**Softmax Activation Function Explained**

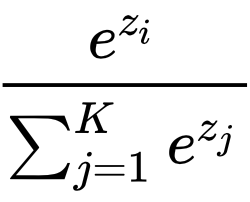
<https://towardsdatascience.com/softmax-activation-function-explained-a7e1bc3ad60>

Activation function(s) used on hidden layers are mostly the same for all hidden layers. It’s unlikely to see ReLU used on the first hidden layer, followed by a Hyperbolic tangent function — it’s usually ReLU or tanh all the way.

The function is great for **classification** problems, especially if you’re dealing with multi-class classification problems, as it will report back the “confidence score” for each class. Since we’re dealing with probabilities here, the scores returned by the softmax function will add up to 1.

The predicted class is, therefore, the item in the list where confidence score is the highest.

# Mathematical representation



It states that we need to apply a standard exponential function to each element of the output layer, and then normalize these values by dividing by the sum of all the exponentials. Doing so ensures the sum of all exponentiated values adds up to 1.

Here are the steps:

1. Exponentiate every element of the output layer and sum the results (around 181.73 in this case)
2. Take each element of the output layer, exponentiate it and divide by the sum obtained in step 1 (exp(1.3) / 181.37 = 3.67 / 181.37 = 0.02)