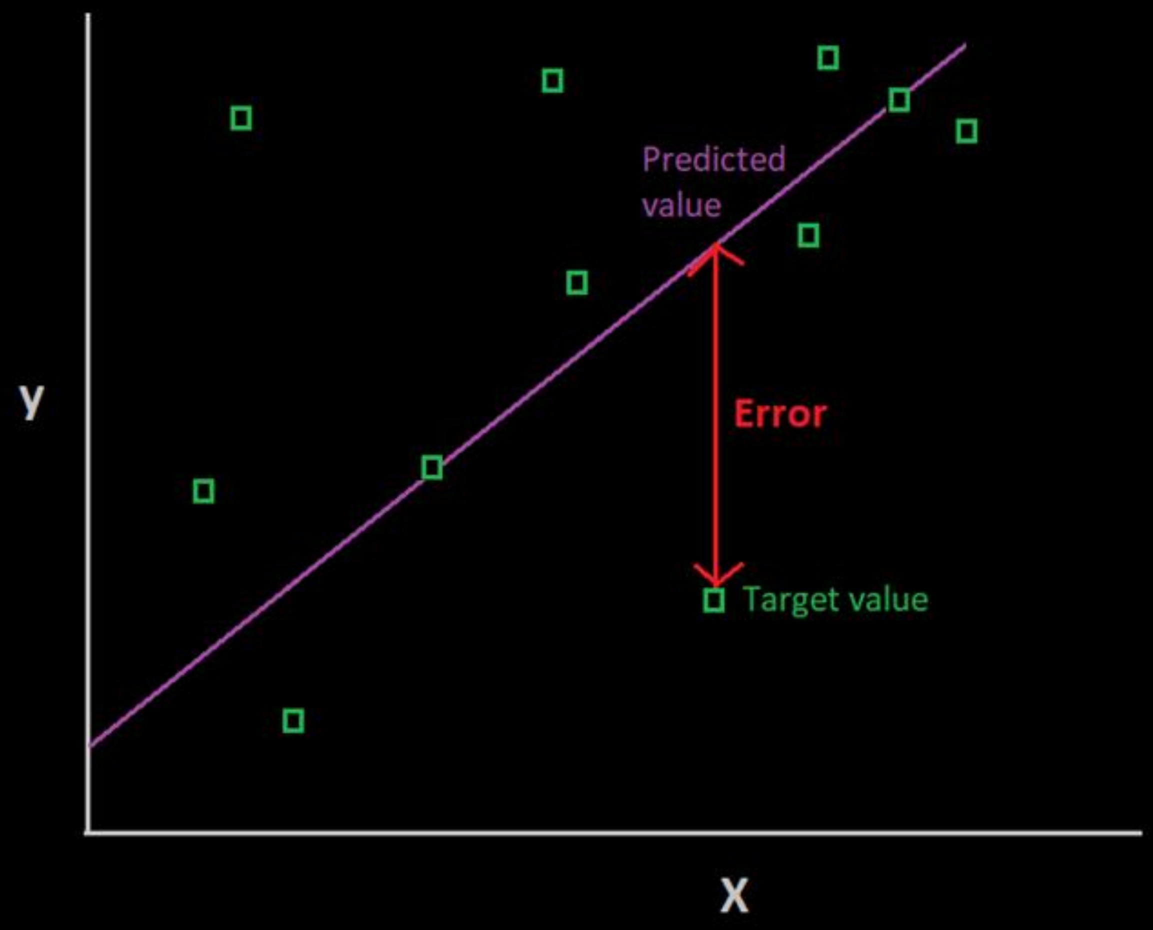
Loss Functions

Topics to be covered

- Mean Square Error (MSE) Loss Function
- Cross Entropy Loss Function
- Softmax Function

Why Loss Function ?



Error = Target value − Predicted Value

Why Loss Function ?

$\boxed{\text{if Error} = 5}$, how to reduce the error ?

$$\boxed{Loss\ Function = f(Error)}$$

$$\text{MSE} = \frac{1}{N} \sum \left(\text{Error} \right)^2$$

$$\text{MSE} = \frac{1}{N} \sum \left(\text{Target value} - \text{Predicted Value} \right)^2$$

Mean Square Error (MSE) Loss Function

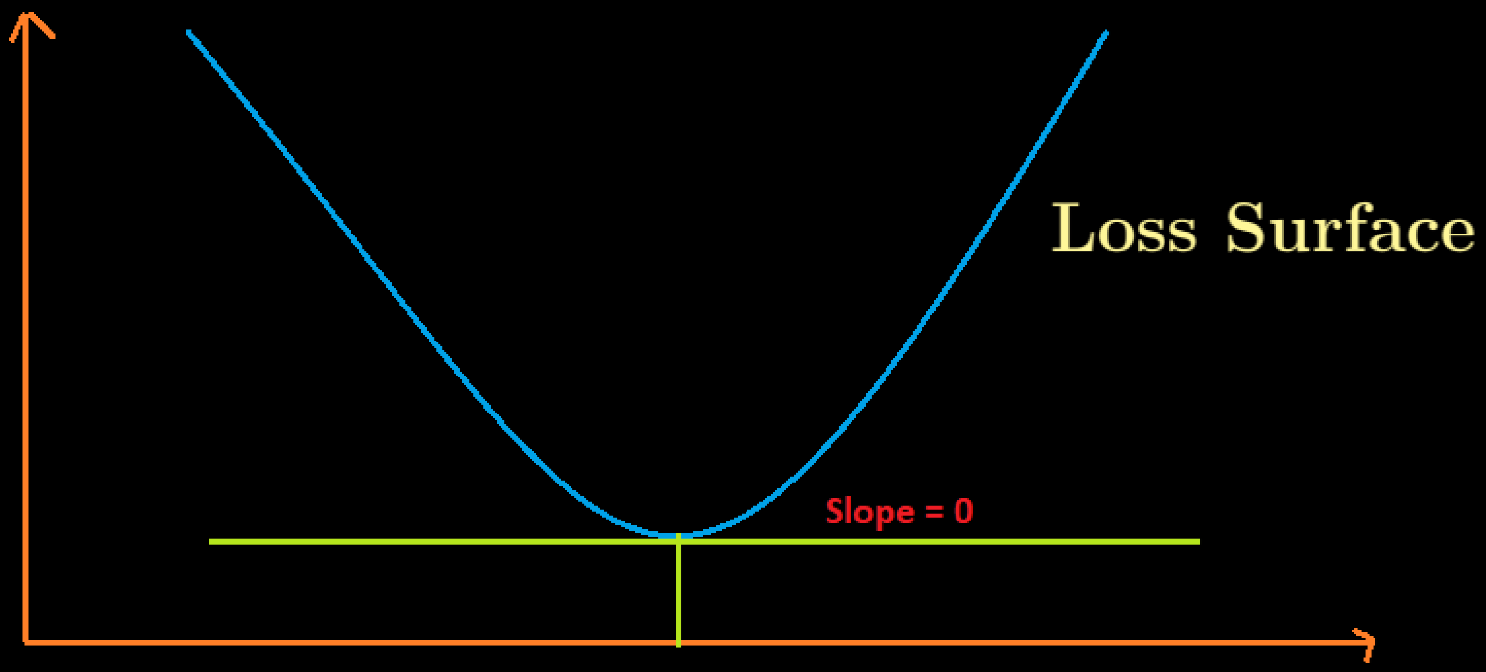
Mean Square Error (MSE) Loss Function

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y - y_{pred})^2$$

samples	y	y_{pred}	$Error = y - y_{pred}$	$SquaredError$
1	60	71	-11	121
2	50	52	-2	4
3	45	40	5	25

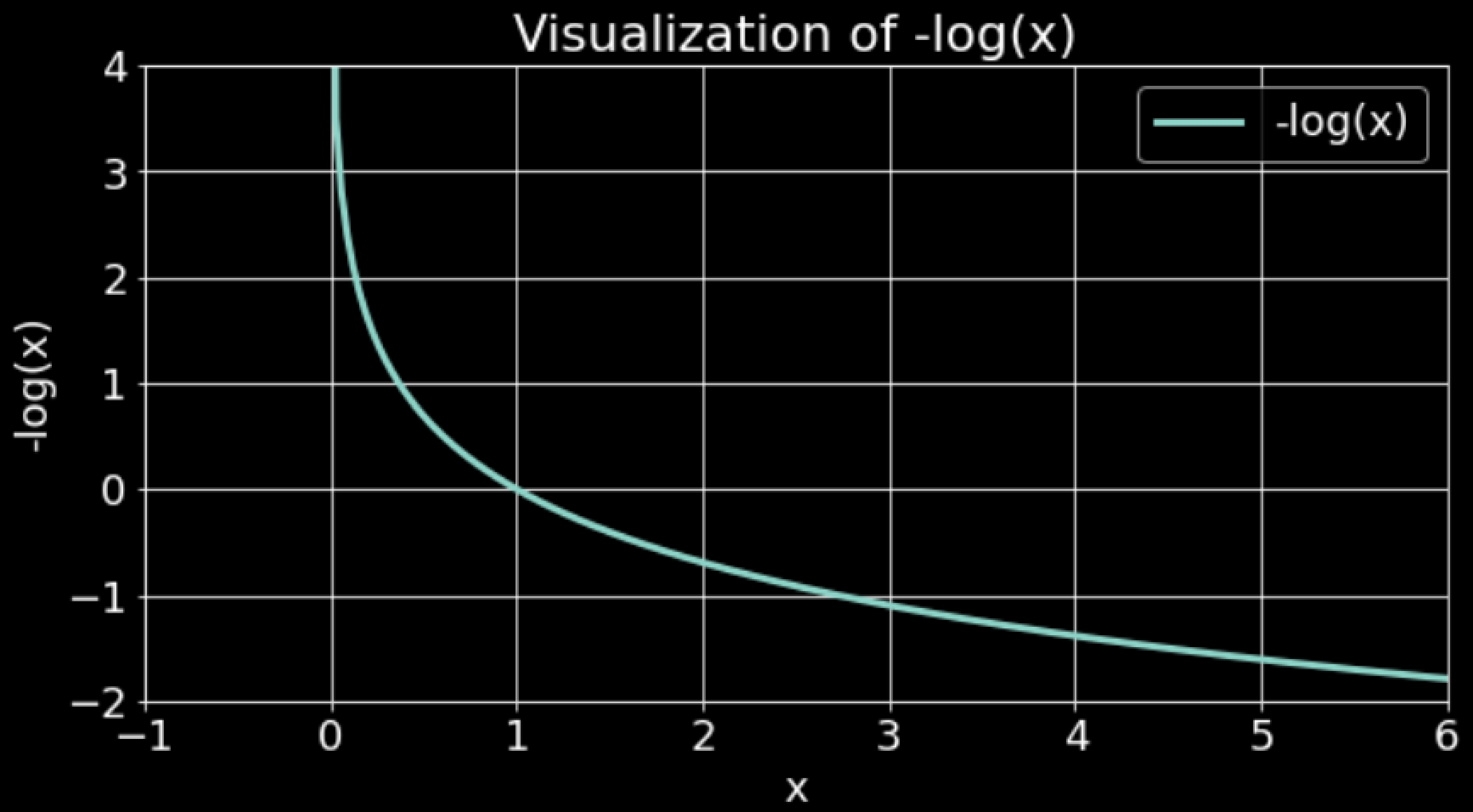
$$MSE = \frac{1}{3}(121 + 4 + 25) = 50$$

MSE Loss Function



Cross Entropy Loss Function

Visualization of Negative Log Function



$$\text{if} \quad x = 0.03$$

$$-log(x) = 3.5$$

$$\text{if} \quad x = 0.9$$

$$-log(x) = 0.1$$

Formula of Cross Entropy Loss

$$Cross\ Entropy = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{c}[(y_i\ log(p_i) + (1 - y_i)\ log(1 - p_i)]$$

where
y = Actual class Label
p = Predicted probability for the class
n = number of samples
c = number of classes

Binary Cross Entropy Loss

For two classes the Cross Entropy loss becomes Binary Cross Entropy loss (BCE) and its equation is given by

$$BCE = -\frac{1}{n}\sum_{i=1}^{n}[(y_i \, log(p_i) + (1 - y_i) \, log(1 - p_i)]$$

For better understanding, suppose $n = 1$, then we have

$$BCE = -[(y_i \, log(p_i) + (1 - y_i) \, log(1 - p_i)]$$

if class label = 1 i.e $y = 1$, then

$$\boxed{BCE = -log(p_i)}$$

if class label = 0 i.e $y = 0$, then

$$\boxed{BCE = -log(1 - p_i)}$$

Case 1 : When loss is high and need to minimize

- $y = 1$ and $p = 0.1$ then $-log(0.1)$ is high
- $y = 0$ and $p = 0.9$ then $-log(1 - 0.9)$ is high

Case 2 : When loss is low

- $y = 1$ and $p = 0.9$ then $-log(0.9)$ is low
- $y = 0$ and $p = 0.1$ then $-log(1 - 0.1)$ is low

Case 3 : Ideal Scenario

- $y = 1$ and $p = 1$ then $-log(1) = 0$
- $y = 0$ and $p = 0$ then $-log(1 - 0) = 0$

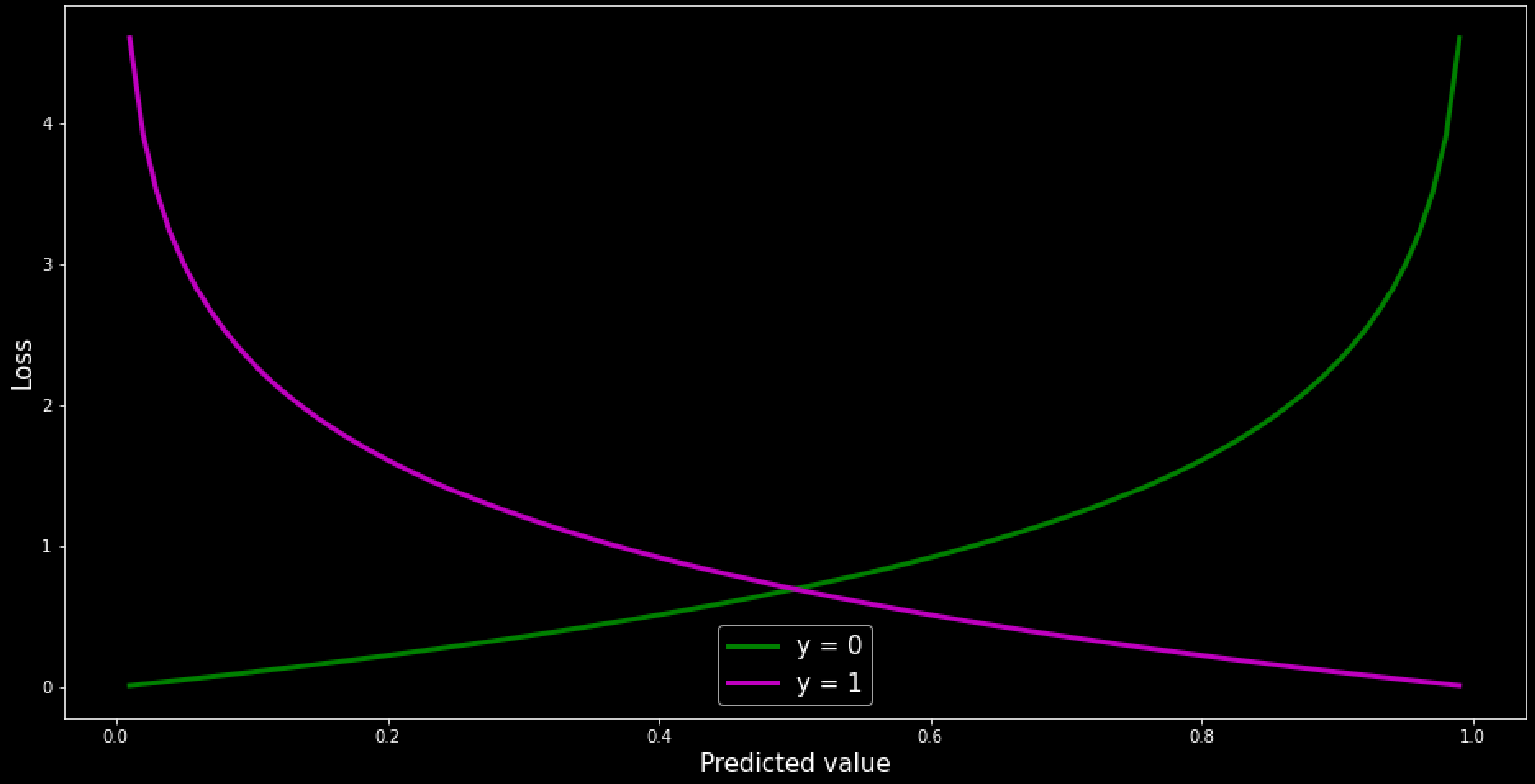
$y = 1$, then

$$\boxed{BCE = -log(p_i)}$$

$y = 0$, then

$$\boxed{BCE = -log(1 - p_i)}$$

Binary Cross Entropy Loss



Softmax function

Softmax function

- Cross Entropy loss uses softmax as an activation function.
- The Softmax turns all the class probabilities to values that sum up to 1.
- It is used for multiclass classification.
- For two categories softmax equals to sigmoid.

Mathematics of Softmax function

$$\boxed{\sigma(z) = \frac{e^{z}}{\sum e^{zi}}}$$

Suppose $z = [-1.28, 0.53, 0.87]$ is vector of output layer

The exponents of each value are:

$e^{-1.28} = 0.278 \quad e^{0.53} = 1.7 \quad e^{0.87} = 2.38$

Adding all the exponents values

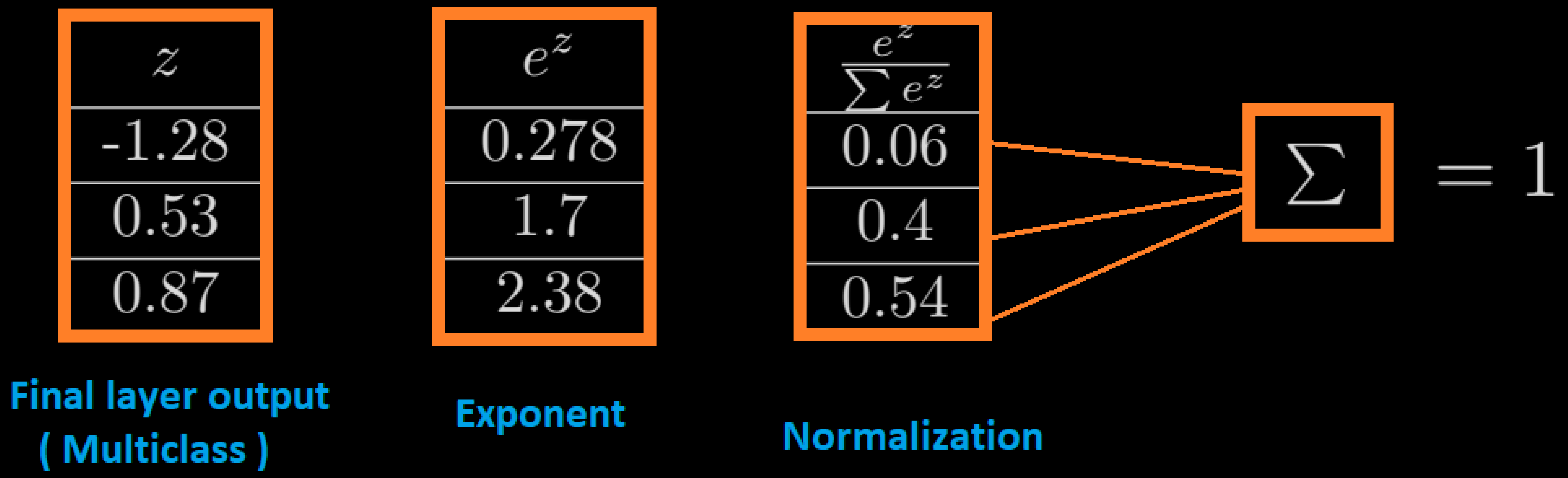
$0.278 + 1.7 + 2.38 = 4.36$

Normalizing each exponent value

$$\frac{0.278}{4.36} = 0.06 \qquad \frac{1.7}{4.36} = 0.4 \qquad \frac{2.38}{4.36} = 0.54$$

Sum of the Normalized values is equal to 1.

$0.06 + 0.4 + 0.54 = 1$


z
-1.28
0.53
0.87
Final layer output
(Multiclass)
e^z
0.278
1.7
2.38
Exponent
$\frac{e^z}{\sum e^z}$
0.06
0.4
0.54
Normalization
$\sum = 1$

Thank you !

Thank you !