

Softmax is a mathematical function that converts a vector of raw scores (logits) into probabilities. It's commonly used in machine learning, particularly in multiclass classification problems. The softmax function takes an input vector and outputs a probability distribution over multiple classes.

Given a vector  $Z = [z_1, z_2, \dots, z_k]$ , the softmax function is defined as follows:

$$\text{Softmax}(Z)_i = \frac{\exp(z_i)}{\sum_{j=1}^k \exp(z_j)}$$

In simpler terms, each element in the softmax output is computed by taking the exponential ( $e^x$ ) of the corresponding element in the input vector, and then normalizing by dividing by the sum of the exponentials of all elements in the input vector.

The softmax function ensures that the resulting probabilities sum to 1, making it suitable for modeling a probability distribution over multiple classes. It's often used as the activation function in the output layer of a neural network for multiclass classification problems.

Softmax is particularly useful when you have multiple classes and you want to interpret the model's output as class probabilities. During training, the softmax output is compared to the actual distribution (one-hot encoded) using a loss function, and the model is optimized to reduce the difference between predicted and actual probabilities.