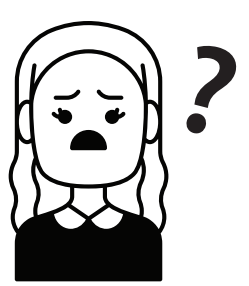


Neural Networks For People Who Get Confused Easily

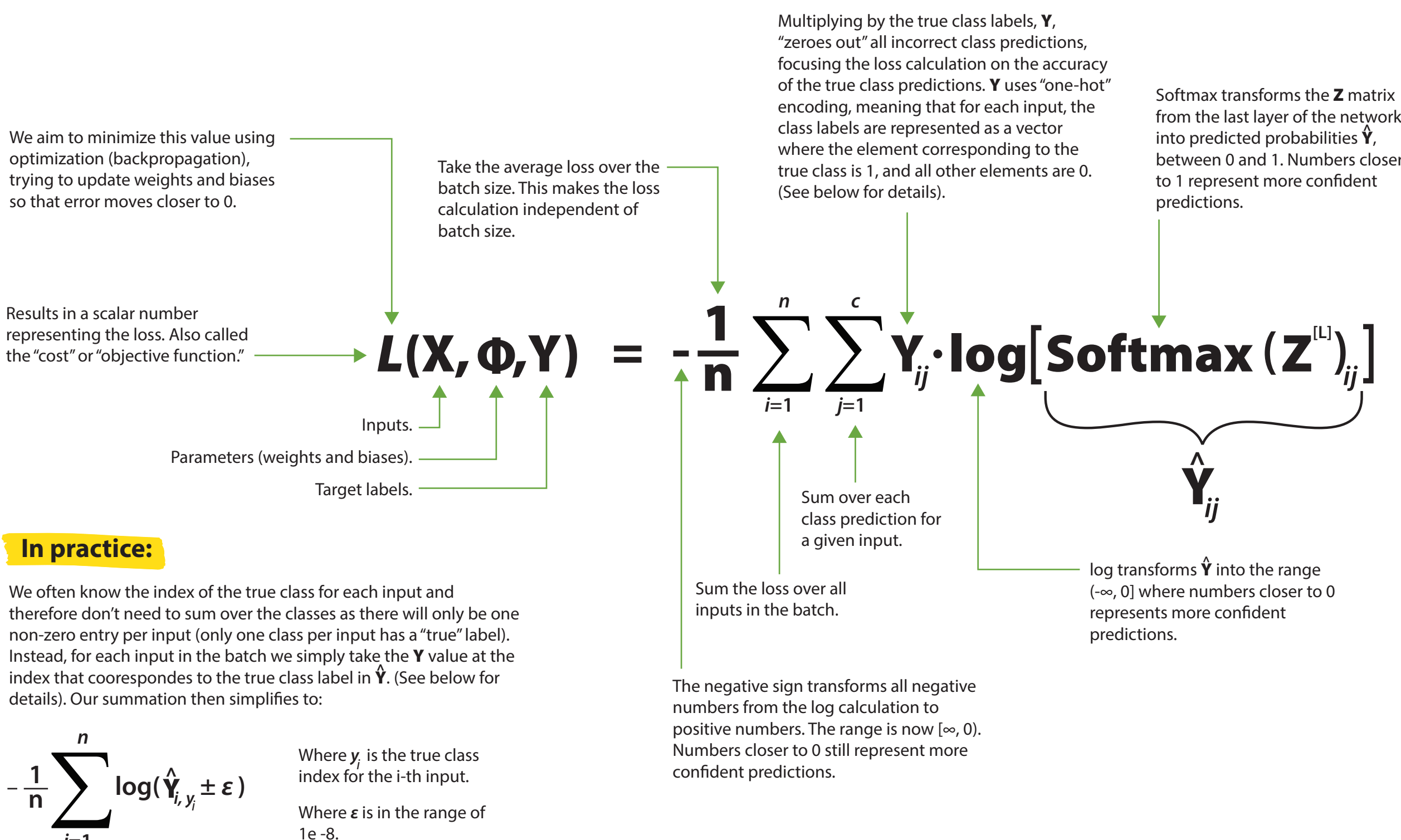
by James McCommon (A fellow who gets confused easily)



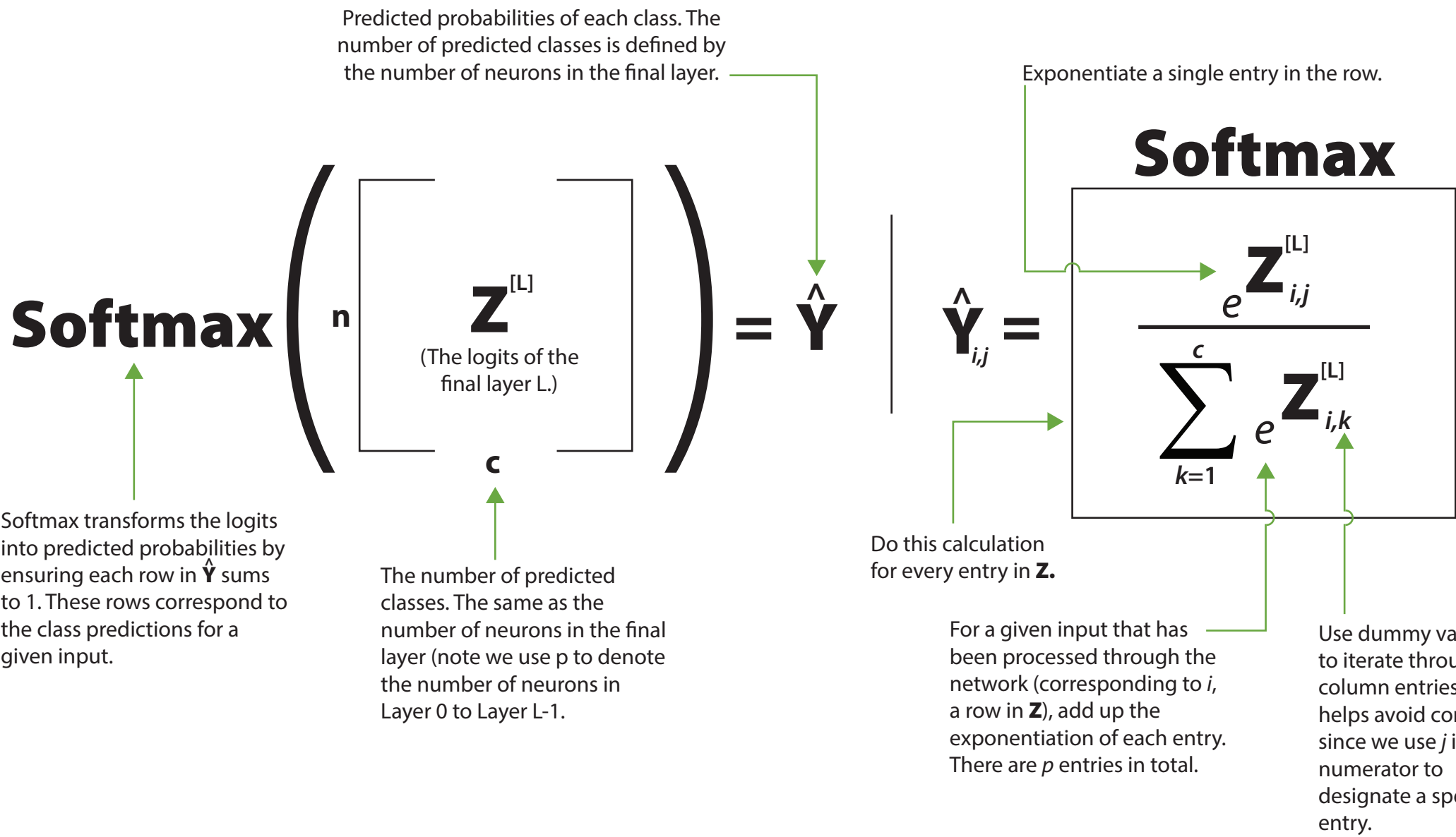
Softmax and cross-entropy loss

The conceptual picture

The cross-entropy loss function evaluates the accuracy of predictions by focusing on the probabilities assigned to the correct classes. Its goal is to minimize the error in predicting the true class, rather than penalizing the confidence in incorrect classes. For example, if the true class for a given input is Class 5 (out of 10 possible classes), the loss function specifically considers the probability assigned to Class 5. It does not concern itself with the probabilities assigned to other classes, such as Class 1 to 4 or Class 6 to 10.



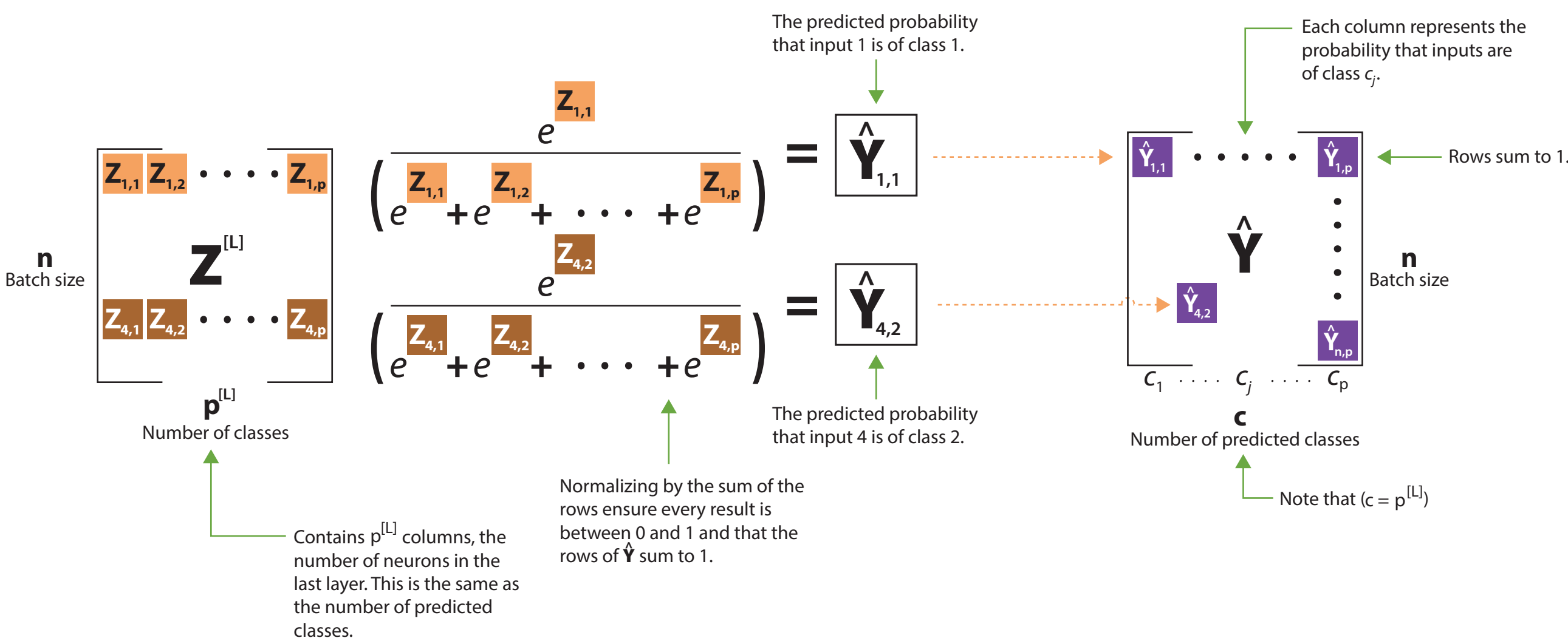
Softmax definition



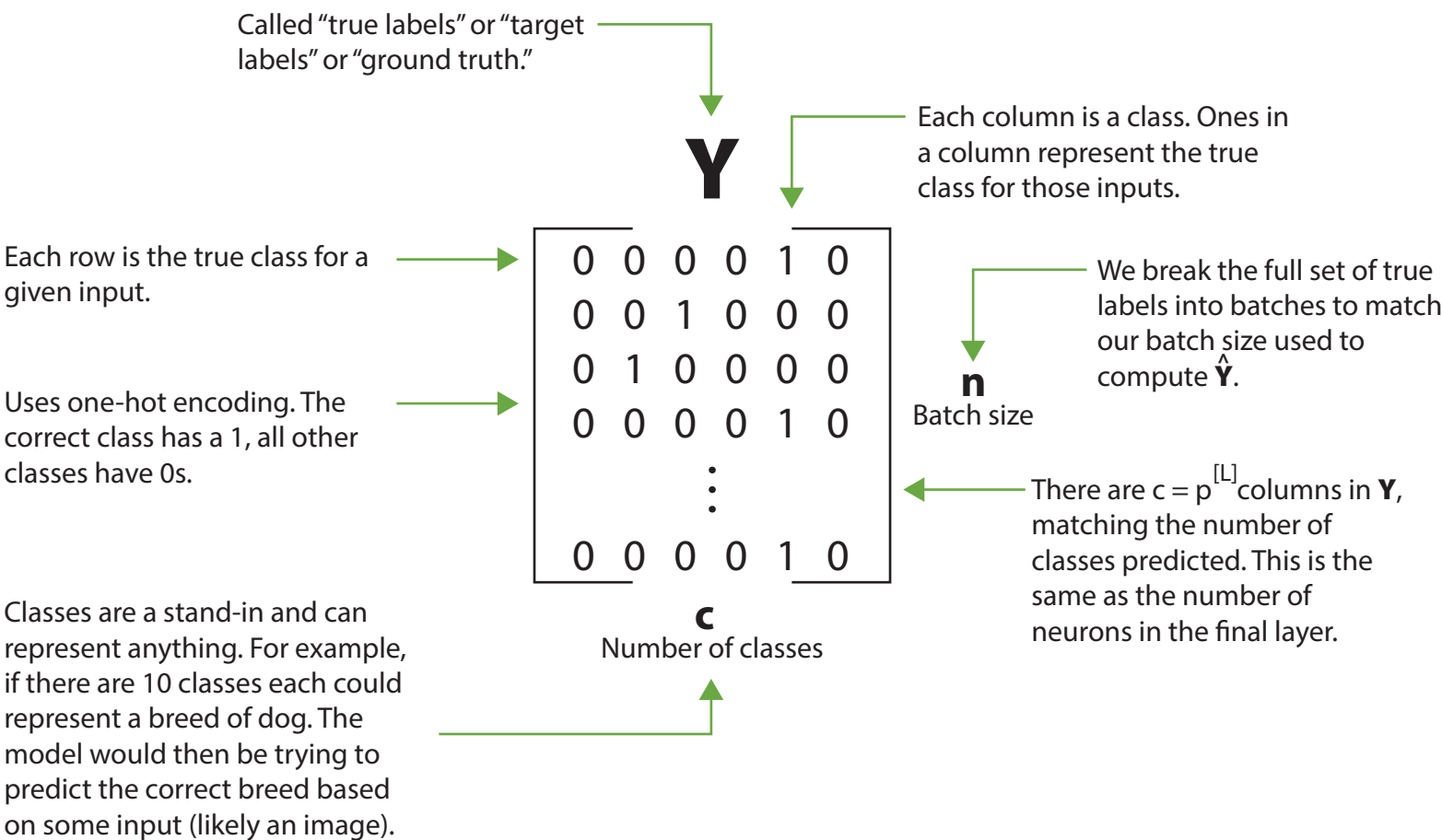
Notes

- Why do we exponentiate?
 - Range of Exponentials:** The exponential function maps any real number to a positive range $(0, \infty]$. This ensures all network outputs are non-negative, allowing them to be interpreted as probabilities, regardless of their initial sign or magnitude.
 - Stability and Saturation:** Exponentiation stabilizes the gradient calculations during backpropagation by providing a smoother gradient over a larger range of input values. This helps prevent issues like gradient vanishing or exploding, which can occur when differences in logits are very small.
 - Amplification of Probabilities:** The exponential function amplifies differences between logits, making the most likely classes stand out more distinctly. This "winner-take-all" behavior is particularly useful in multi-class classification settings, ensuring a clear distinction in probabilities for accurate class prediction.
- Softmax as a Generalization of the Logistic Function:** In cases with two classes, softmax reduces to the logistic sigmoid function, which also uses exponentials for transforming logits into probabilities. This highlights softmax as a natural extension of logistic regression to multiple classes.

Visualization of softmax



True labels



Cross-entropy loss

The cross-entropy loss function specifically targets the accuracy of the predictions by focusing on the probabilities assigned to the correct classes because the goal is to minimize the error in actual class predictions, not necessarily to penalize confidence in incorrect classes.

