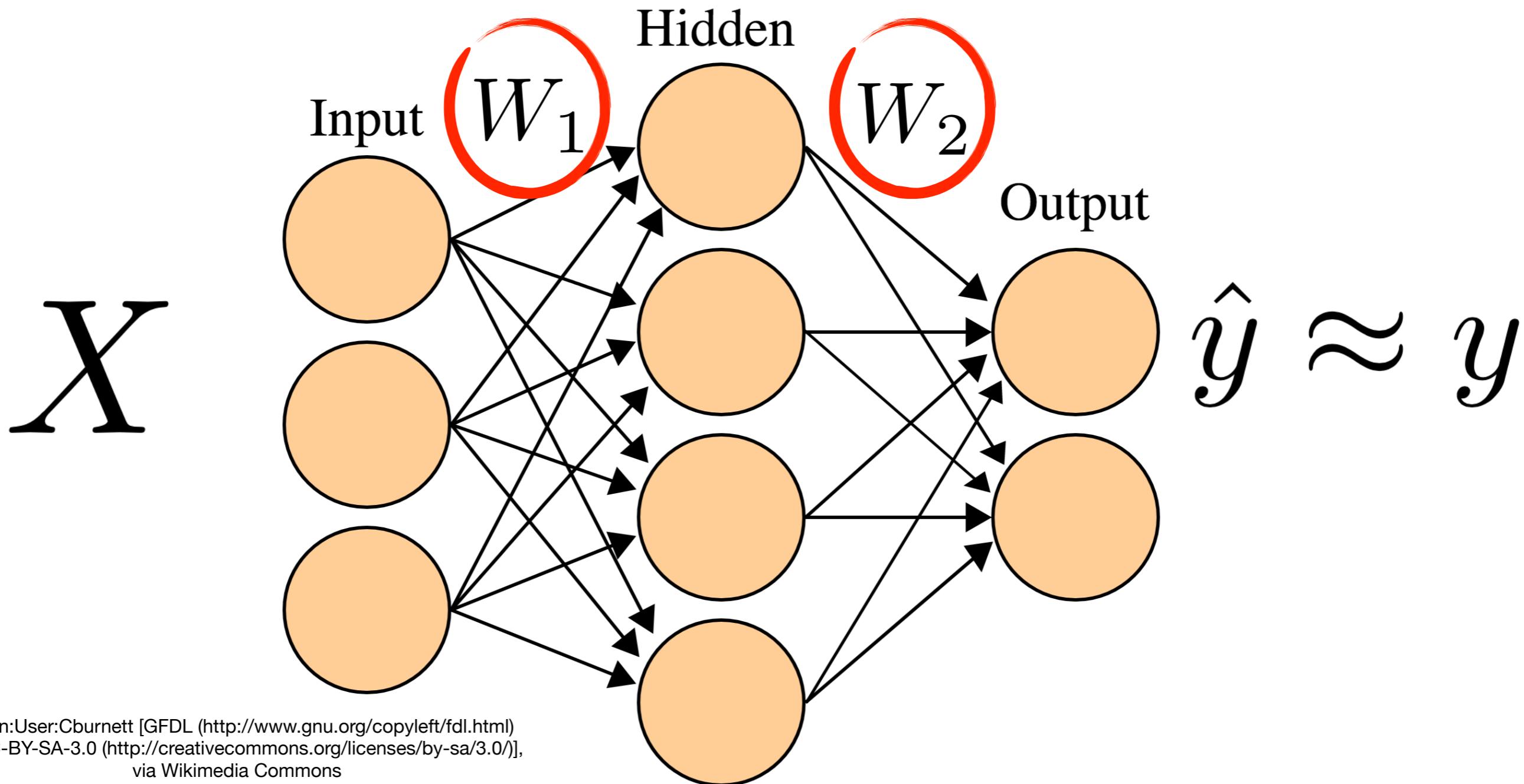
$$f(x) = \text{sigmoid}(x)$$

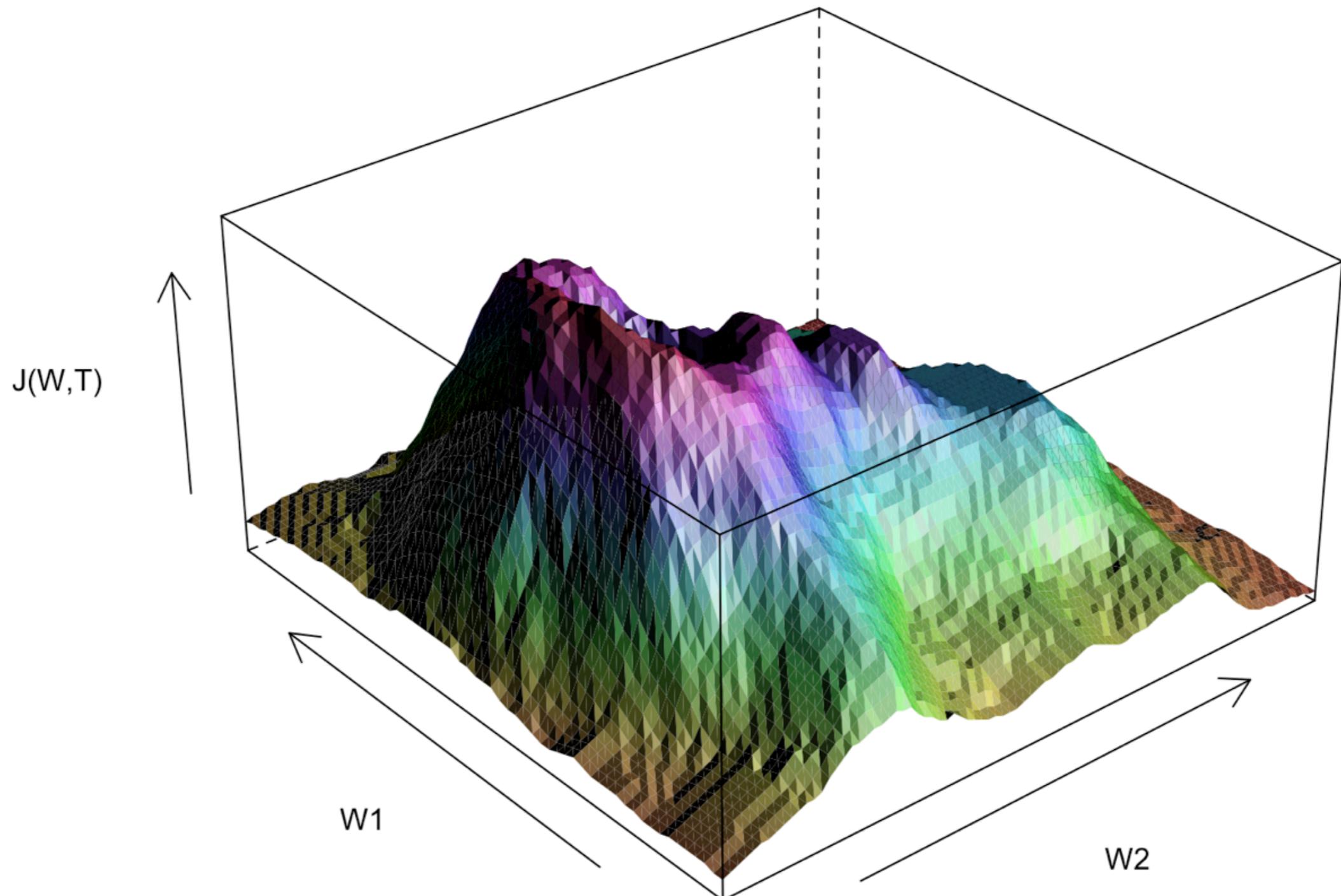
$$z_2 = XW_1$$

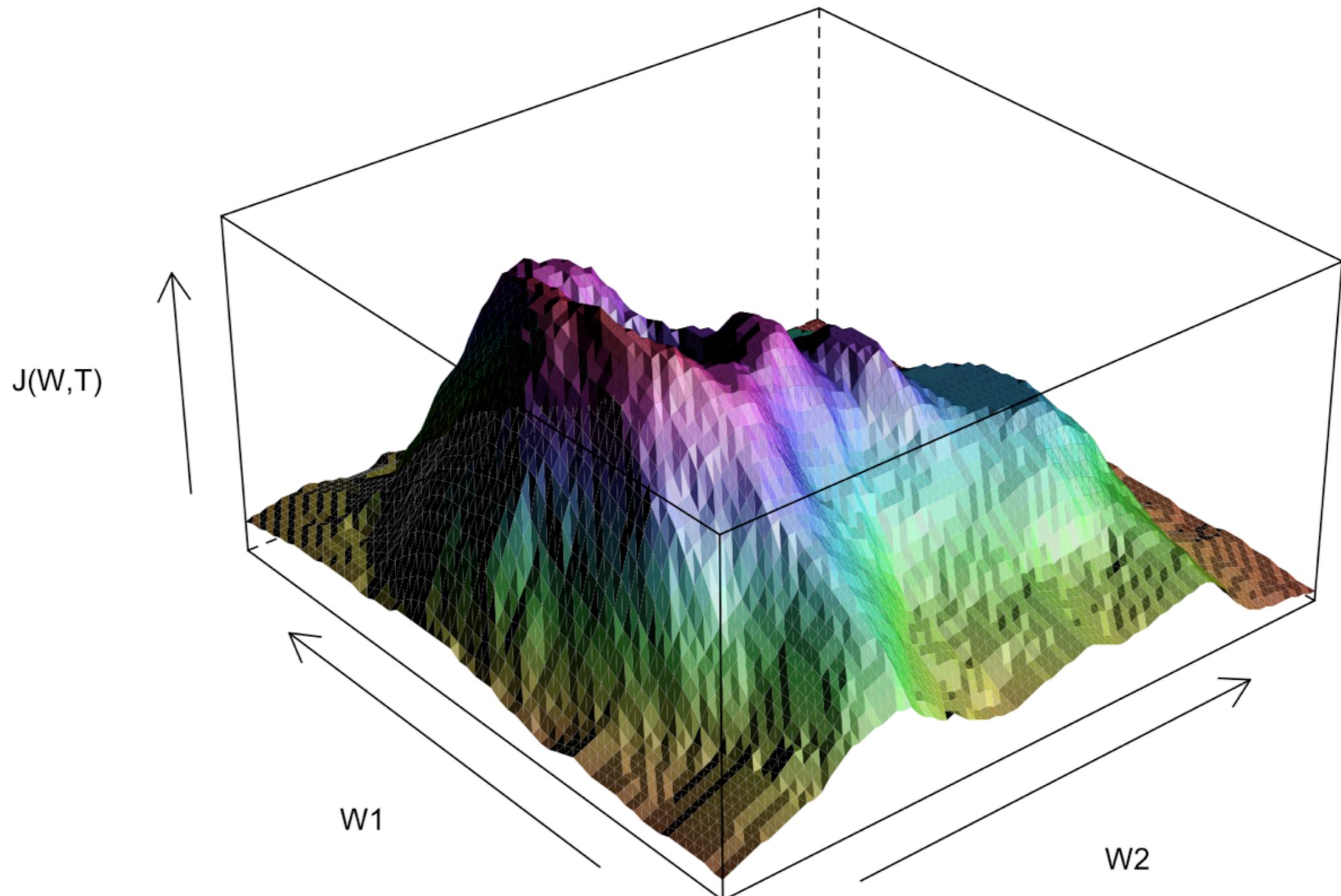
$$a_2 = f(z_2)$$

$$z_3 = a_2 W_2$$

$$\hat{y} = a_3 = f(z_3)$$







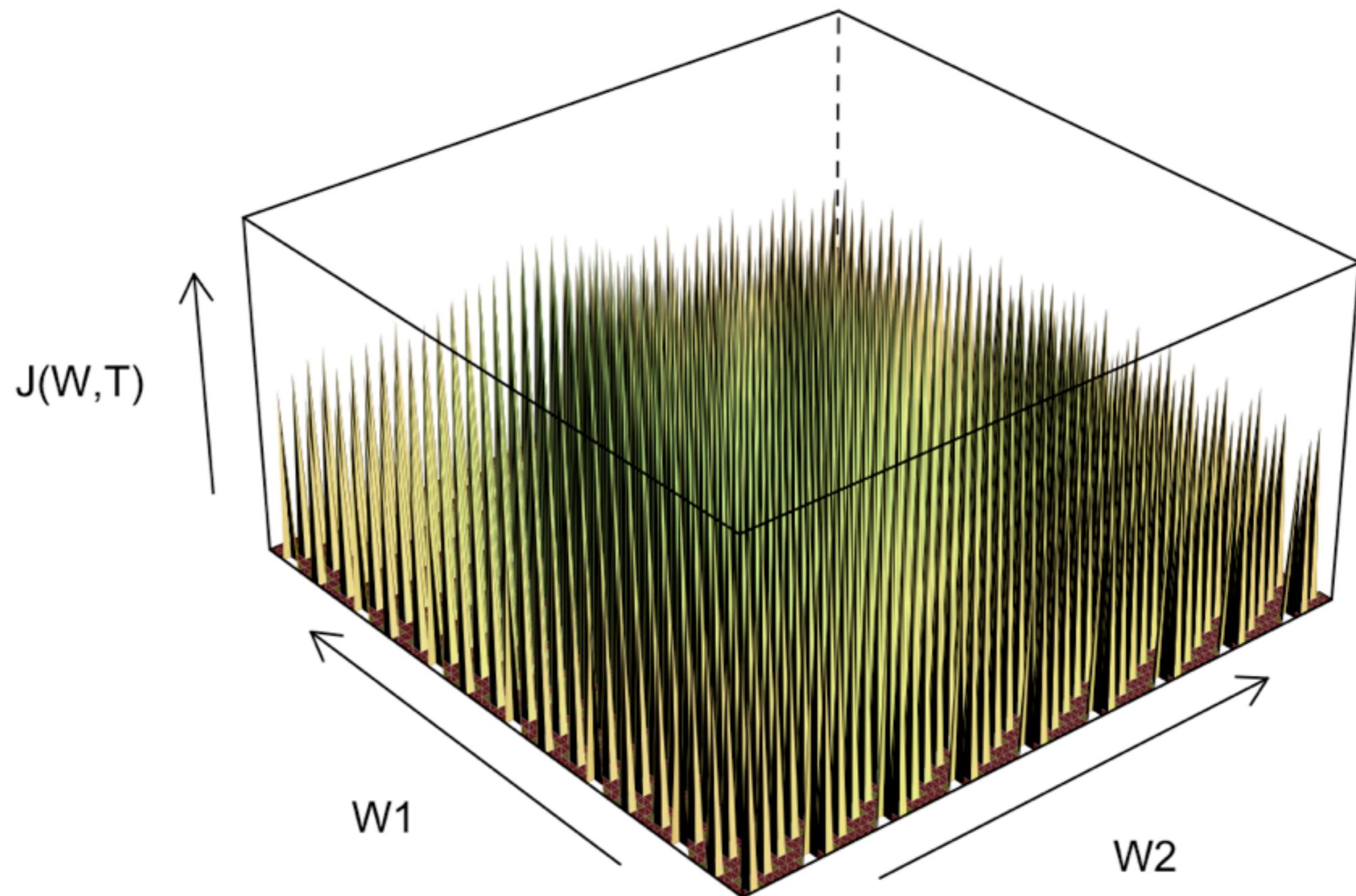
$v = [-1, 1]$

$v = [-1, 1]$

$grid = [-1, -0.9, -0.8, -0.7, -0.6, -0.5, -0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1]$

$v = [-1, 1]$

$grid = [-1, -0.9, -0.8, -0.7, -0.6, -0.5, -0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1]$



$v = [-1, 1]$

$grid = [-1, -0.9, -0.8, -0.7, -0.6, -0.5, -0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1]$

$dim(1) = 21$

$v = [-1, 1]$

$grid = [-1, -0.9, -0.8, -0.7, -0.6, -0.5, -0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1]$

$dim(1) = 21$

$dim(2) = 411$

$v = [-1, 1]$

$grid = [-1, -0.9, -0.8, -0.7, -0.6, -0.5, -0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1]$

$dim(1) = 21$

$dim(2) = 411$

$dim(3) = 9261$

$v = [-1, 1]$

$grid = [-1, -0.9, -0.8, -0.7, -0.6, -0.5, -0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1]$

$dim(1) = 21$

$dim(2) = 411$

$dim(3) = 9261$

$dim(4) = 194481$

$v = [-1, 1]$

$grid = [-1, -0.9, -0.8, -0.7, -0.6, -0.5, -0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1]$

$dim(1) = 21$

$dim(2) = 411$

$dim(3) = 9261$

$dim(4) = 194481$

$dim(5) = 4084101$

$v = [-1, 1]$

$grid = [-1, -0.9, -0.8, -0.7, -0.6, -0.5, -0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1]$

$dim(1) = 21$

$dim(2) = 411$

$dim(3) = 9261$

$dim(4) = 194481$

$dim(5) = 4084101$

$dim(6) = 85766121$

$v = [-1, 1]$

$grid = [-1, -0.9, -0.8, -0.7, -0.6, -0.5, -0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1]$

$dim(1) = 21$

$dim(2) = 411$

$dim(3) = 9261$

$dim(4) = 194481$

$dim(5) = 4084101$

$dim(6) = 85766121$

$dim(7) = 1801088541$

$v = [-1, 1]$

$grid = [-1, -0.9, -0.8, -0.7, -0.6, -0.5, -0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1]$

$dim(1) = 21$

$dim(2) = 411$

$dim(3) = 9261$

$dim(4) = 194481$

$dim(5) = 4084101$

$dim(6) = 85766121$

$dim(7) = 1801088541$

$dim(8) = 37822859361$

$v = [-1, 1]$

$grid = [-1, -0.9, -0.8, -0.7, -0.6, -0.5, -0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1]$

$dim(1) = 21$

$dim(2) = 411$

$dim(3) = 9261$

$dim(4) = 194481$

$dim(5) = 4084101$

$dim(6) = 85766121$

$dim(7) = 1801088541$

$dim(8) = 37822859361$

$dim(9) = 794280046581$

$v = [-1, 1]$

$grid = [-1, -0.9, -0.8, -0.7, -0.6, -0.5, -0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1]$

$dim(1) = 21$

$dim(2) = 411$

$dim(3) = 9261$

$dim(4) = 194481$

$dim(5) = 4084101$

$dim(6) = 85766121$

$dim(7) = 1801088541$

$dim(8) = 37822859361$

$dim(9) = 794280046581$

$dim(10) = 16679880978201$

$v = [-1, 1]$

$grid = [-1, -0.9, -0.8, -0.7, -0.6, -0.5, -0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1]$

$dim(1) = 21$

$dim(2) = 411$

$dim(3) = 9261$

$dim(4) = 194481$

$dim(5) = 4084101$

$dim(6) = 85766121$

$dim(7) = 1801088541$

$dim(8) = 37822859361$

$dim(9) = 794280046581$

$dim(10) = 16679880978201$

$dim(11) = 350277500542221$

$v = [-1, 1]$

$grid = [-1, -0.9, -0.8, -0.7, -0.6, -0.5, -0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1]$

$dim(1) = 21$

$dim(2) = 411$

$dim(3) = 9261$

$dim(4) = 194481$

$dim(5) = 4084101$

$dim(6) = 85766121$

$dim(7) = 1801088541$

$dim(8) = 37822859361$

$dim(9) = 794280046581$

$dim(10) = 16679880978201$

$dim(11) = 350277500542221$

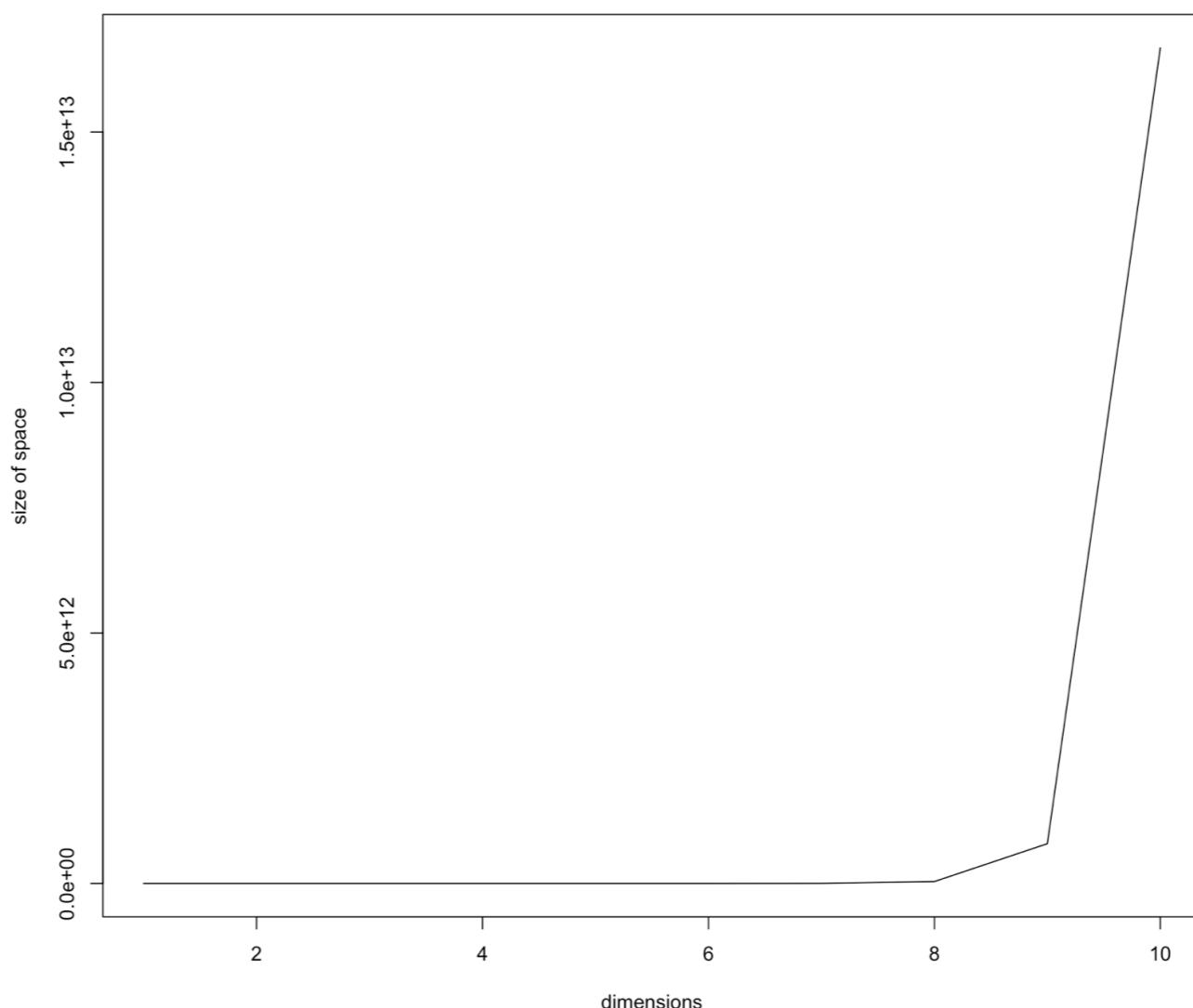
$dim(12) = 7355827511386641$

```
v = [-1, 1]
grid = [-1, -0.9, -0.8, -0.7, -0.6, -0.5, -0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1]
dim(1) = 21
dim(2) = 411
dim(3) = 9261
dim(4) = 194481
dim(5) = 4084101
dim(6) = 85766121
dim(7) = 1801088541
dim(8) = 37822859361
dim(9) = 794280046581
dim(10) = 16679880978201
dim(11) = 350277500542221
dim(12) = 7355827511386641
dim(13) = 1.544723777391195e17
```

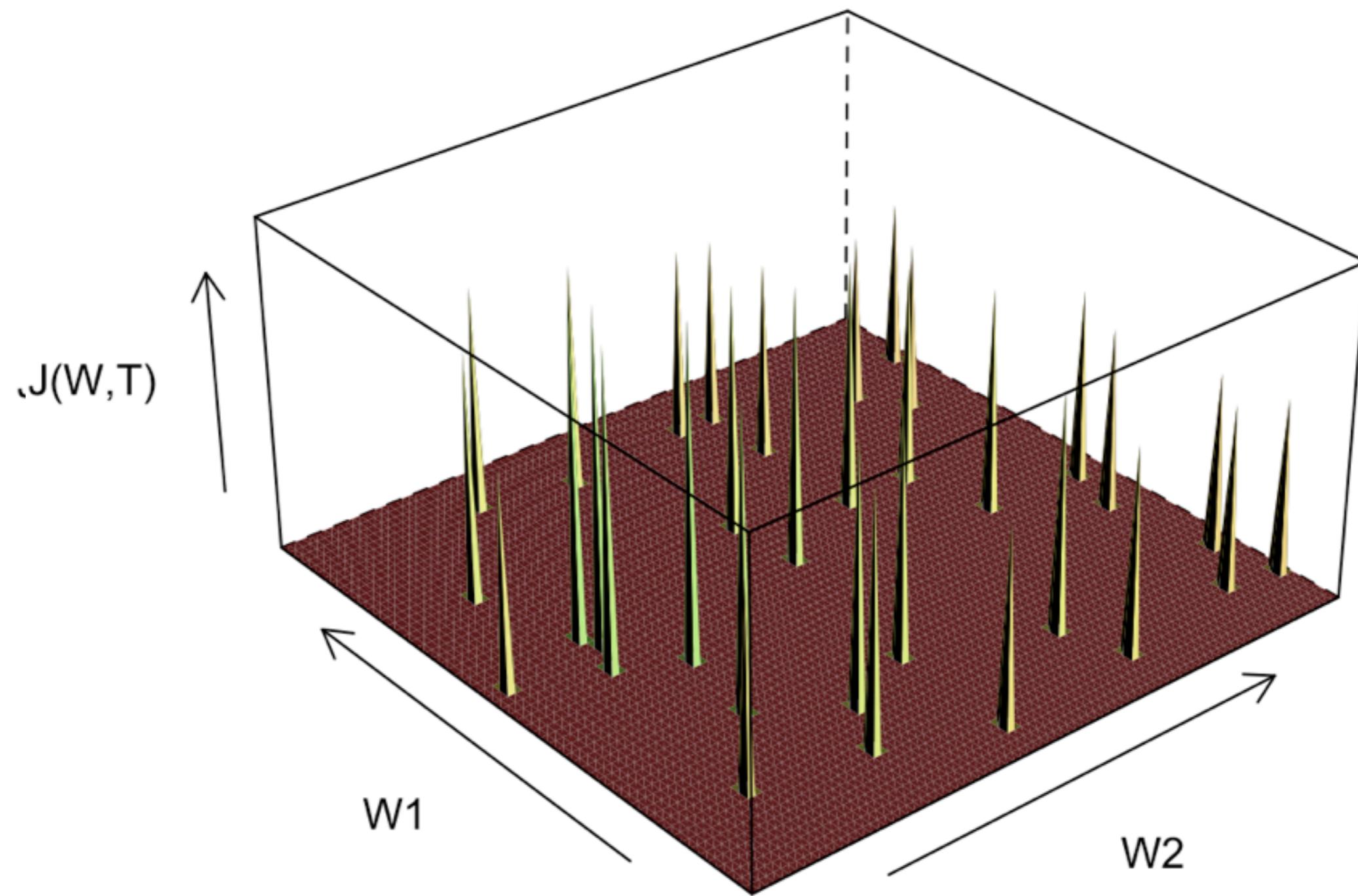
```
v = [-1, 1]
grid = [-1, -0.9, -0.8, -0.7, -0.6, -0.5, -0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1]
dim(1) = 21
dim(2) = 411
dim(3) = 9261
dim(4) = 194481
dim(5) = 4084101
dim(6) = 85766121
dim(7) = 1801088541
dim(8) = 37822859361
dim(9) = 794280046581
dim(10) = 16679880978201
dim(11) = 350277500542221
dim(12) = 7355827511386641
dim(13) = 1.544723777391195e17
dim(14) = 3.243919932521508e18
```

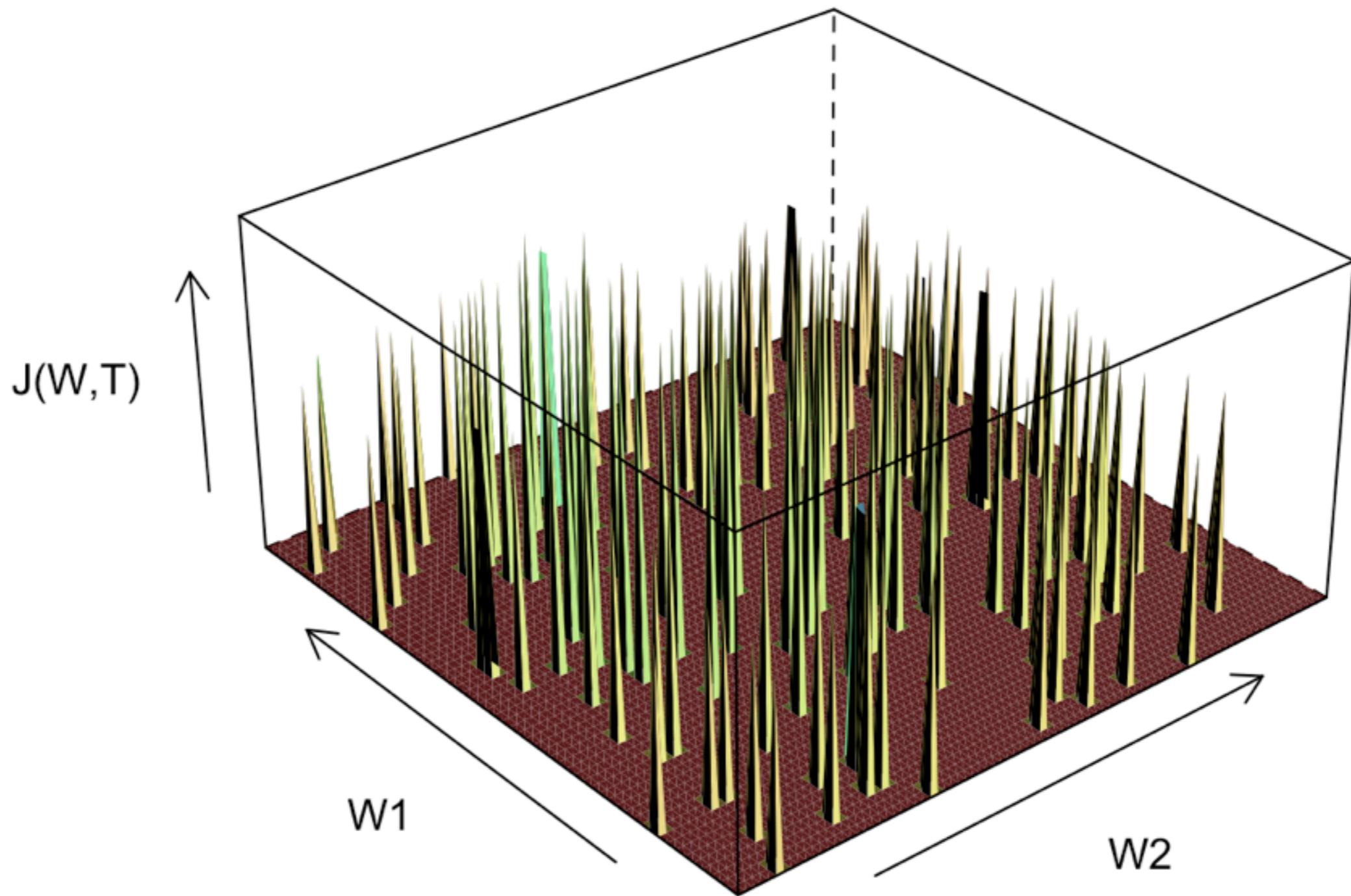
```
v = [-1, 1]
grid = [-1, -0.9, -0.8, -0.7, -0.6, -0.5, -0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1]
dim(1) = 21
dim(2) = 411
dim(3) = 9261
dim(4) = 194481
dim(5) = 4084101
dim(6) = 85766121
dim(7) = 1801088541
dim(8) = 37822859361
dim(9) = 794280046581
dim(10) = 16679880978201
dim(11) = 350277500542221
dim(12) = 7355827511386641
dim(13) = 1.544723777391195e17
dim(14) = 3.243919932521508e18
dim(15) = 6.812231858295167e19
```

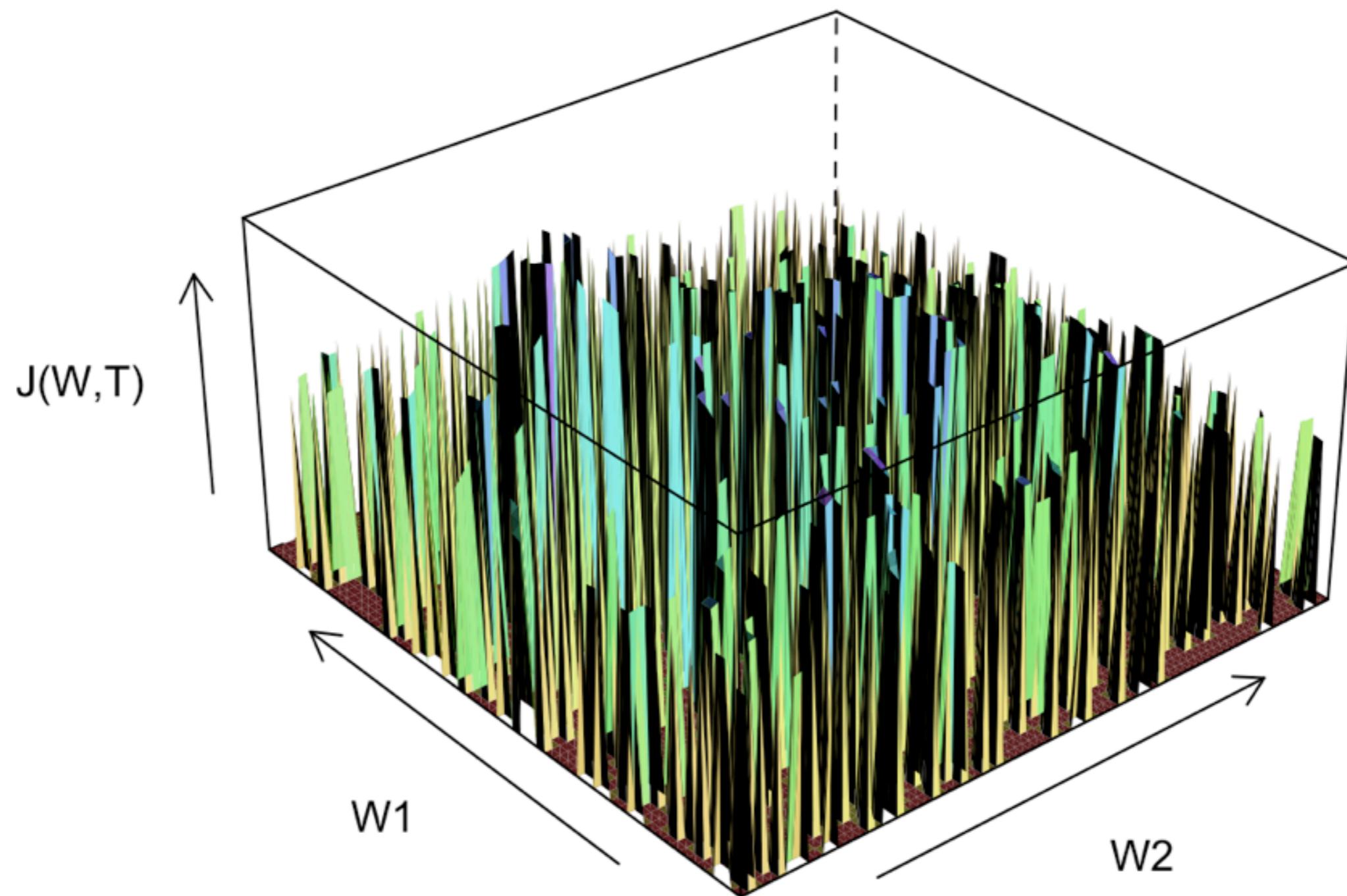
$v = [-1, 1]$   
 $grid = [-1, -0.9, -0.8, -0.7, -0.6, -0.5, -0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1]$   
 $dim(1) = 21$   
 $dim(2) = 411$   
 $dim(3) = 9261$   
 $dim(4) = 194481$   
 $dim(5) = 4084101$   
 $dim(6) = 85766121$   
 $dim(7) = 1801088541$   
 $dim(8) = 37822859361$   
 $dim(9) = 794280046581$   
 $dim(10) = 16679880978201$   
 $dim(11) = 350277500542221$   
 $dim(12) = 7355827511386641$   
 $dim(13) = 1.544723777391195e17$   
 $dim(14) = 3.243919932521508e18$   
 $dim(15) = 6.812231858295167e19$   
 $dim(16) = 1.430568690241985e21$

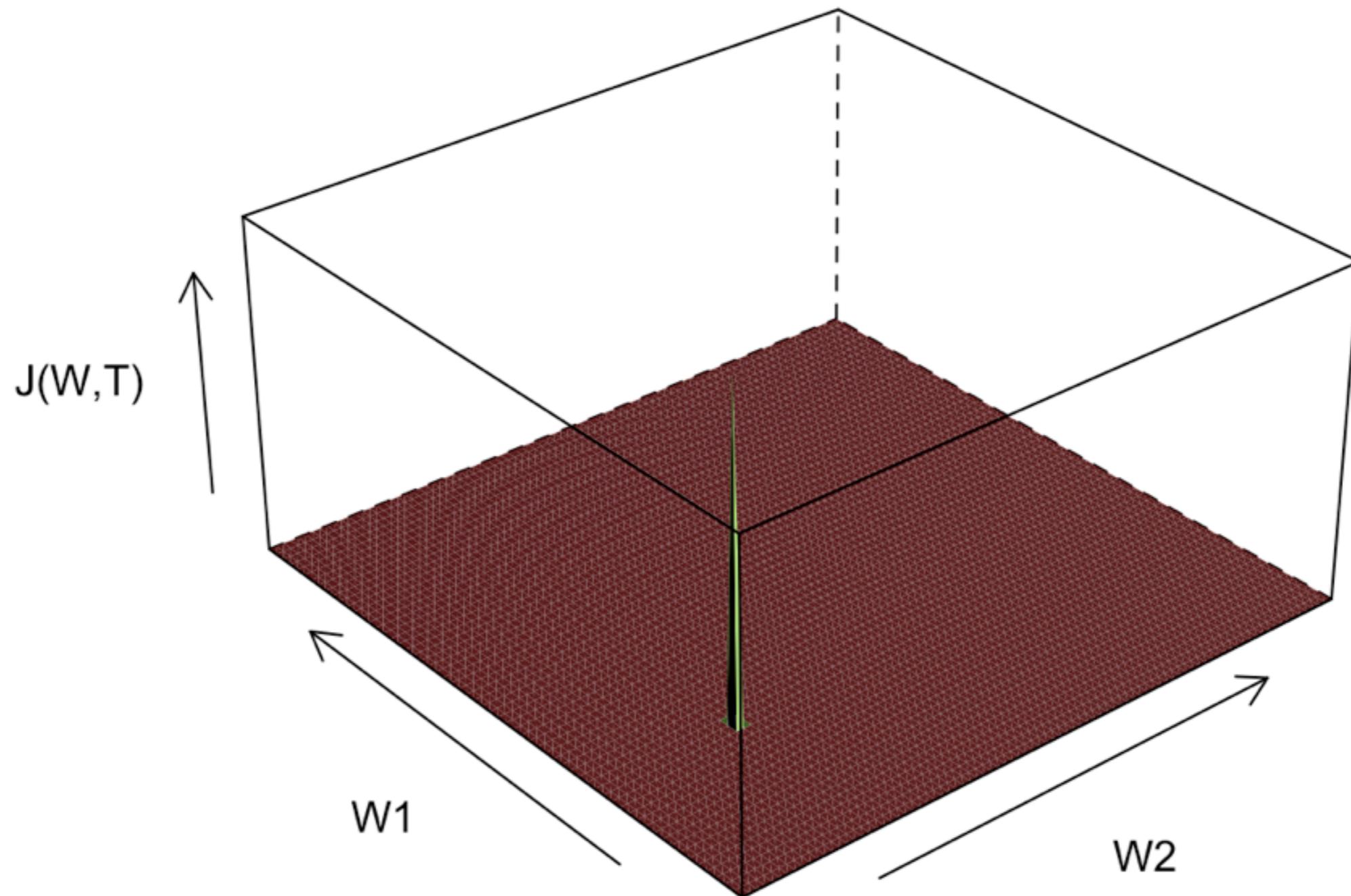


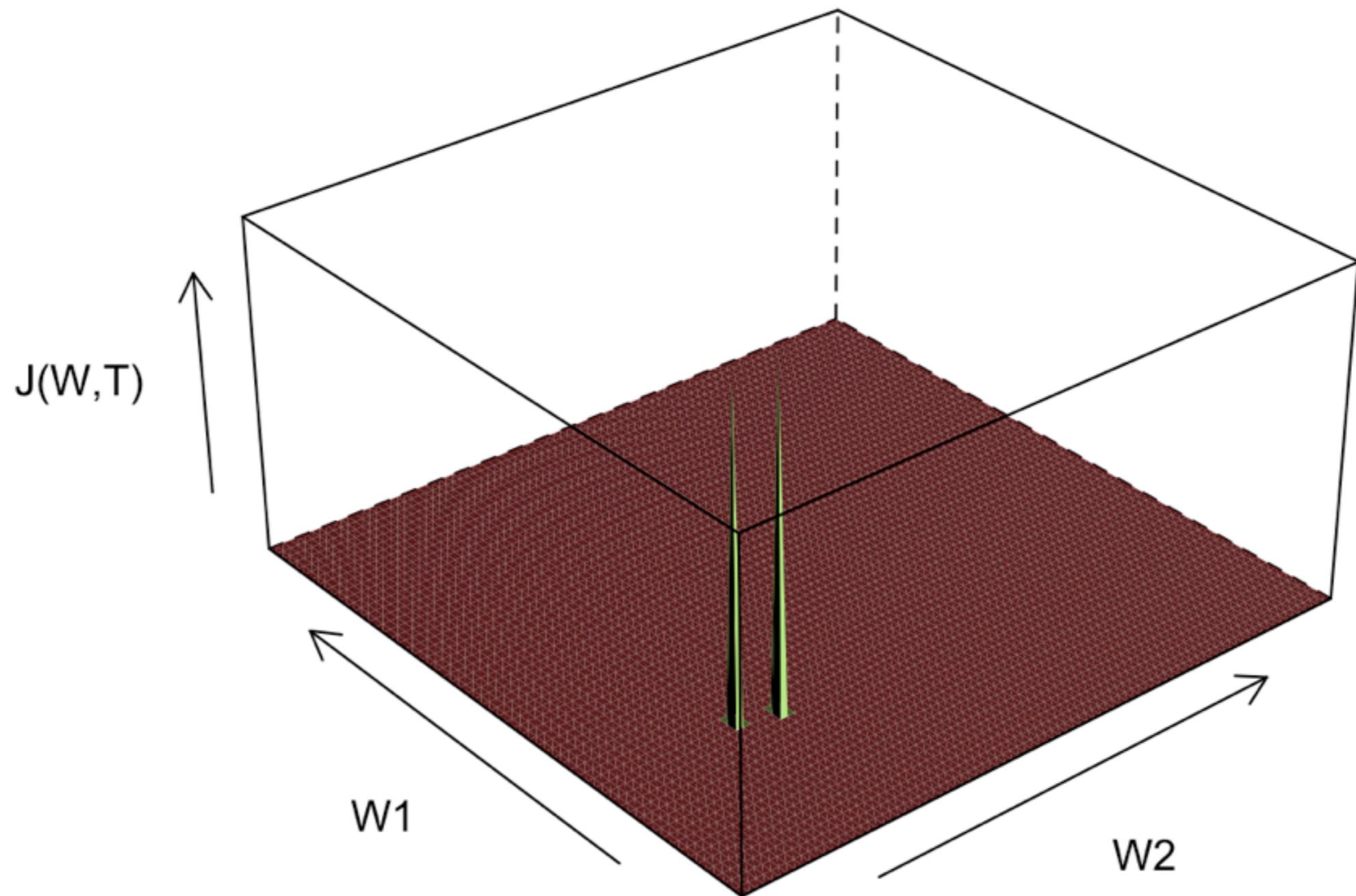
$$\dim(100) = 21^{100} = 1.66697648439634e132$$

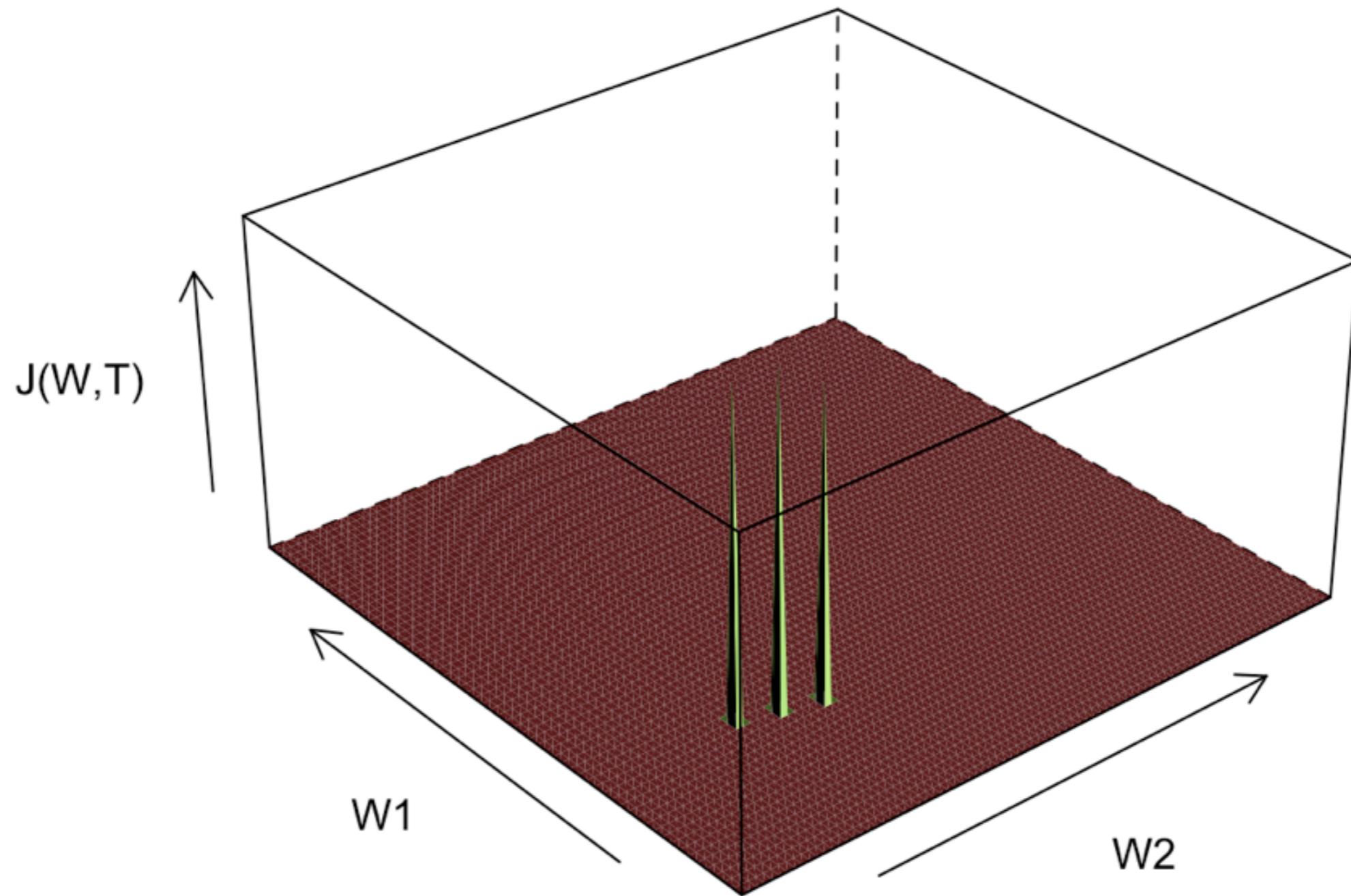


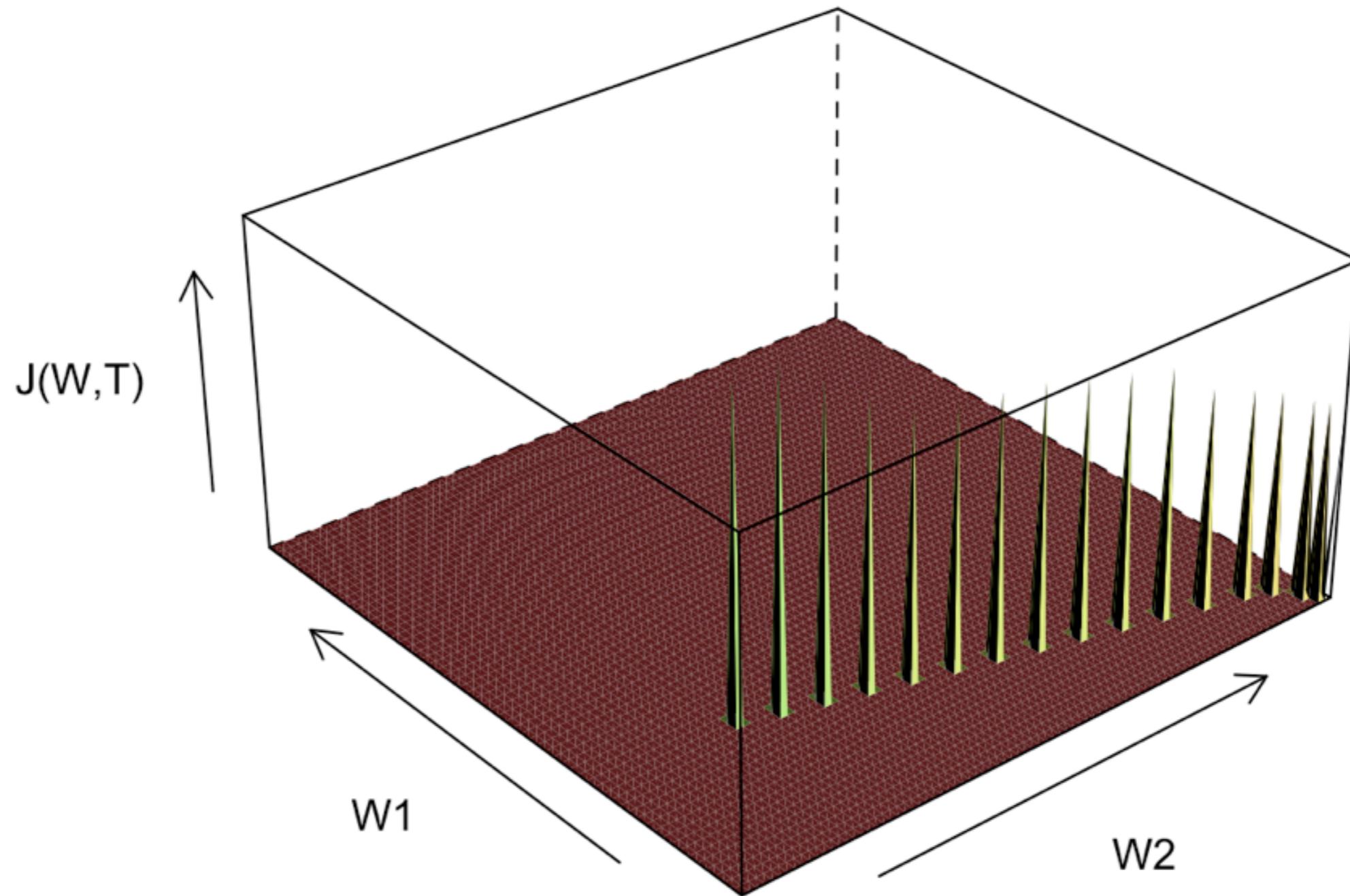












$$J = \sum \frac{1}{2} (y - \hat{y})^2$$

$$J = \sum \frac{1}{2} (y - \hat{y})^2$$

# cross entropy

$$J = \sum \frac{1}{2} (y - \hat{y})^2$$

**cross entropy  
exponential cost**

$$J = \sum \frac{1}{2} (y - \hat{y})^2$$

**cross entropy**

**exponential cost**

**hellinger distance**

$$J = \sum \frac{1}{2} (y - \hat{y})^2$$

$$\hat{y} = a_3 = f(z_3)$$

$$J = \sum \frac{1}{2} (y - \hat{y})^2$$

$$\hat{y} = a_3 = f(z_3)$$

$$\hat{y} = a_3 = f(a_2 W_2)$$

$$J = \sum \frac{1}{2} (y - \hat{y})^2$$

$$\hat{y} = a_3 = f(z_3)$$

$$\hat{y} = a_3 = f(a_2 W_2)$$

$$\hat{y} = a_3 = f(f(z_2) W_2)$$

$$J = \sum \frac{1}{2} (y - \hat{y})^2$$

$$\hat{y} = a_3 = f(z_3)$$

$$\hat{y} = a_3 = f(a_2 W_2)$$

$$\hat{y} = a_3 = f(f(z_2) W_2)$$

$$\hat{y} = a_3 = f(f(XW_1) W_2)$$

$$J = \sum \frac{1}{2} (y - \hat{y})^2$$

$$\hat{y} = a_3 = f(z_3)$$

$$\hat{y} = a_3 = f(a_2 W_2)$$

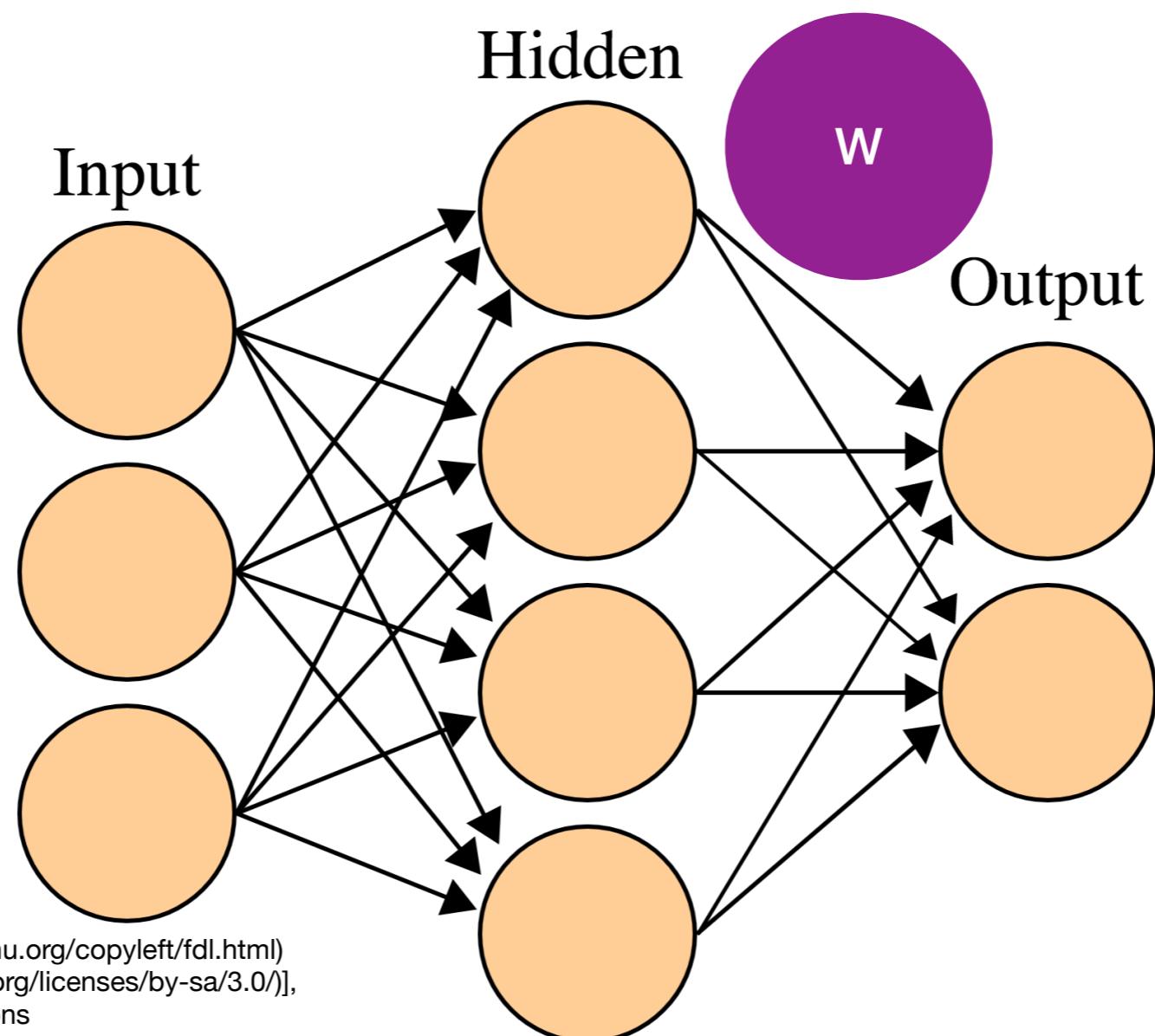
$$\hat{y} = a_3 = f(f(z_2) W_2)$$

$$\hat{y} = a_3 = f(f(XW_1) W_2)$$

$$J = \sum \frac{1}{2} (y - f(f(XW_1) W_2))^2$$

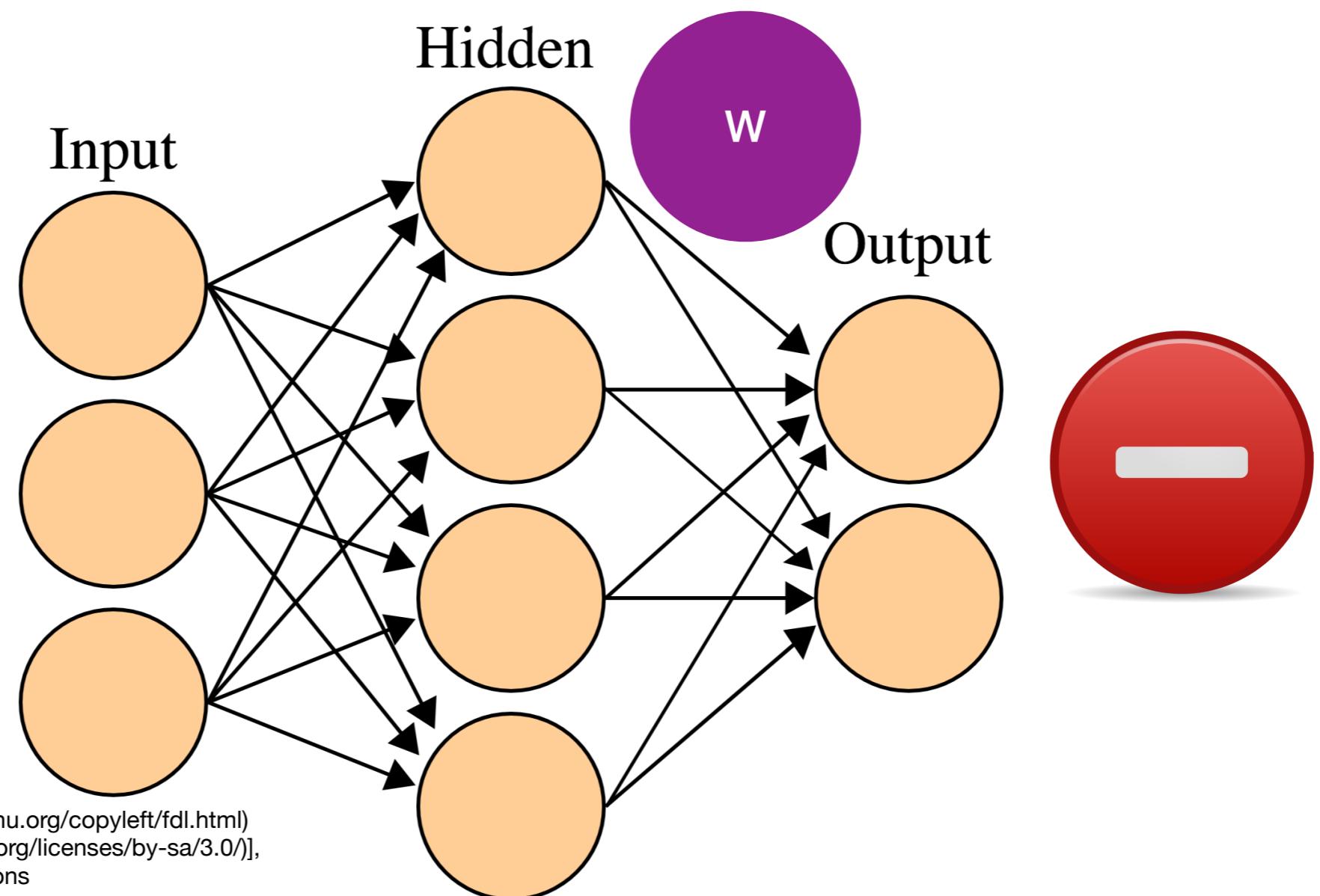
$$\sigma_3 = -(y - \hat{y}) f'(z_3)$$

$$\sigma_3 = -(y - \hat{y}) f'(z_3)$$



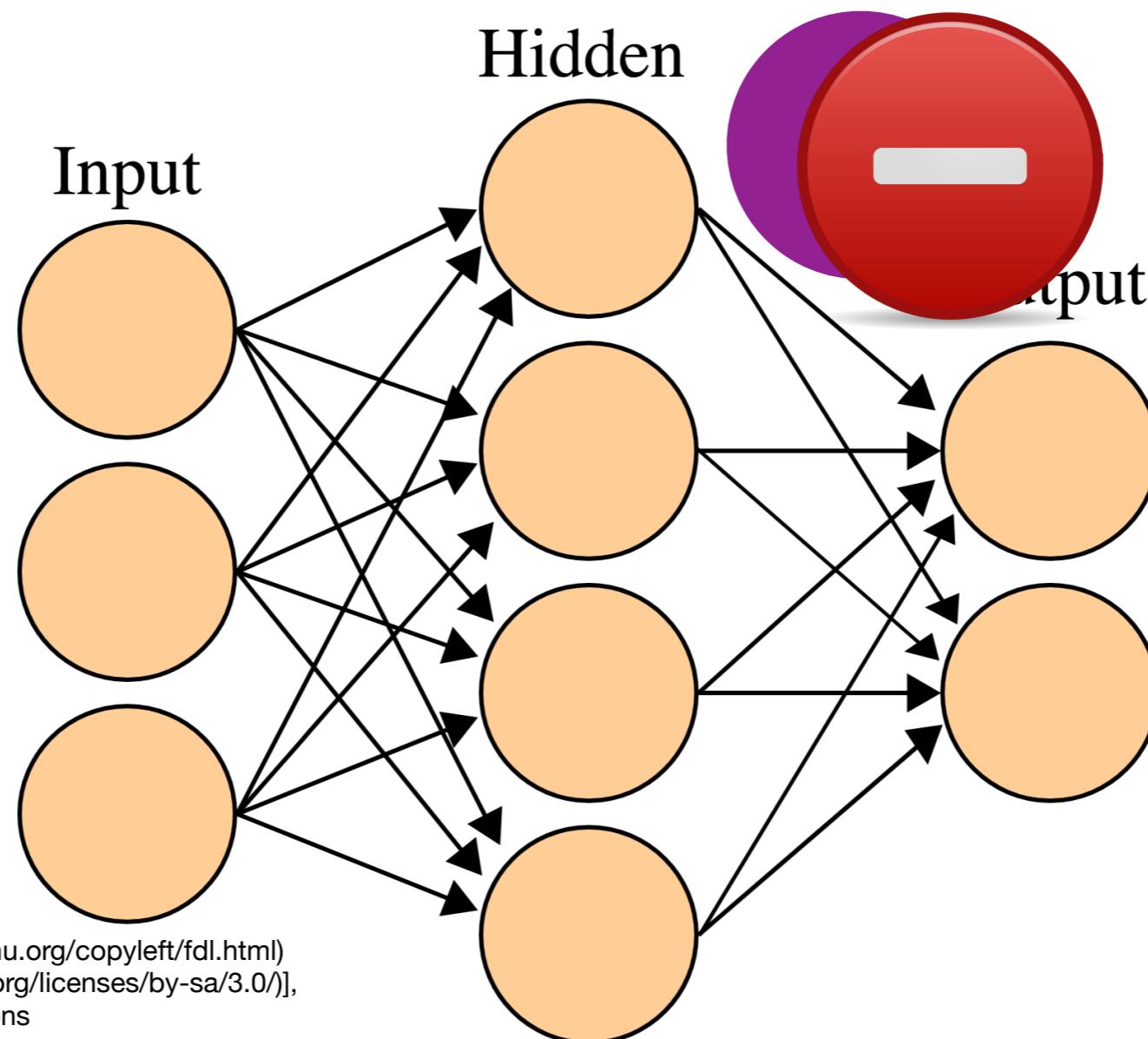
By en:User:Cburnett [GFDL (<http://www.gnu.org/copyleft/fdl.html>)  
or CC-BY-SA-3.0 (<http://creativecommons.org/licenses/by-sa/3.0/>)],  
via Wikimedia Commons

$$\sigma_3 = -(y - \hat{y}) f'(z_3)$$



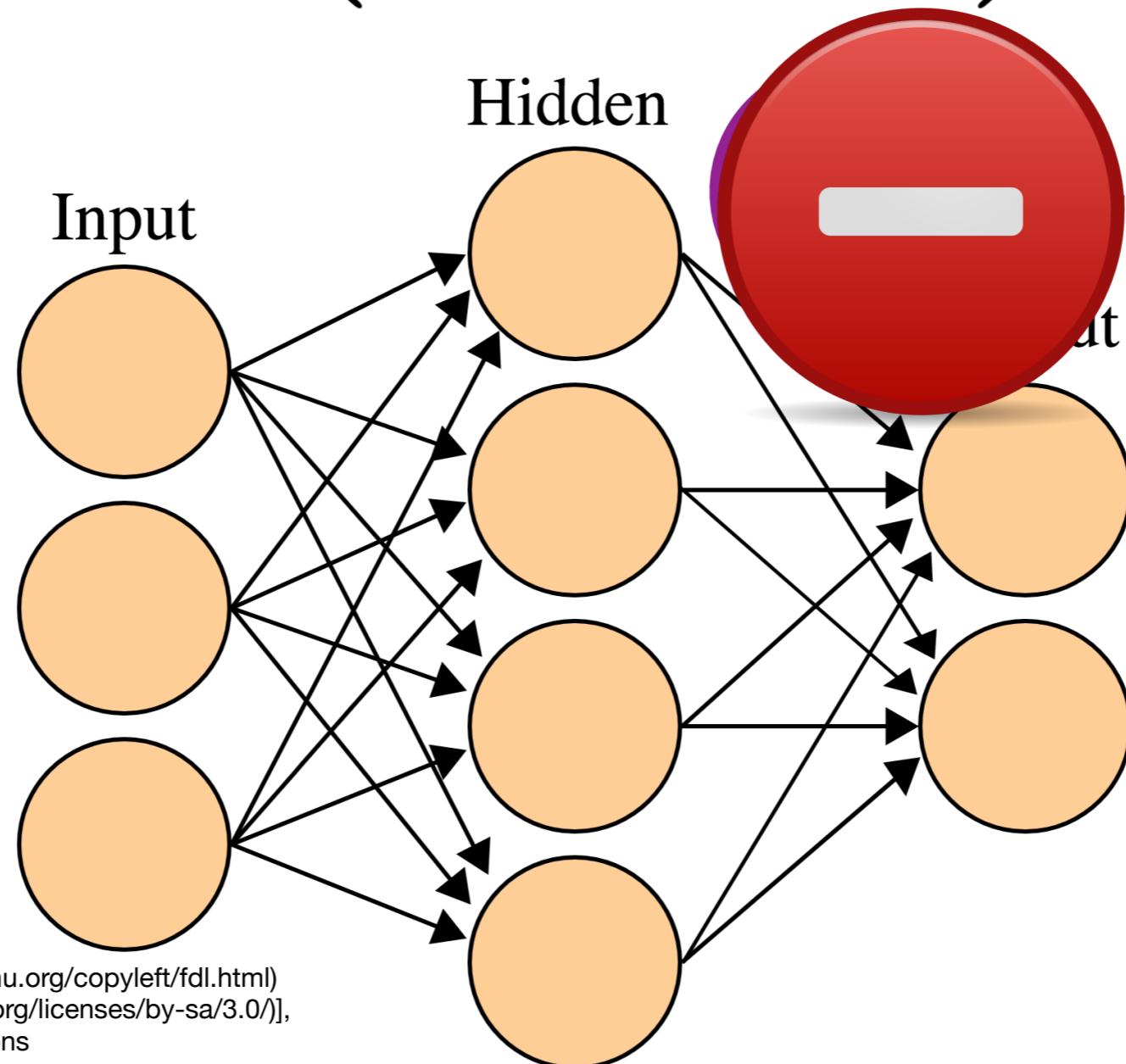
By en:User:Cburnett [GFDL (<http://www.gnu.org/copyleft/fdl.html>)  
or CC-BY-SA-3.0 (<http://creativecommons.org/licenses/by-sa/3.0/>)],  
via Wikimedia Commons

$$\sigma_3 = -(y - \hat{y}) f'(z_3)$$



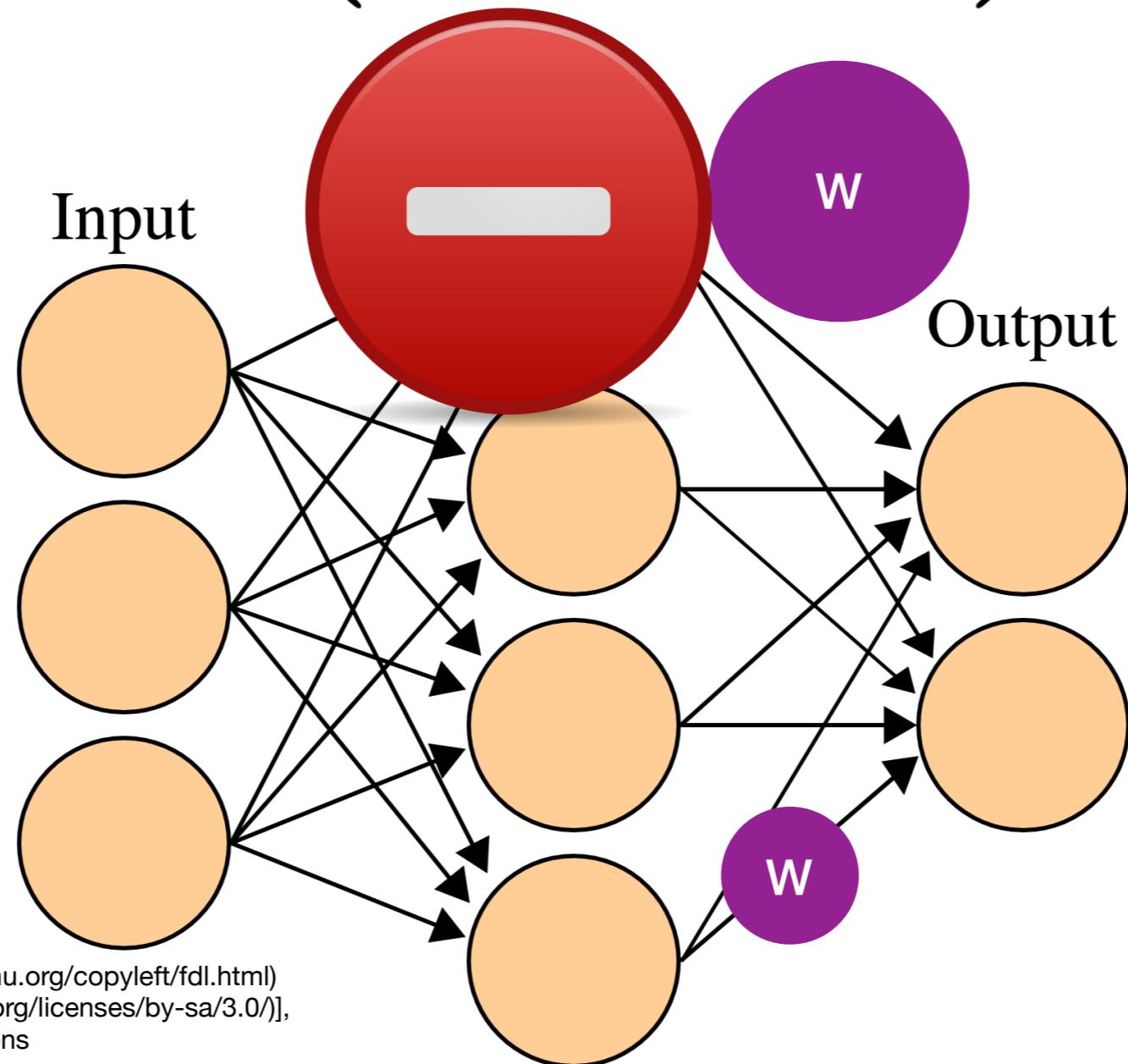
By en:User:Cburnett [GFDL (<http://www.gnu.org/copyleft/fdl.html>)  
or CC-BY-SA-3.0 (<http://creativecommons.org/licenses/by-sa/3.0/>)],  
via Wikimedia Commons

$$\sigma_3 = -(y - \hat{y}) f'(z_3)$$



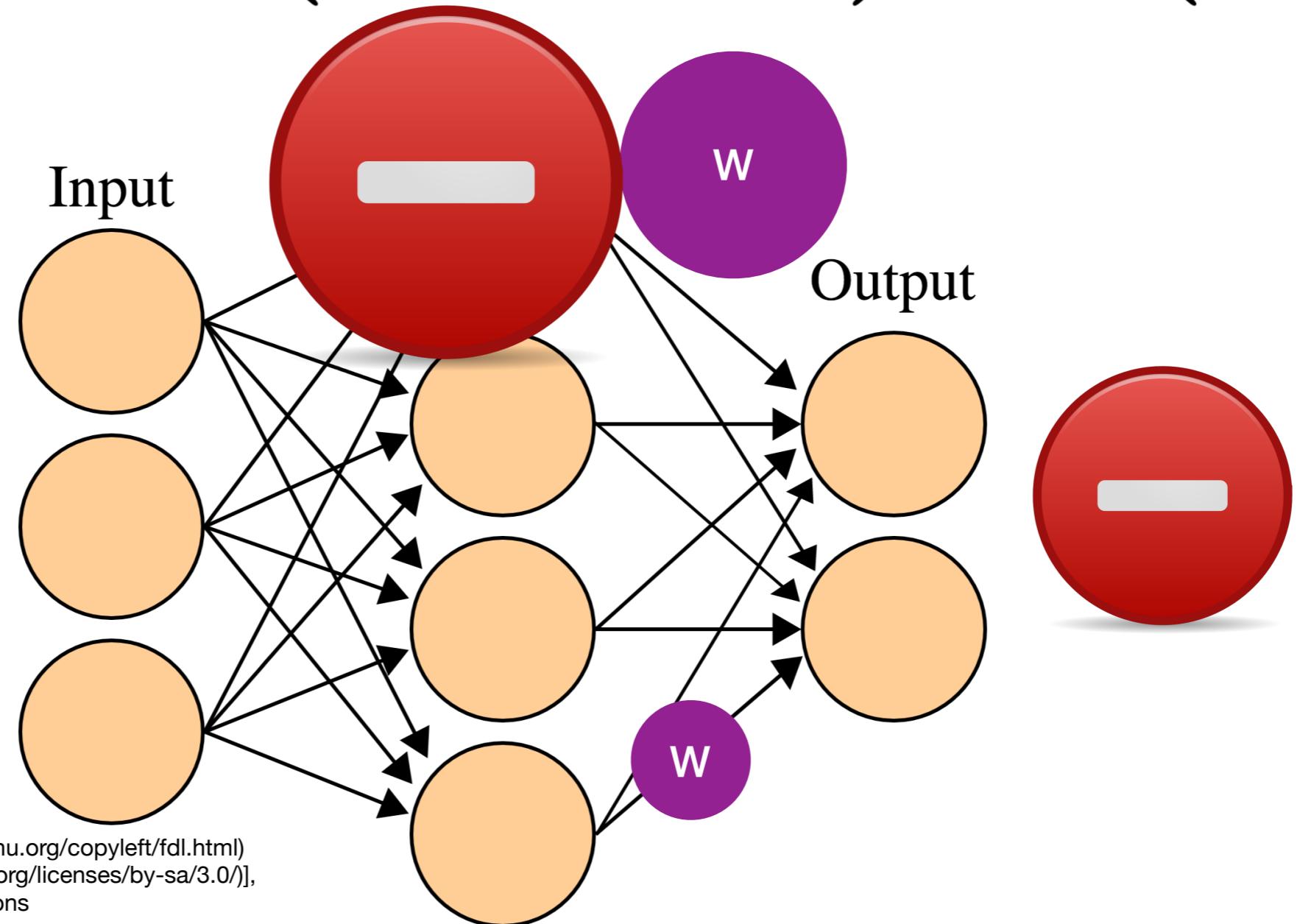
By en:User:Cburnett [GFDL (<http://www.gnu.org/copyleft/fdl.html>)  
or CC-BY-SA-3.0 (<http://creativecommons.org/licenses/by-sa/3.0/>)],  
via Wikimedia Commons

$$\sigma_3 = -(y - \hat{y}) f'(z_3)$$



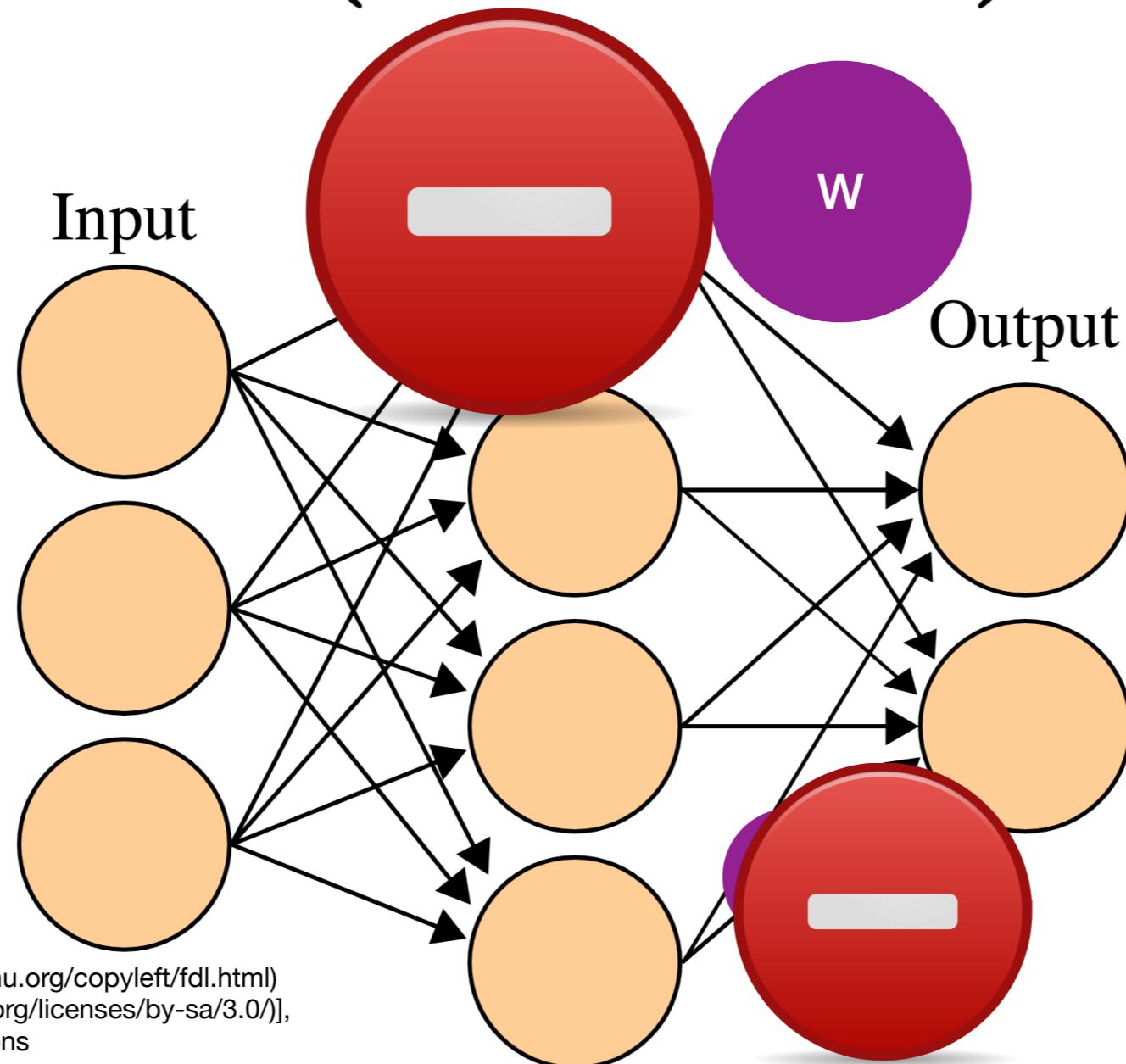
By en:User:Cburnett [GFDL (<http://www.gnu.org/copyleft/fdl.html>)  
or CC-BY-SA-3.0 (<http://creativecommons.org/licenses/by-sa/3.0/>)],  
via Wikimedia Commons

$$\sigma_3 = -(y - \hat{y}) f'(z_3)$$



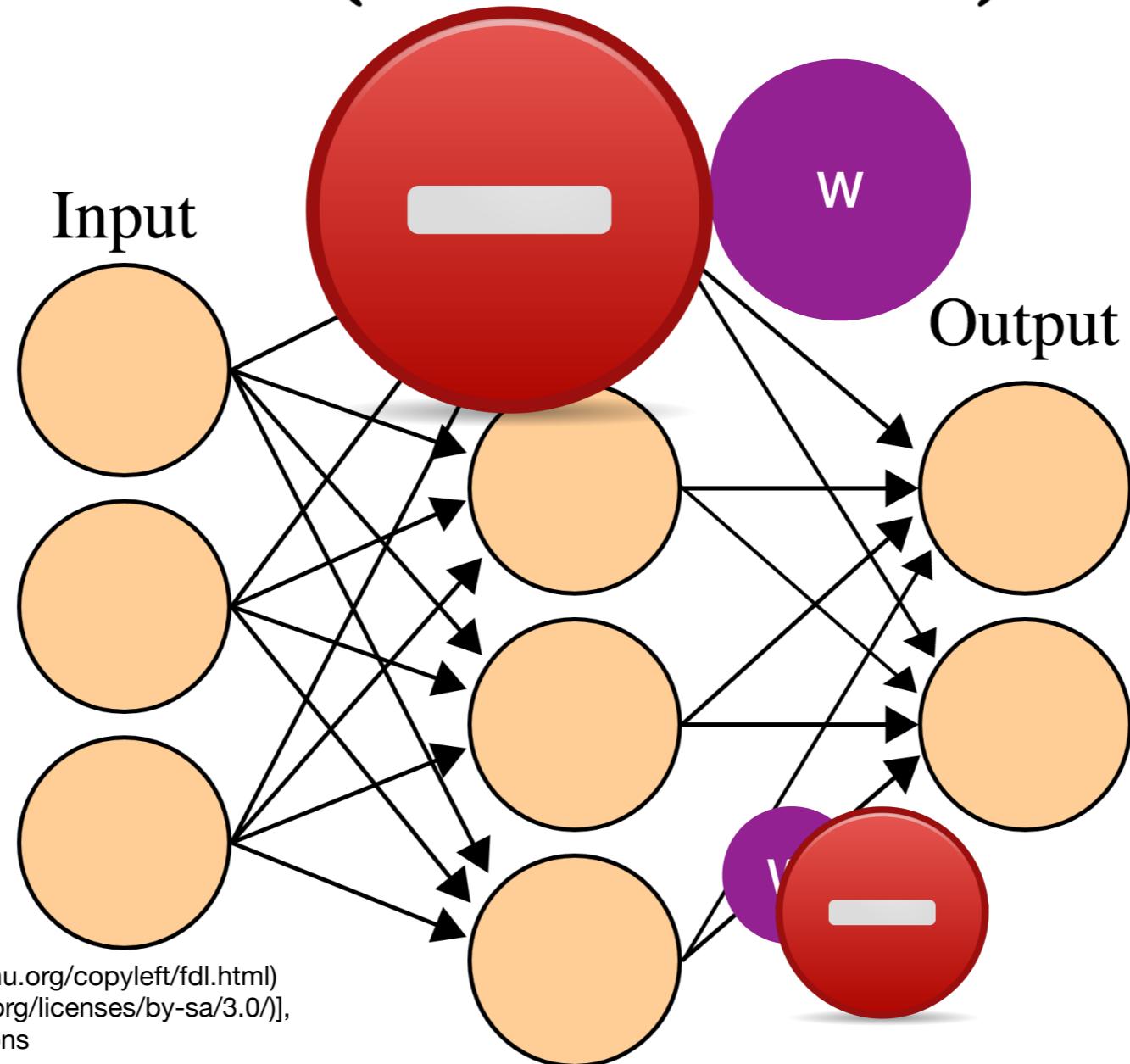
By en:User:Cburnett [GFDL (<http://www.gnu.org/copyleft/fdl.html>)  
or CC-BY-SA-3.0 (<http://creativecommons.org/licenses/by-sa/3.0/>)],  
via Wikimedia Commons

$$\sigma_3 = -(y - \hat{y}) f'(z_3)$$



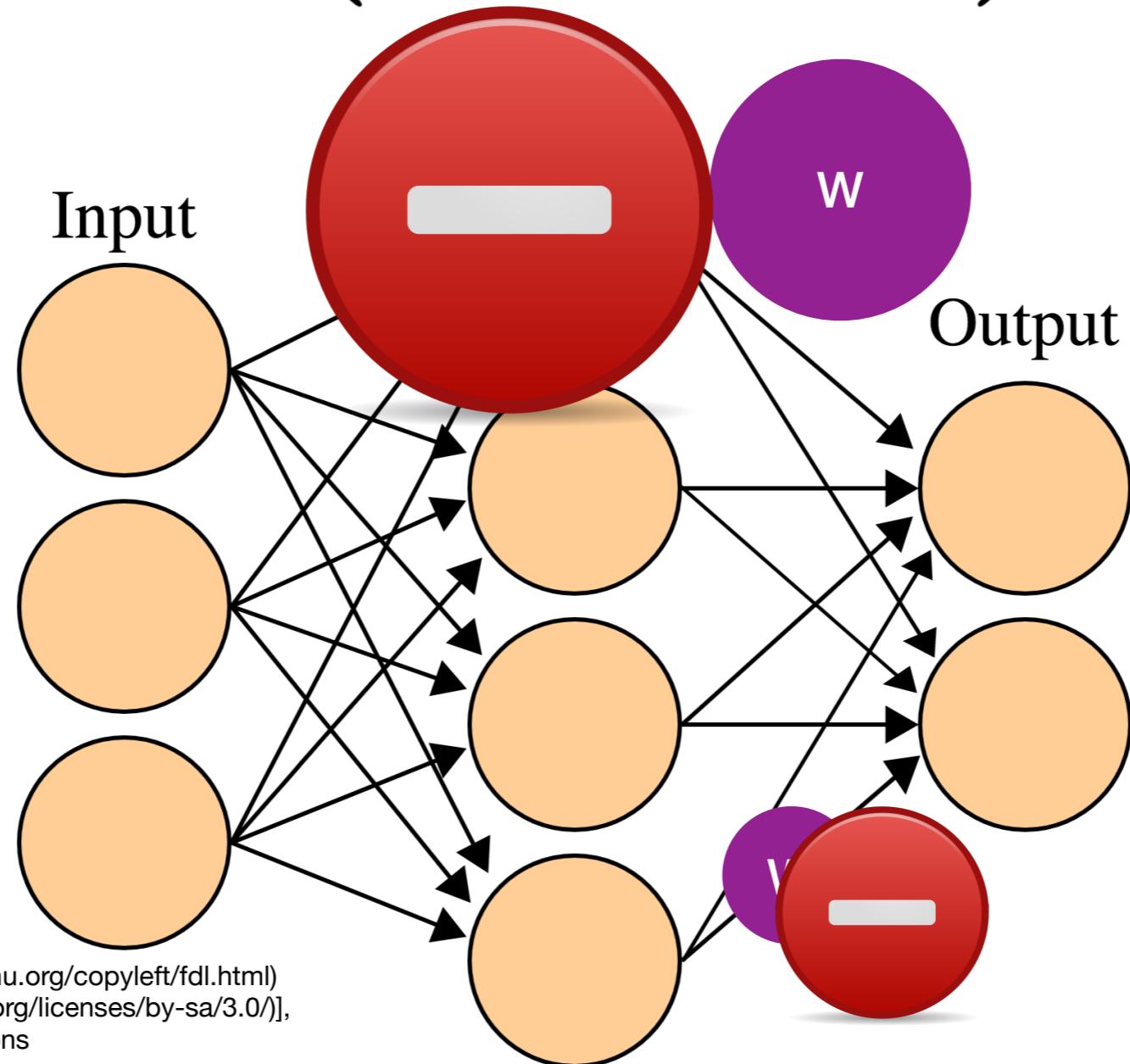
By en:User:Cburnett [GFDL (<http://www.gnu.org/copyleft/fdl.html>)  
or CC-BY-SA-3.0 (<http://creativecommons.org/licenses/by-sa/3.0/>)],  
via Wikimedia Commons

$$\sigma_3 = -(y - \hat{y}) f'(z_3)$$



By en:User:Cburnett [GFDL (<http://www.gnu.org/copyleft/fdl.html>)  
or CC-BY-SA-3.0 (<http://creativecommons.org/licenses/by-sa/3.0/>)],  
via Wikimedia Commons

$$\sigma_3 = -(y - \hat{y}) f'(z_3)$$

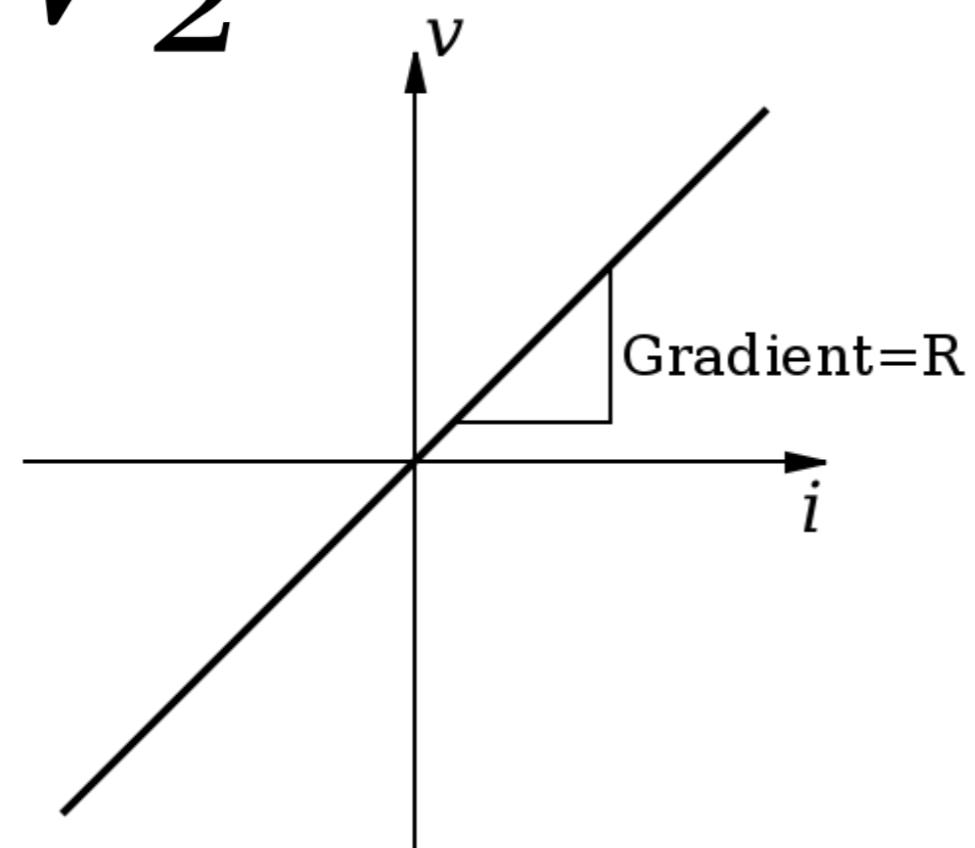


By en:User:Cburnett [GFDL (<http://www.gnu.org/copyleft/fdl.html>)  
or CC-BY-SA-3.0 (<http://creativecommons.org/licenses/by-sa/3.0/>)],  
via Wikimedia Commons

$$\sigma_3 = -(y - \hat{y}) f'(z_3)$$
$$\frac{\partial J}{\partial W_2} = a_2^T \sigma_3$$

$$\sigma_3 = -(y - \hat{y}) f'(z_3)$$

$$\frac{\partial J}{\partial W_2} = a_2^T \sigma_3$$



By inductiveload (From scratch in Inkscape) [Public domain], via Wikimedia Commons

$$\sigma_3 = -(y - \hat{y}) f'(z_3)$$

$$\frac{\partial J}{\partial W_2} = a_2^T \sigma_3$$

$$\sigma_2 = \sigma_3 W_2^T f'(z_2)$$

$$\sigma_3 = -(y - \hat{y}) f'(z_3)$$

$$\frac{\partial J}{\partial W_2} = a_2^T \sigma_3$$

$$\sigma_2 = \sigma_3 W_2^T f'(z_2)$$

$$\frac{\partial J}{\partial W_1} = X^T \sigma_2$$

# Summary

# Introduction to TensorFlow