



# Sylhet International University

Department of Computer Science & Engineering

Course Title: Machine Learning

Course Code : CSE 463

Md. Umor Faruk Jahangir

B.Sc(Engg) in CSE

Sylhet International University

M.Sc(Thesis) in CSE (Ongoing)

Shahjalal University of Science and Technology

# Types of Machine Learning

```
graph TD; A[Types of Machine Learning] --> B[Supervised ML]; A --> C[Unsupervised ML]; A --> D[Reinforcement ML];
```

Supervised ML

Unsupervised ML

Reinforcement ML

# Supervised Machine Learning

Supervised Learning is the one, where you can consider the learning is guided by a teacher.

We have a dataset which acts as a teacher and its role is to train the model or the machine.

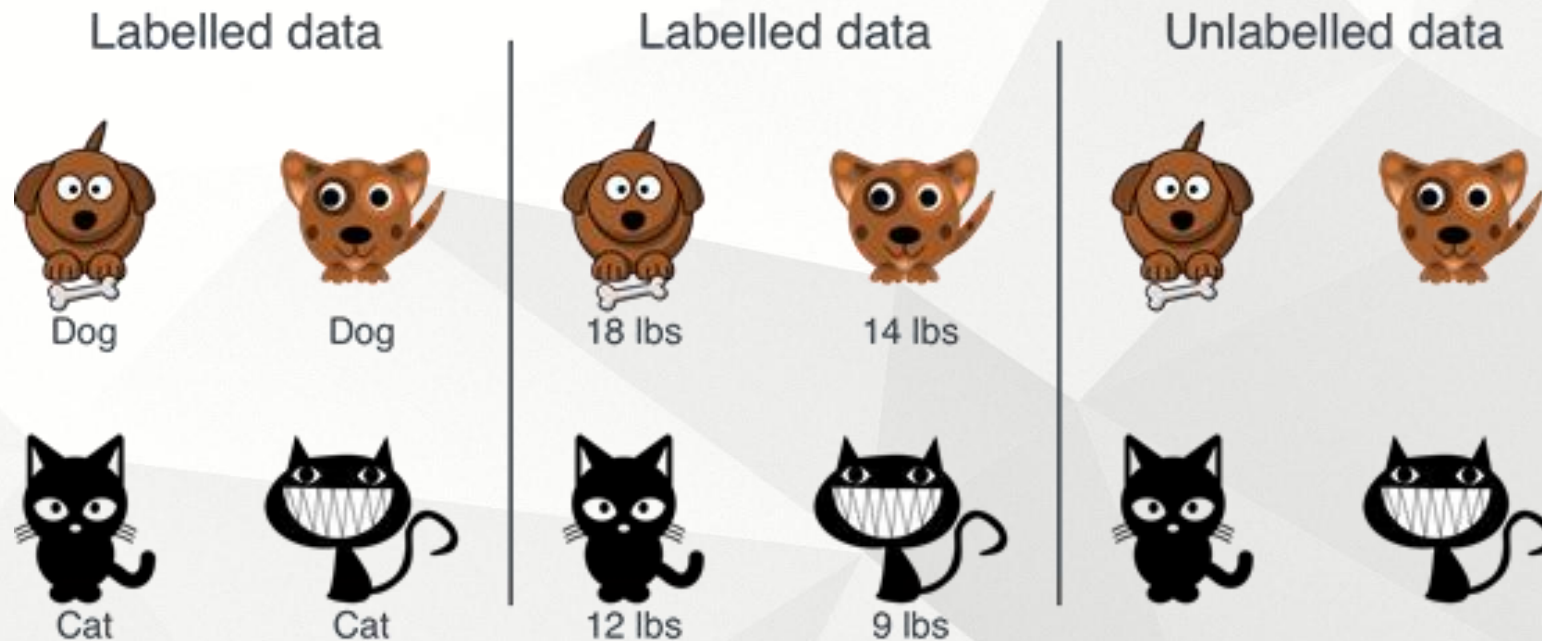
Once the model gets trained it can start making a prediction or decision when new data is given to it.

## Note:

- ❖ Labeled Data
- ❖ Direct Feedback
- ❖ Predict Outcome

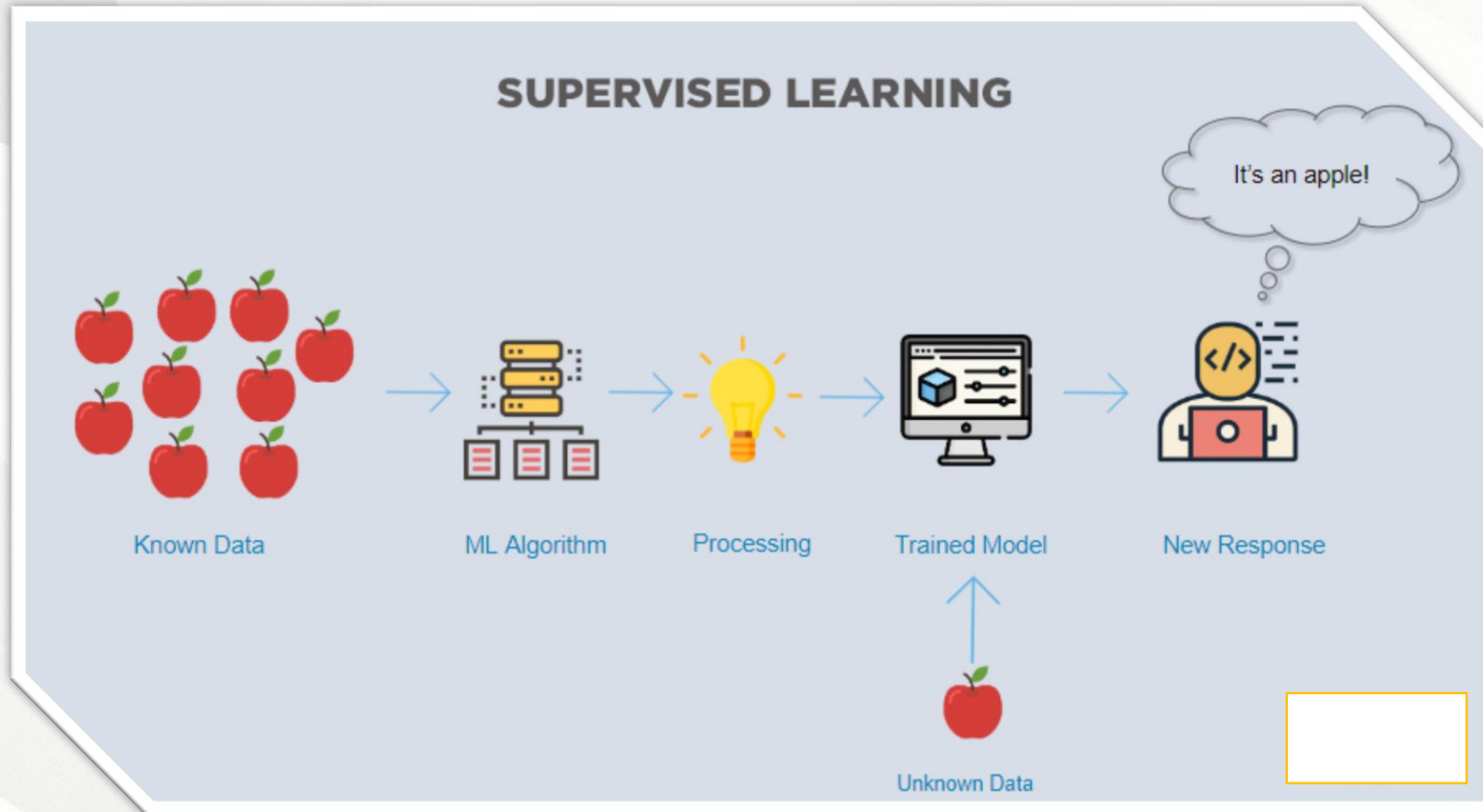
# Supervised Machine Learning

Labeled data: Data consisting of a set of training examples, where each example is a pair consisting of an input and a desired output value (also called the supervisory signal, labels, etc)



# Supervised Machine Learning

Slide - 02



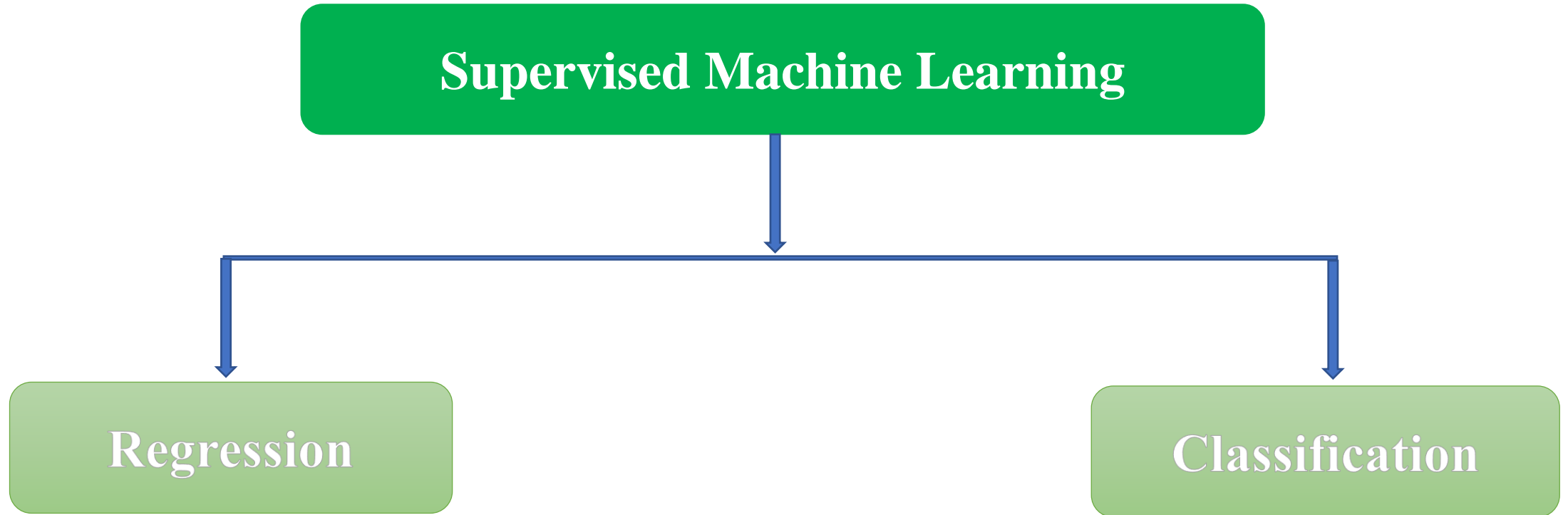


# Supervised Machine Learning

- ❑ Cortana or any speech automated system in your mobile phone trains your voice and then starts working based on this training.
- ❑ Based on various features (past record of head-to-head, pitch, toss, player-vs-player) predicts the winning % of both teams.
- ❑ Train your handwriting to OCR system and once trained, it will be able to convert your hand-writing images into text (till some accuracy obviously)
- ❑ Based on past information about spams, filtering out a new incoming email into **Inbox** (normal) or **Junk folder** (Spam)

.....

# Supervised Machine Learning



# Supervised Machine Learning

**Regression** analysis is a statistical method to model the relationship between a dependent (target) and independent (predictor) variables with one or more independent variables. More specifically, Regression analysis helps us to understand how the value of the dependent variable is changing corresponding to an independent variable when other independent variables are held fixed. It predicts continuous/real values such as **temperature, age, salary, price**, etc.



# Supervised Machine Learning

Terminologies Related to the **Regression** Analysis:

**Dependent Variable:** The main factor in Regression analysis which we want to predict or understand is called the dependent variable. It is also called **target variable**.

**Independent Variable:** The factors which affect the dependent variables or which are used to predict the values of the dependent variables are called independent variable, also called as a **predictor**.

**Outliers:** Outlier is an observation which contains either very low value or very high value in comparison to other observed values. An outlier may hamper the result, so it should be avoided.

# Supervised Machine Learning

Terminologies Related to the **Regression** Analysis:

**Multi-collinearity:** If the independent variables are highly correlated with each other variables, then such condition is called Multi-collinearity. It should not be present in the dataset, because it creates problem while ranking the most affecting variable.

**Underfitting and Overfitting:** If our algorithm works well with the training dataset but not well with test dataset, then such problem is called **Overfitting**. And if our algorithm does not perform well even with training dataset, then such problem is called **underfitting**.

# Supervised Machine Learning

## Regression

Advertisement	Sales
\$90	\$1000
\$120	\$1300
\$150	\$1800
\$100	\$1200
\$130	\$1380
\$200	??

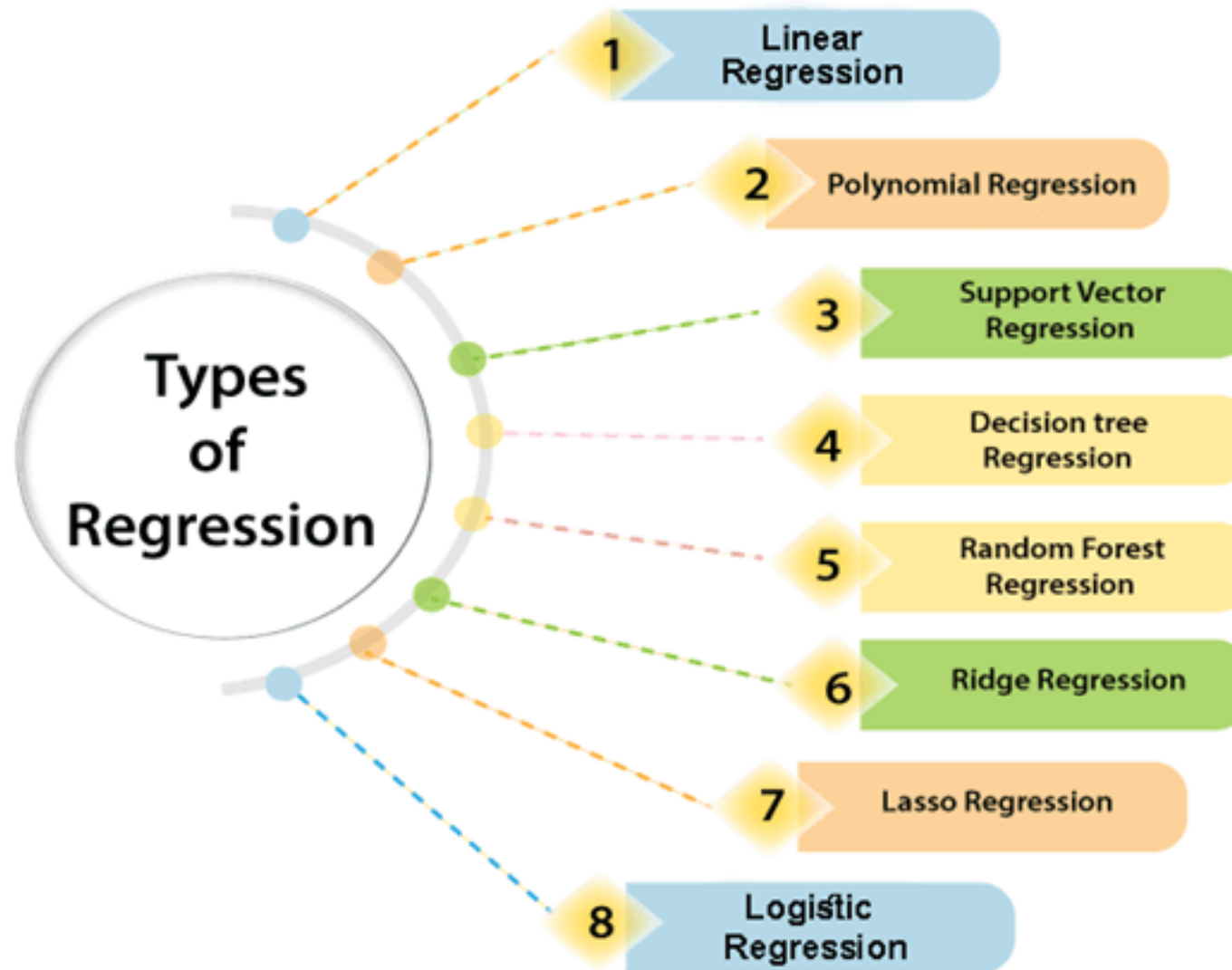
# **Supervised Machine Learning**

## **Regression**

Some examples of regression can be as:

- ☐ Prediction of rain using temperature and other factors
- ☐ Determining Market trends
- ☐ Prediction of road accidents due to rash driving.

# Supervised Machine Learning



**Linear regression** algorithm shows a linear relationship between a dependent ( $y$ ) and one or more independent ( $x$ ) variables, hence called as linear regression. Since linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable.

### **Simple Linear Regression:**

If a single independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Simple Linear Regression.



## Simple Linear Regression:

Slide - 02

$$y = m * x + c$$

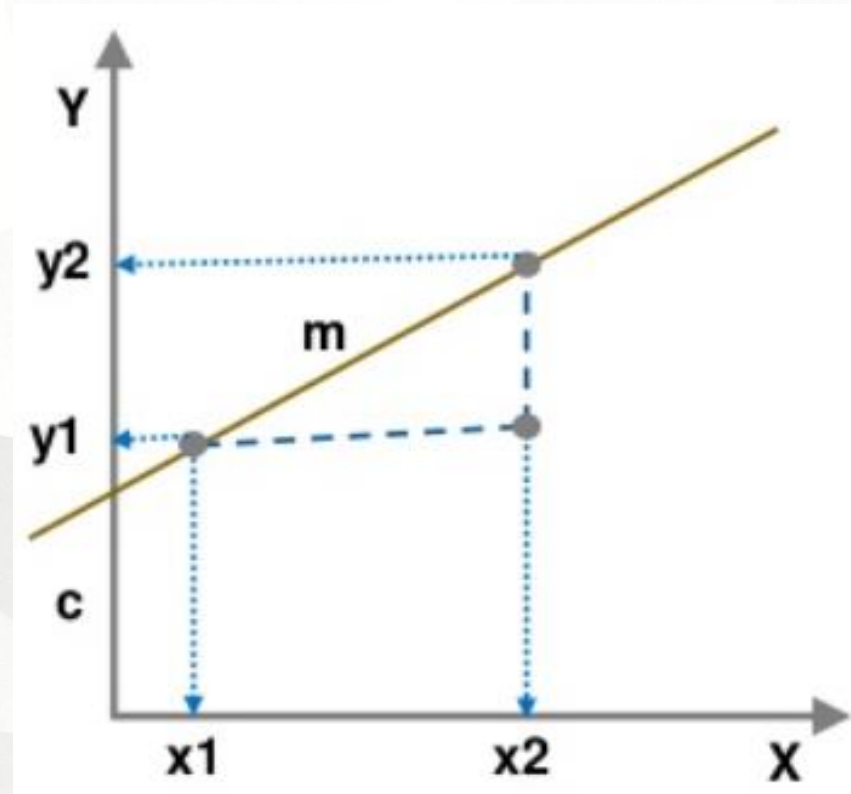
$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

y ---> Dependent Variable

x ---> Independent Variable

m ---> Slope of the line

c ---> Coefficient of the line



# Application of Linear Regression:

Slide - 02



Economic Growth



Product price



Score Prediction



Housing sales

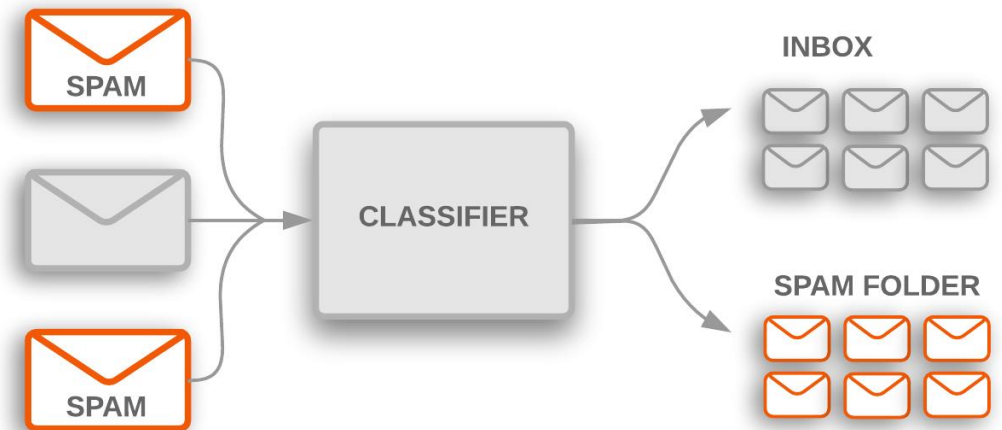
# Supervised Machine Learning

**Classification** is the process of predicting the class of given data points. Classes are sometimes called as targets/ labels or categories. Classification predictive modeling is the task of approximating a mapping function ( $f$ ) from input variables ( $X$ ) to discrete output variables ( $y$ ).

# Supervised Machine Learning

## Classification

- Speech Recognition
- Identifications of Cancer tumor cells, etc



## **Supervised Machine Learning(Classification)**

The algorithm which implements the classification on a dataset is known as a **classifier**.

**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:

**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 learners take less time in training and more time in prediction. **Example:** Decision Trees, Naïve Bayes, ANN.



# Model evaluation procedures

❖ Training and testing on the same data

❖ Train/test split

1. Split the dataset into two pieces, so that the model can be trained and tested on different data
2. Better estimate of out-of-sample performance, but still a "high variance" estimate
3. Useful due to its speed, simplicity, and flexibility

# Cross-validation

Cross-validation is a technique in which we train our model using the subset of the data-set and then evaluate using the complementary subset of the data-set.

The three steps involved in cross-validation are as follows :

- Reserve some portion of sample data-set.
- Using the rest data-set train the model.
- Test the model using the reserve portion of the data-set.

# K-Fold Cross Validation

We split the data-set into k number of subsets(known as folds) then we perform training on the all the subsets but leave one( $k-1$ ) subset for the evaluation of the trained model. In this method, we iterate k times with a different subset reserved for testing purpose each time.



## Special Thanks

- <https://towardsdatascience.com/>
- <https://www.edureka.co/>
- <https://www.simplilearn.com/>
- <https://www.geeksforgeeks.org/>
- <https://www.kdnuggets.com/>
- <https://www.javatpoint.com/>

**THANK YOU!**