CLASSIFICATION LOSS FUNCTIONS

LOSS FUNCTION FOR CLASSIFICATION: CROSS-ENTROPY

Cross-entropy loss is also referred to as “logarithmic loss,” “logistic loss,” or “log loss”.

Cross-entropy loss, or log loss; measures the performance of a classification model that outputs a probability value between 0 and 1.

During model training the goal is to minimize loss. A model that predicts perfect probabilities has a cross entropy or log loss of 0.0 --- But not at the cost of the model memorizing the data.

The common Cross-Entropy related loss functions are as follows -

1. Binary Cross Entropy Loss Function: Used for a two-class classification problem
2. Multi-Class Cross-Entropy Loss: Used when true labels are one-hot encoded
3. Sparse Multiclass Cross-Entropy Loss: Used when true labels are integer value

CROSS-ENTROPY: BINARY CLASSIFICATION

$$-(y\ log(\hat{y}) + (1 - y)\ log(1 - \hat{y}))$$

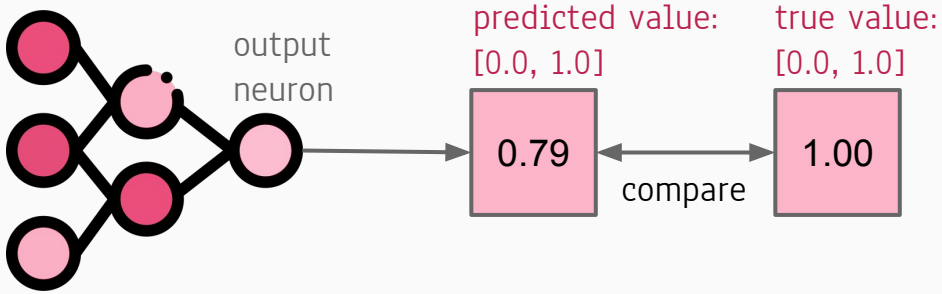
where

$\hat{y}$ - predicted value

y - expected value

Binary classification is a prediction algorithm where the output can be either one of two items, indicated by 0 or 1.

Binary cross entropy measures how far away the true value (which is either 0 or 1) is from the prediction value for each of the classes and then averages these class-wise errors to obtain the final loss.



SureStart

CROSS-ENTROPY: MULTI-CLASS CROSS-ENTROPY LOSS

Both categorical cross entropy and sparse categorical cross-entropy have the same loss function. The only difference between the two is on how truth labels are defined.

Target Feature	Encoding	Tiger	Cat	Airplane
Category		□	□	□
Tiger	→	1	0	0
Cat	→	0	1	0
Airplane	→	0	0	1
Cat	→	0	1	0

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:
tiger, cat, airplane = [1,0,0], [0,1,0], [0,0,1]

Multi-Class Cross-Entropy Loss has a sparse variant.

In **sparse categorical cross-entropy**, truth labels are integer encoded.

No one-hot encoding of the target variable is required for this loss function. This will perform the same cross-entropy calculation of error, without requiring that the target variable be one hot encoded prior to training.

SureStart