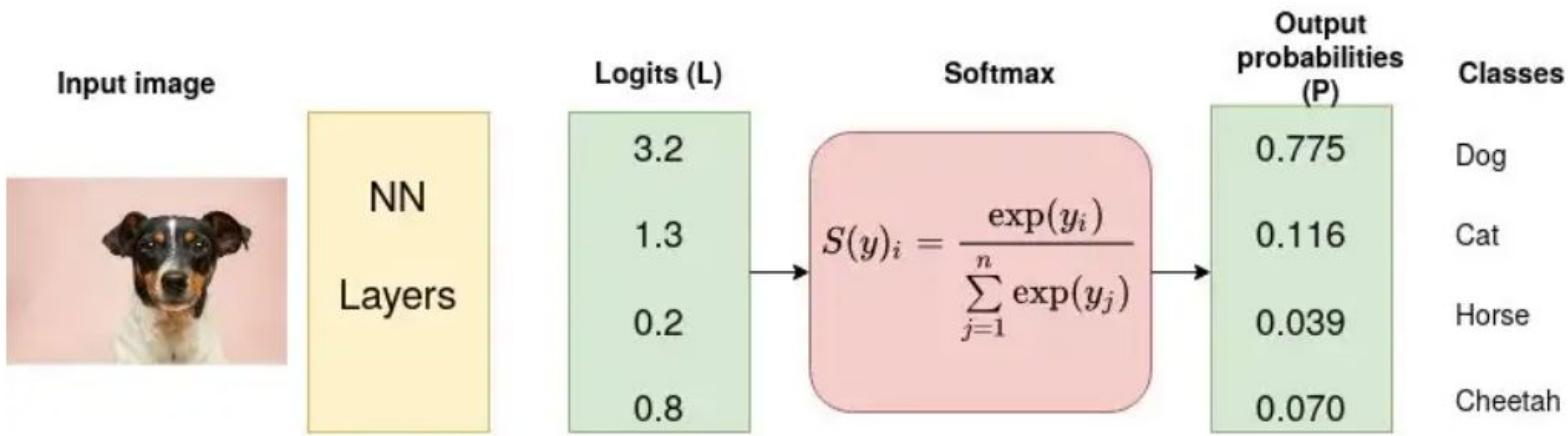
Multiclass Classification

# Softmax

# Sparse categorical cross-entropy

Wikipedia

# Softmax: From Logits to Probabilities




Input image source: Photo by Victor Grabarczyk on Unsplash . Diagram by author.

https://towardsdatascience.com/cross-entropy-loss-function-f38c4ec8643e

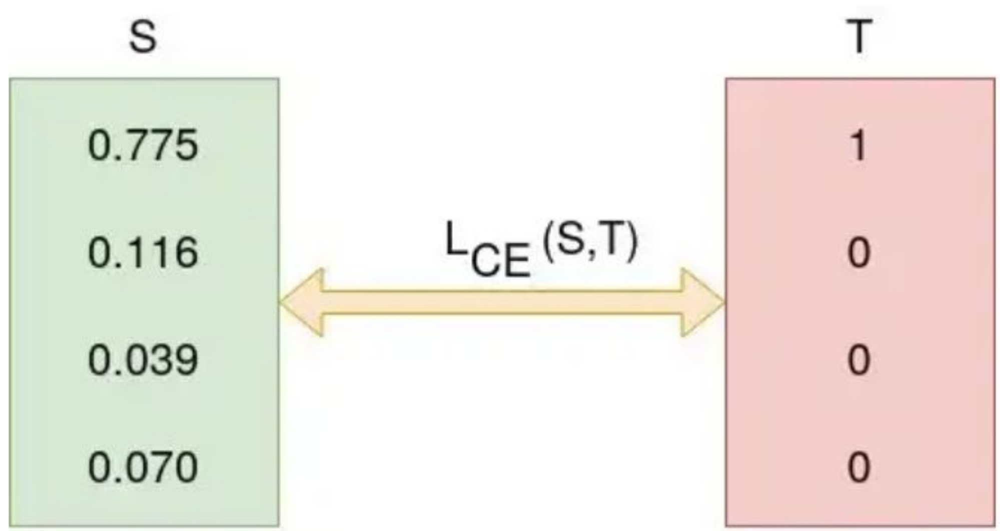
# Softmax

$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

- The softmax function takes as input a vector of K real numbers and normalizes it into a probability distribution consisting of K probabilities proportional to the exponentials of the input numbers. That is:
  - Prior to applying softmax, some vector components could be negative, or greater than one; and might not sum to 1; but after applying softmax, each component will be in the interval (0,1)
  - The components will add up to 1, so that they can be interpreted as probabilities
  - The larger input components will correspond to larger probabilities.

# Categorical cross-entropy

- Computes the *cross-entropy loss* between the labels and predictions.
- Cross-entropy is a quantity from the field of information theory that measures the distance between probability distributions or, in this case, between the ground-truth distribution and the model's predictions.



# Categorical cross-entropy

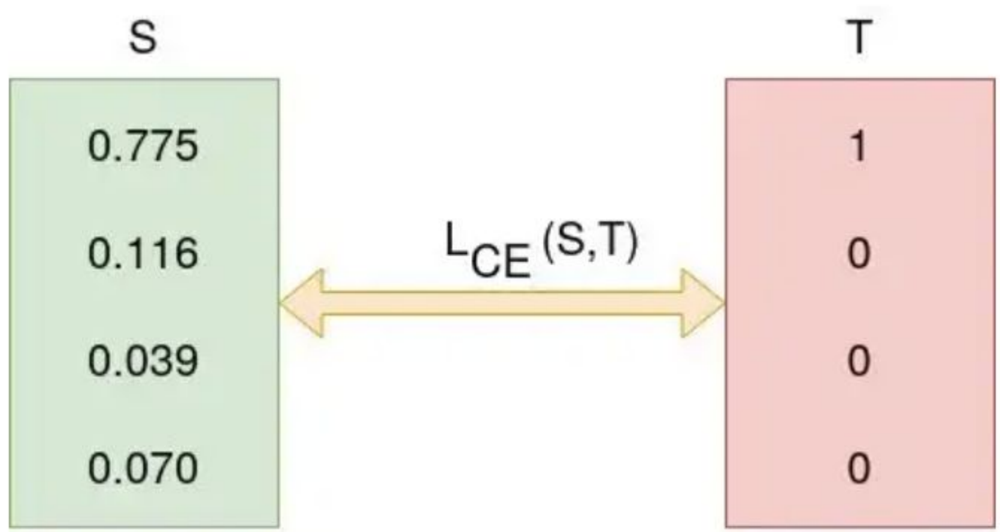
$$\boxed{L_{\text{CE}} = -\sum_{i=1}^{n} t_i \log(p_i), \quad \text{for n classes,}}$$

where $t_i$ is the truth label and $p_i$ is the Softmax probability for the $i^{th}$ class.

# Categorical cross-entropy

$$L_{\text{CE}} = -\sum_{i=1}^{n} t_i \log(p_i), \quad \text{for n classes,}$$

$$\begin{aligned} L_{CE} &= -\sum_{i=1} T_i \log(S_i) \\ &= -\left[1\log_2(0.775) + 0\log_2(0.126) + 0\log_2(0.039) + 0\log_2(0.070)\right] \\ &= -\log_2(0.775) \\ &= 0.3677 \end{aligned}$$



# Sparse Categorical cross-entropy

- Categorical cross-entropy is used when true labels are one-hot encoded, for example, we have the following true values for 3-class classification problem [1,0,0], [0,1,0] and [0,0,1].
- In **sparse categorical cross-entropy** , truth labels are integer encoded, for example, [1], [2] and [3] for 3-class problem.