

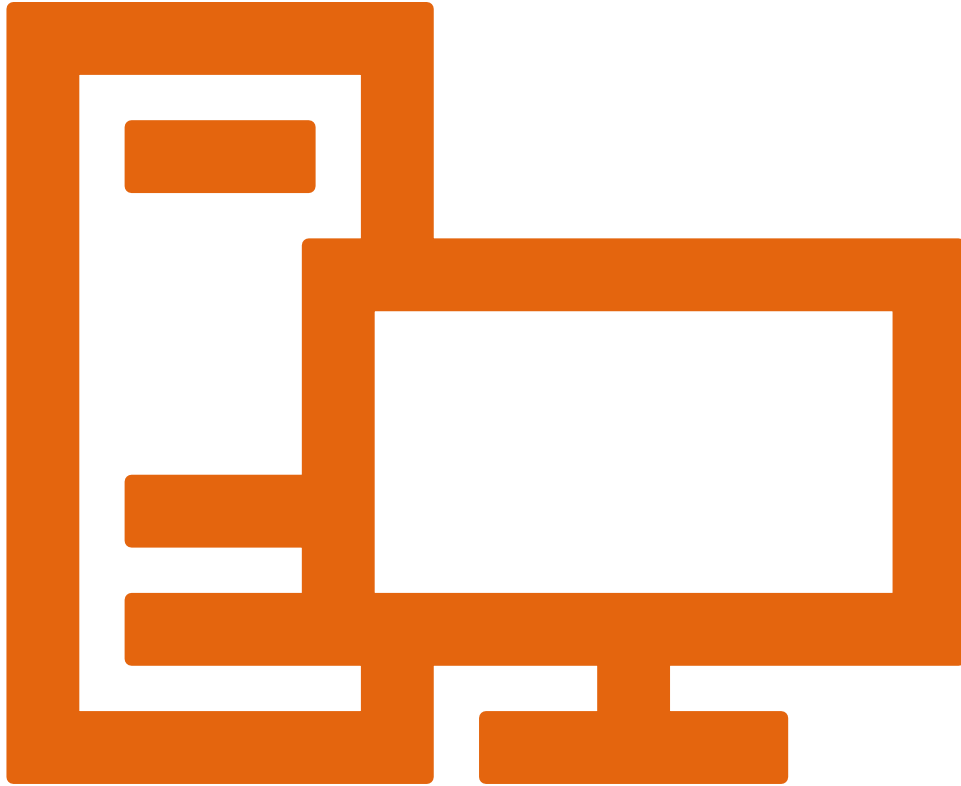
---

- **Classification Algorithm in Machine Learning**

Dr. Jagendra Singh



**Machine Learning**



# Classification Algorithm

---

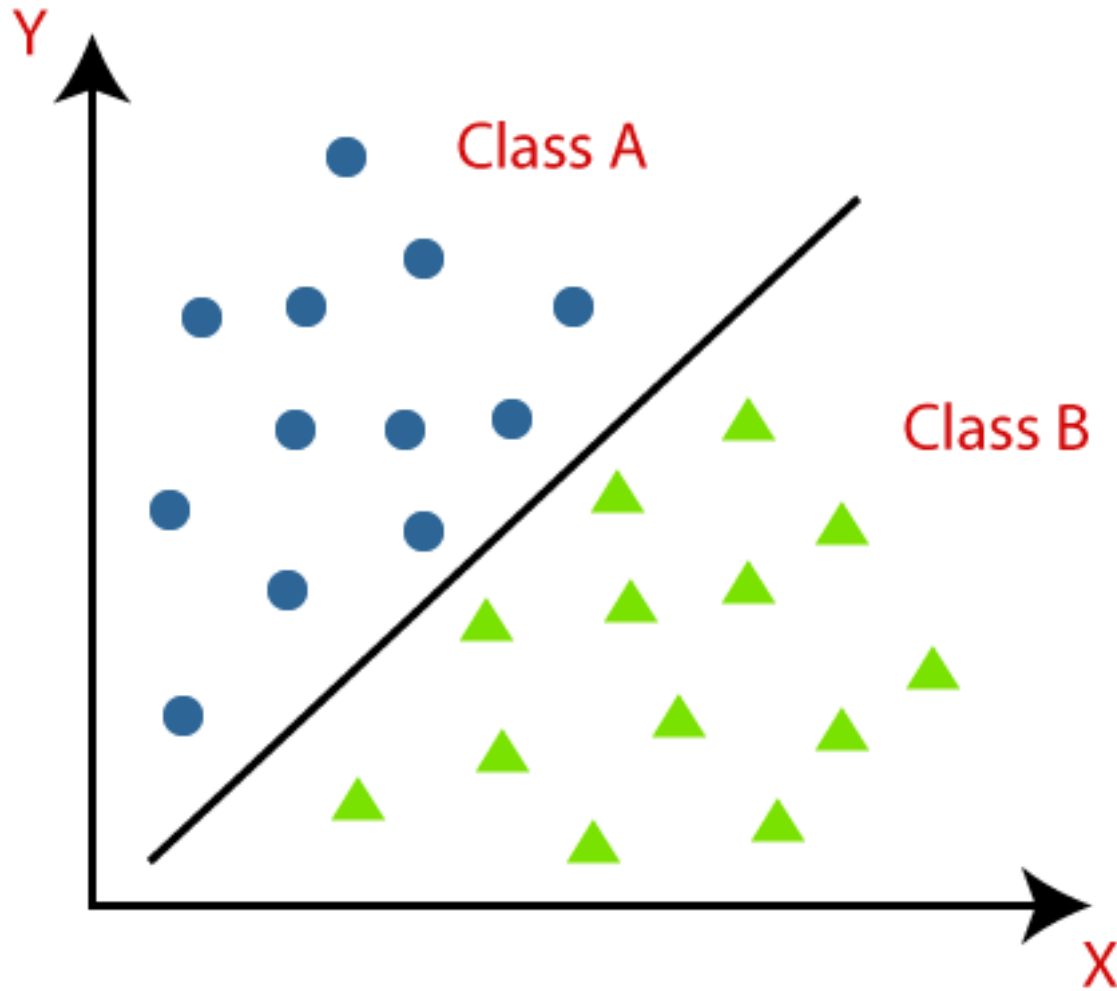
- As we know, the Supervised Machine Learning algorithm can be broadly classified into Regression and Classification Algorithms.
- In Regression algorithms, we have predicted the output for continuous values, but to predict the categorical values, we need Classification algorithms.



# CLASSIFICATION ALGORITHM

---

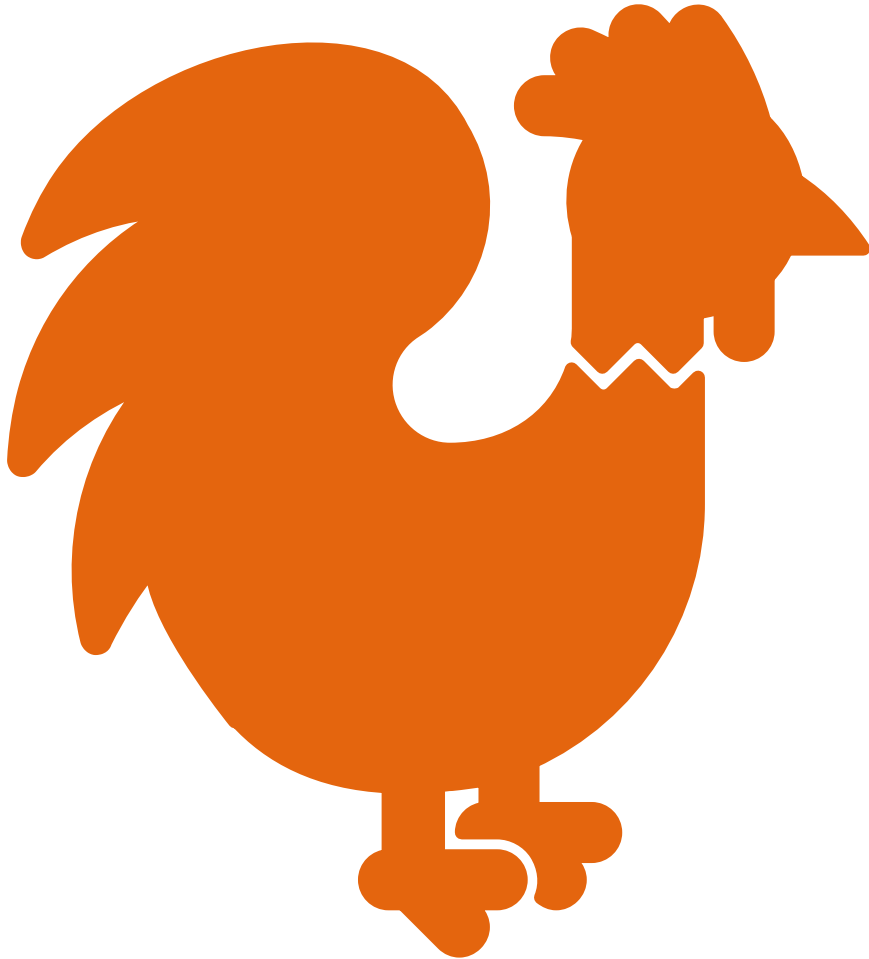
- In Classification, a program learns from the given dataset or observations and then classifies new observation into a number of classes or groups. Such as, **Yes or No, 0 or 1, Spam or Not Spam, cat or dog**, etc.
- Classes can be called as targets/labels or categories
- Unlike regression, the output variable of Classification is a category, not a value, such as "Green or Blue", "fruit or animal", etc.



## CLASSIFICATION ALGORITHM

---

- The best example of an ML classification algorithm is **Email Spam Detector**.
- Classification algorithms can be better understood using the below diagram.
- In this diagram, there are two classes, class A and Class B. These classes have features that are similar to each other and dissimilar to other classes.



# CLASSIFICATION ALGORITHM

---

- The algorithm which implements the classification on a dataset is known as a classifier. There are two types of Classifications:
  - **Binary Classifier:** If the classification problem has only two possible outcomes, then it is called as Binary Classifier.  
**Examples:** YES or NO, MALE or FEMALE, SPAM or NOT SPAM, CAT or DOG, etc.
  - **Multi-class Classifier:** If a classification problem has more than two outcomes, then it is called as Multi-class Classifier.  
**Example:** Classifications of types of crops, Classification of types of music.

# LEARNERS IN CLASSIFICATION PROBLEMS

---

In the classification problems, there are two types of learners:

- Lazy Learners: Lazy Learner firstly stores the training dataset and wait until it receives the test dataset. In Lazy learner case, classification is done on the basis of the most related data stored in the training dataset. It takes less time in training but more time for predictions.  
Example: K-NN algorithm, Case-based reasoning
- Eager Learners: Eager Learners develop a classification model based on a training dataset before receiving a test dataset. Opposite to Lazy learners, Eager Learner takes more time in learning, and less time in prediction. Example: Decision Trees, Naïve Bayes, ANN.

# TYPES OF ML CLASSIFICATION ALGORITHMS

Classification Algorithms can be further divided into the Mainly two category:

## Linear Models

- Logistic Regression
- Support Vector Machines

## Non-linear Models

- K-Nearest Neighbours
- Kernel SVM
- Naïve Bayes
- Decision Tree Classification
- Random Forest Classification

# EVALUATING A CLASSIFICATION MODEL

---

Once our model is completed, it is necessary to evaluate its performance; either it is a Classification or Regression model. So for evaluating a Classification model, we have the following ways:

## 1. Log Loss or Cross-Entropy Loss:

- It is used for evaluating the performance of a classifier, whose output is a probability value between the 0 and 1.
- For a good binary Classification model, the value of log loss should be near to 0.
- The value of log loss increases if the predicted value deviates from the actual value.
- The lower log loss represents the higher accuracy of the model.



# EVALUATING A CLASSIFICATION MODEL

- Cross-entropy loss is used when adjusting model weights during training. The aim is to minimize the loss, i.e, the smaller the loss the better the model. A perfect model has a cross-entropy loss of 0.
- For Binary classification, cross-entropy can be calculated as:

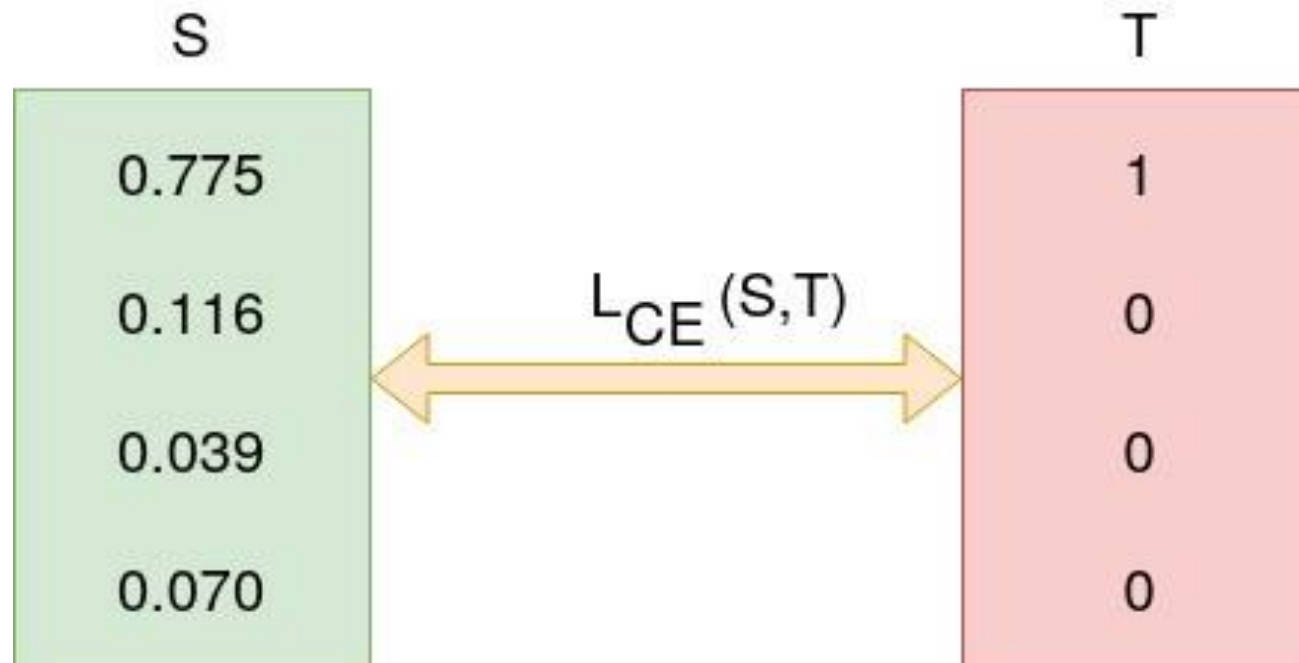
$$L_{\text{CE}} = - \sum_{i=1}^n t_i \log(p_i), \text{ for } n \text{ classes,}$$

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

# EVALUATING A CLASSIFICATION MODEL

- **Example**

- Consider the classification problem with the following Softmax probabilities (S) and the labels (T). The objective is to calculate for cross-entropy loss given these information.



## EVALUATING A CLASSIFICATION MODEL

- The categorical cross-entropy is computed as follows:

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

# EVALUATING A CLASSIFICATION MODEL: CONFUSION MATRIX

---

|                           | <b>Actual Positive</b> | <b>Actual Negative</b> |
|---------------------------|------------------------|------------------------|
| <b>Predicted Positive</b> | True Positive          | False Positive         |
| <b>Predicted Negative</b> | False Negative         | True Negative          |

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{Total Population}}$$

- The confusion matrix provides us a matrix/table as output and describes the performance of the model.
- It is also known as the error matrix.
- The matrix consists of predictions result in a summarized form, which has a total number of correct predictions and incorrect predictions. The matrix looks like as this table:

# EVALUATING A CLASSIFICATION MODEL: AUC-ROC CURVE

---

- ROC curve stands for **Receiver Operating Characteristics Curve** and AUC stands for **Area Under the Curve**.
- It is a graph that shows the performance of the classification model at different thresholds.
- To visualize the performance of the multi-class classification model, we use the AUC-ROC Curve.
- The ROC curve is plotted with TPR and FPR, where TPR (True Positive Rate) on Y-axis and FPR(False Positive Rate) on X-axis.

# USE CASES OF CLASSIFICATION ALGORITHMS

- Classification algorithms can be used in different places. Below are some popular use cases of Classification Algorithms:
  - Email Spam Detection
  - Speech Recognition
  - Identifications of Cancer tumor cells.
  - Drugs Classification
  - Biometric Identification, etc.



THANK YOU

---