Choose Loss Functions When Training Deep Learning Neural Networks

As part of the optimization algorithm, the error for the current state of the model requires to choose of an error function, conventionally called a **loss function**.

Neural network models learn a mapping from inputs to outputs from examples and the choice of loss function must match the framing of the specific predictive modeling problem, such as classification or regression.

In which situations should we use a specific loss function like categorical, sparse, binary.

This example of loss functions and their cases:

Problem Type	Output Type	Final Activation Function	Loss Function
Regression	Numerical value	Linear	Mean Squared Error (MSE)
Classification	Binary outcome	Sigmoid	Binary Cross Entropy
Classification	Single label, multiple classes	Softmax	Cross Entropy
Classification	Multiple labels, multiple classes	Sigmoid	Binary Cross Entropy

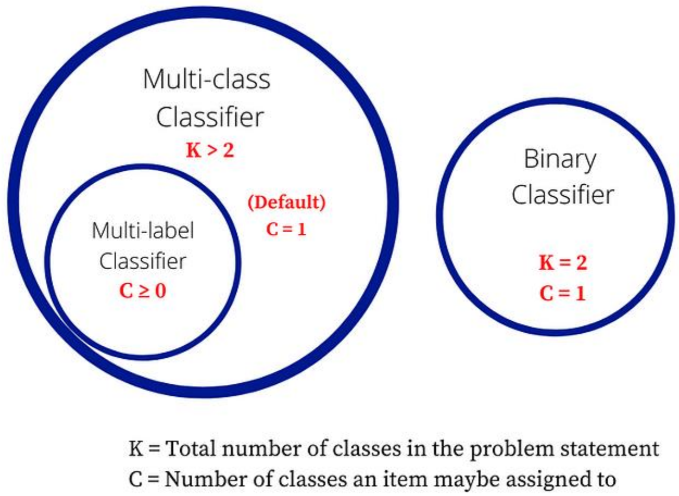
Regression, numerical value

The activation function is linear and loss function can be MSE (Mean Square Error).

Classification:

Classification task can be sub-divided into:

- binary classification (2 classes)
- Single label, multiclass classification.
- Multiple label, multiclass classification.



For all these cases, we will define the **Cross-Entropy Loss**. The main difference will be the input (output of the activation functions) to this loss.

The formulation of the **Cross-Entropy** Loss is the following:

Binary: 2 Classes

$$\text{Loss} = -\frac{1}{\substack{\text{output}\\\text{size}}} \sum_{i=1}^{\substack{\text{output}\\\text{size}}} y_i \cdot \log \hat{y}_i + (1 - y_i) \cdot \log (1 - \hat{y}_i)$$

Binary Cross-Entropy

Multi-Classes:

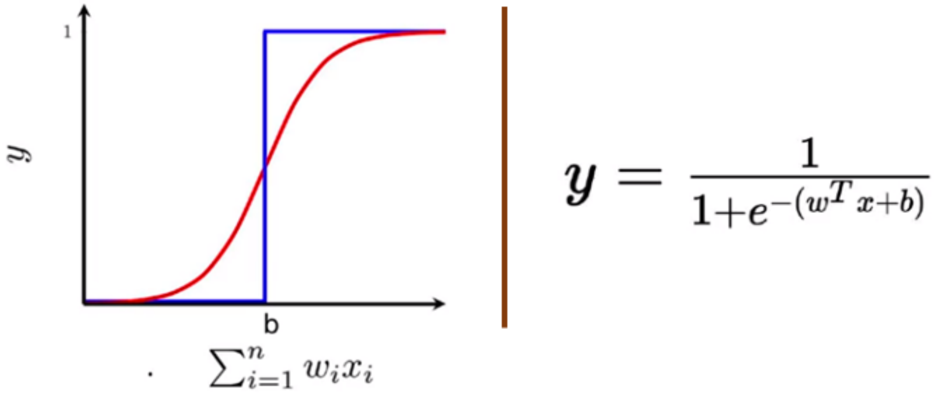
$$\text{Loss} = -\sum_{i=1}^{\substack{\text{output}\\\text{size}}} y_i \cdot \log \hat{y}_i$$

Multiclass Cross-Entropy

Logistic Loss and **Multinomial Logistic Loss** are other names for **Cross-Entropy loss**.

Binary classification, binary cross-entropy loss function

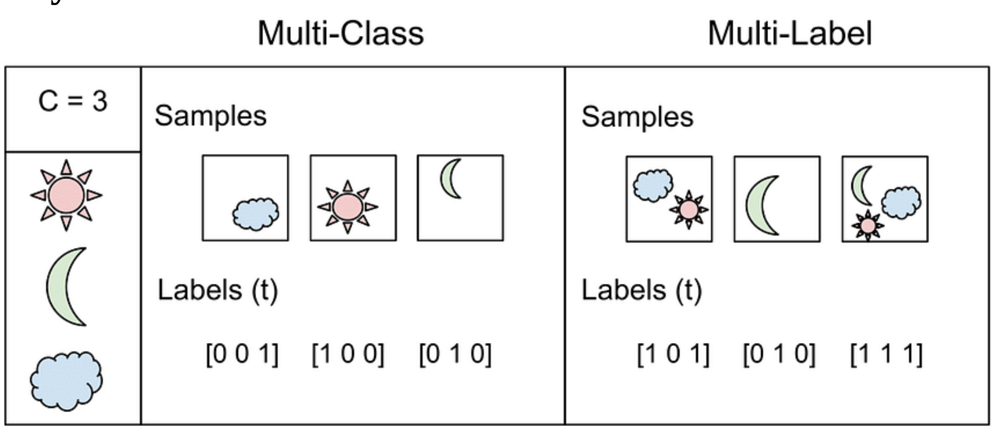
In this case, the vanilla classification (2 classes). This takes **only one** output value between -1 or 1. Binary classification with one output should takes the **sigmoid** function as activation.



Sigmoid function

Also called **Sigmoid Cross-Entropy loss**. It is a **Sigmoid activation** plus a **Cross-Entropy loss**.

Non-Binary classification



Difference between Multi-class and Multi-label

Multi-Class only classify one object from multiples objects in one sample.

Multi-Label can classify multiples objects in one sample.

Multi-Class Single Label:

In this case, we can calculate using two different methods: **Categorical Cross-Entropy** and **Sparse Categorical Cross-Entropy**. We can explain the following way:

- categorical_crossentropy (cce) produces a one-hot array containing the probable match for each category,
- sparse_categorical_crossentropy (scce) produces a category index of the most likely matching category.

Binary_crossentropy is utilized when dealing with a single class,
while categorical_crossentropy is used in case of multiple classes, and

the loss is applied based on one-hot encoded values such as [1,0,0], [0,1,0], [0,0,1] .

The sparse_categorical_crossentropy operates on integers, specifically class indices rather than actual values. This loss function computes logarithm only for the output index that is indicated by the ground truth. For instance, if the model's output is [0.1, 0.3, 0.7] and the ground truth is 3 (assuming indexing from 1), then the loss function will only compute the logarithm of 0.7 . Despite this modification, the final value remains the same as in the regular version of categorical crossentropy, where other values are multiplied by zero due to one-hot encoding.

Difference between categorical and binary cross entropy

we don’t need any one-hot-encoded vector labels when we are going to train using binary_cross_entrpoy. Some have suggested to represent one_hot vectors as [0. 1.] (if class is 1) or [1. 0.] (if class is 0) for binary_cross_entropy.

References

fmorenovr.medium.com/sparse-categorical-cross-entropy-vs-categorical-cross-entropy-ea01d0392d28

https://copyprogramming.com/howto/what-is-the-difference-between-sparse-categorical-crossentropy-and-categorical-crossentropy

https://machinelearningmastery.com/how-to-choose-loss-functions-when-training-deep-learning-neural-networks/

https://towardsdatascience.com/understanding-binary-cross-entropy-log-loss-a-visual-explanation-a3ac6025181a