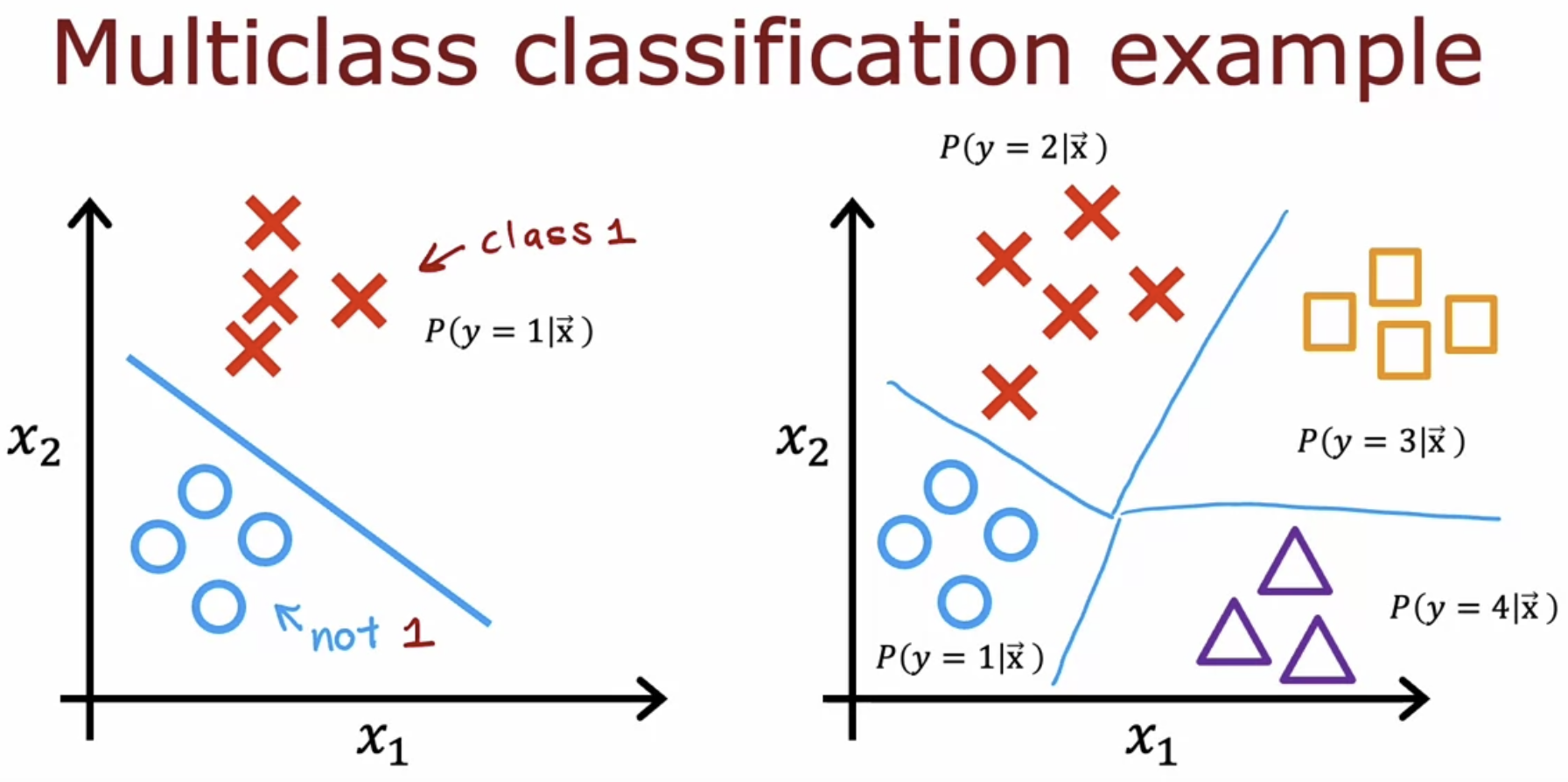
# **Multiclass**



# **SoftMax**

Graphical user interface, text, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Apply SoftMax regression with n=2, so there are only two possible output classes then SoftMax regression ends up computing basically the same thing as logistic regression.

The parameters end up being a little bit different, but it ends up reducing to logistic regression model.

But that's why the SoftMax regression model is the generalization of logistic regression.

Graphical user interface, application

Description automatically generated

aj and loss are inversely proportional.

If a j was very close to 1, then you beyond this part of the curve and the loss will be very small.

But if it thought, say, a j had only a 50% chance then the loss gets a little bit bigger.

The smaller a j is, the bigger the loss.

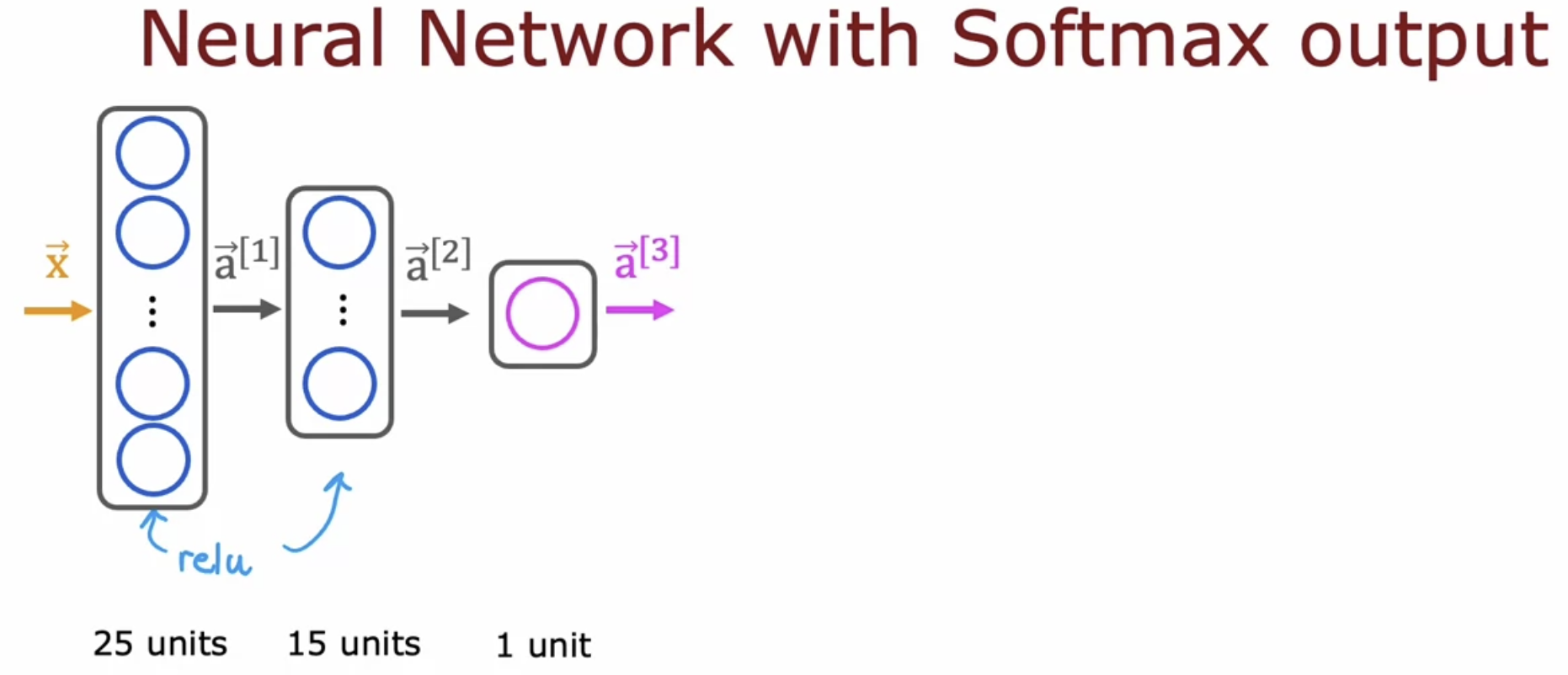
This incentivizes the algorithm to make a j as large as possible, as close to 1 as possible.

Because whatever the actual value y was, you want the algorithm to say hopefully that the chance of y being that value was pretty large.

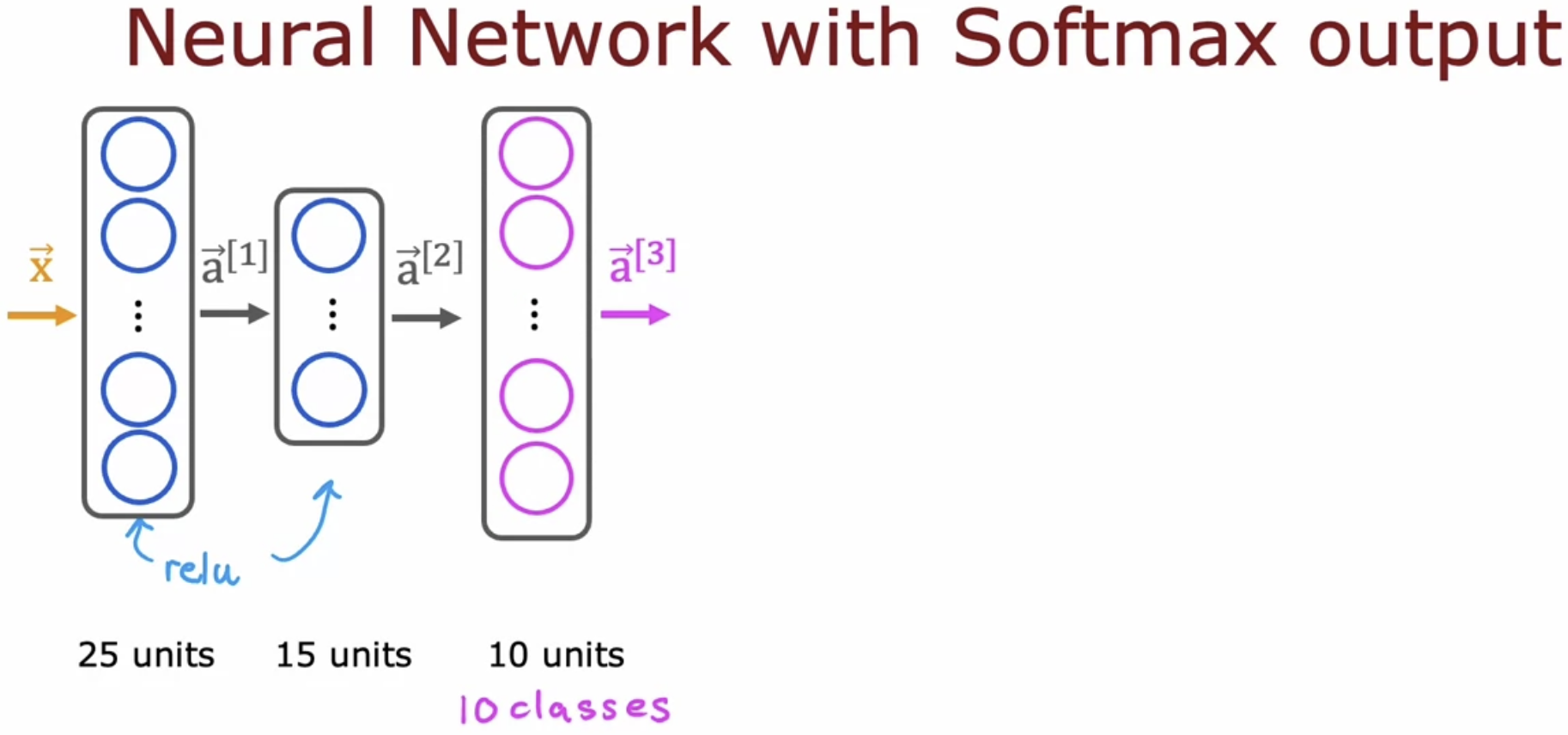
AIM 🡪 aj to make as large as possible, so that the loss is minimized.

For example, if y was equal to 2, you end up computing negative log of a2, but not any of the other negative log of a1 or the other terms here.

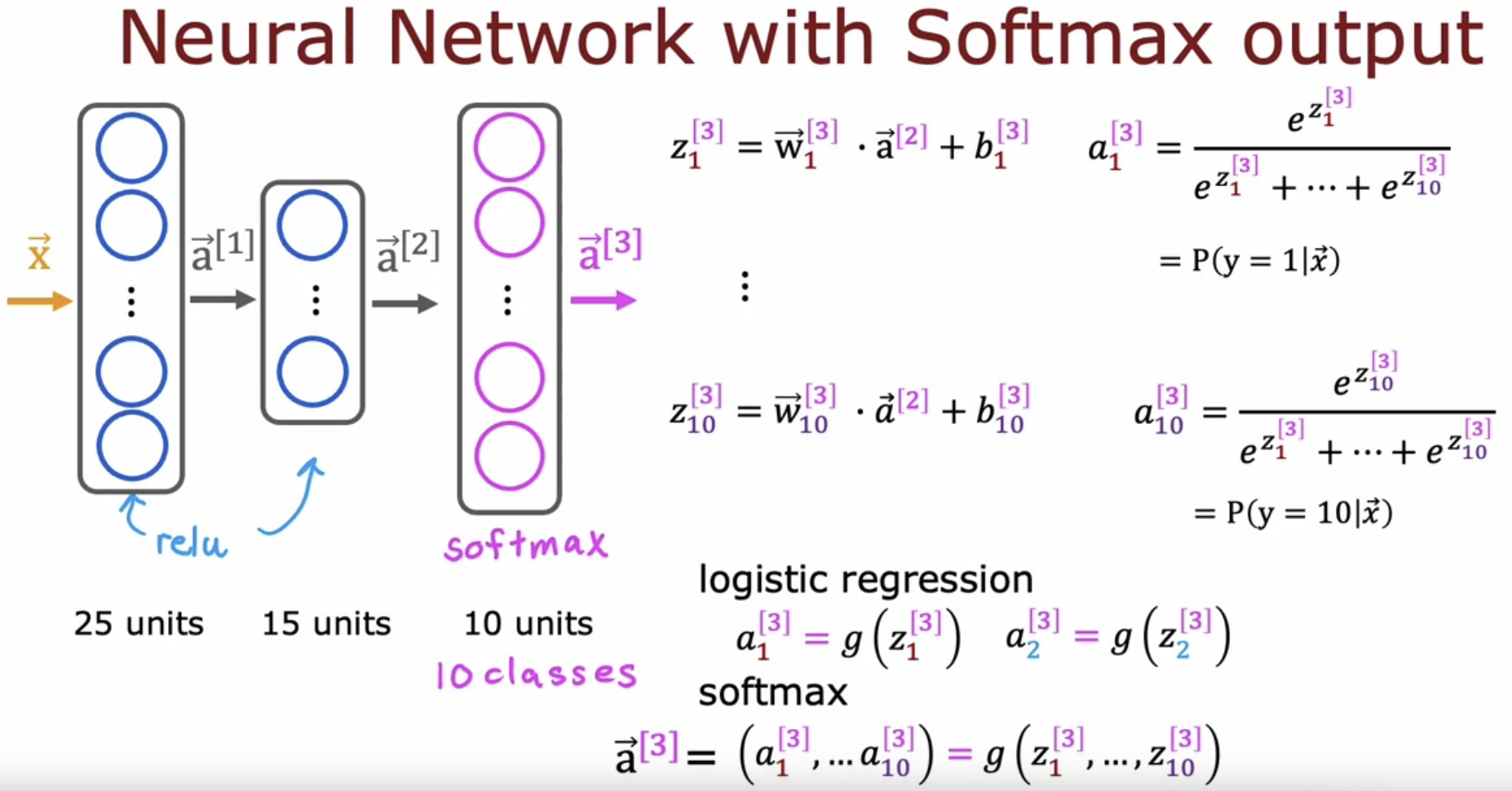
# **Neural Network with SoftMax output**



For 2 class

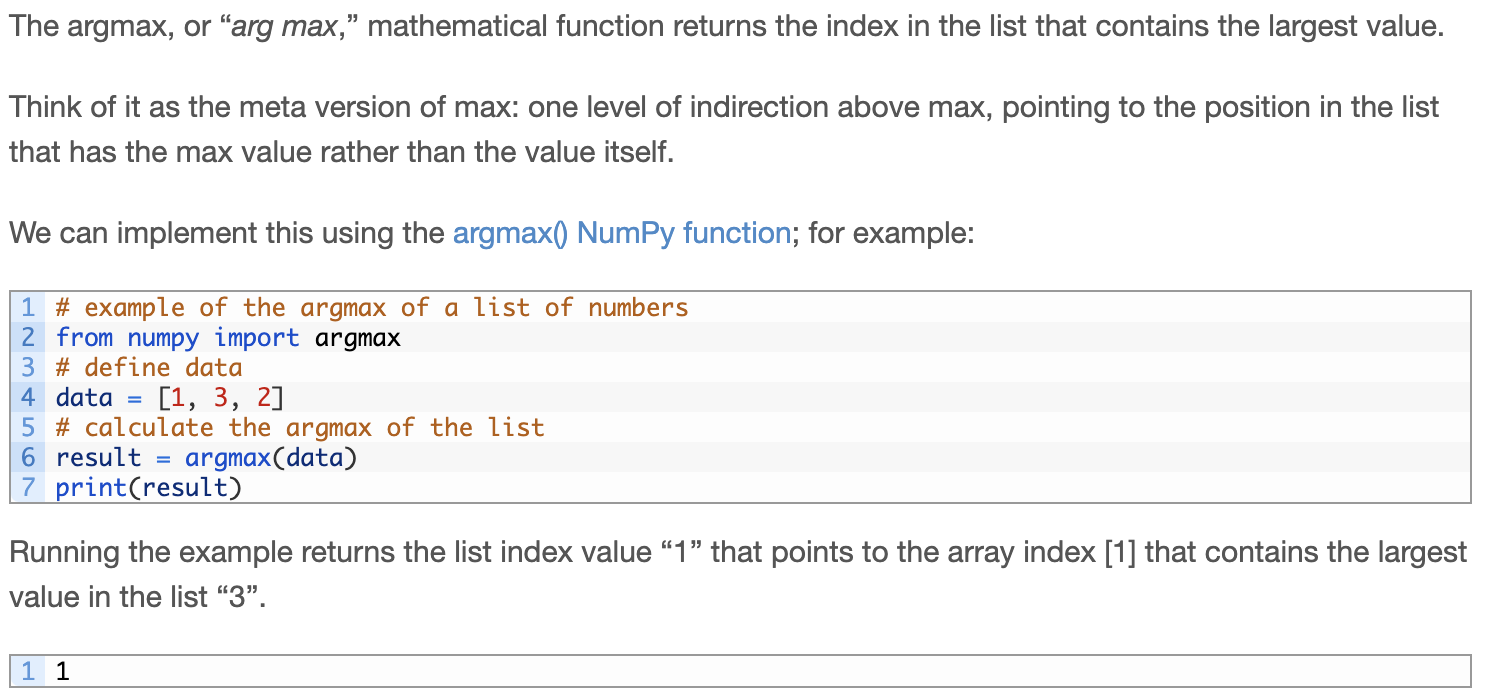


For 10 classes



## **Max function** **Graphical user interface, text, application, email Description automatically generated**

## **Argmax function**



## **SoftMax function**

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

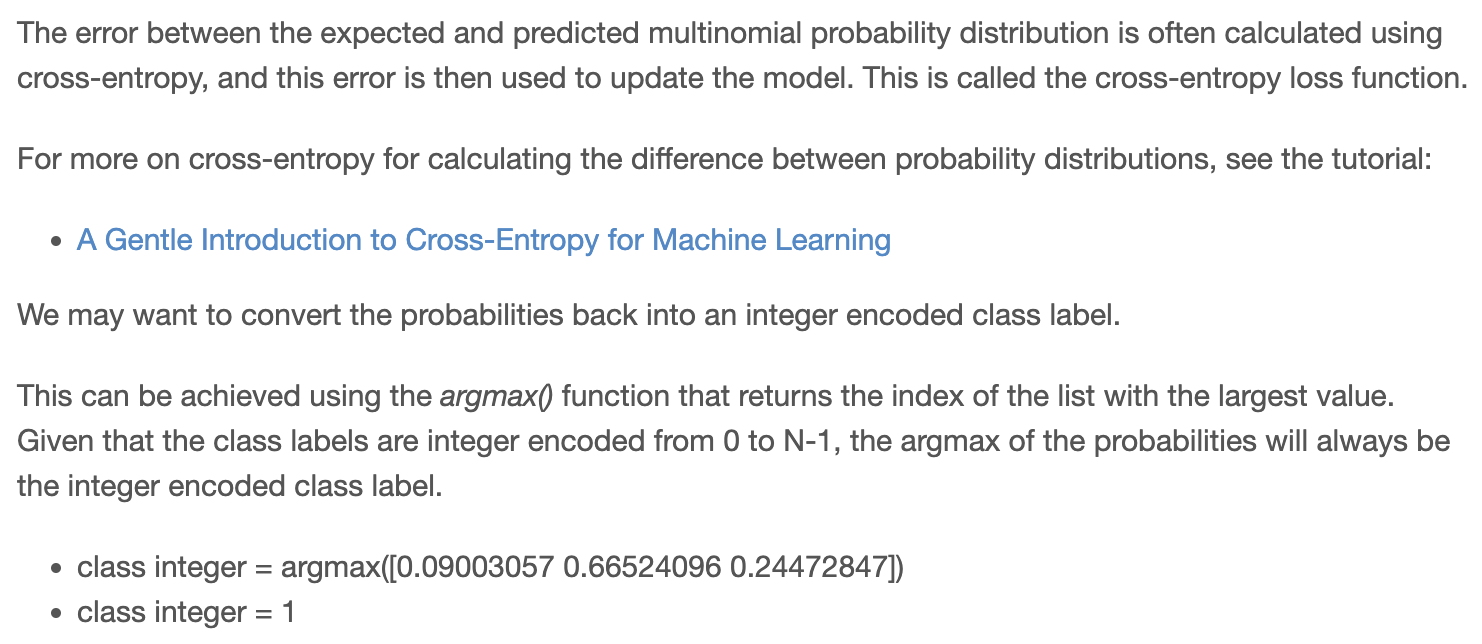
A picture containing table

Description automatically generated

## **SoftMax Activation Function**

Text, letter

Description automatically generated



Text

Description automatically generated

Logistic regression 🡪 Binary Cross Entropy Function

SoftMax regression 🡪 Sparse Categorical Cross Entropy Function

Category refers to, it is still classified y into categories. So it's categorical. This takes on values from 1 to 10.

Sparse refers to that y can only take on one of these 10 values.   
So each image is either 0 or 1 or 2 or so on up to 9. You're not going to see a picture that is simultaneously the number two and the number seven so sparse refers to that each digit is only one of these categories.

Cross Entropy function also called as intensive though the sparse categorical cross entropy loss function.