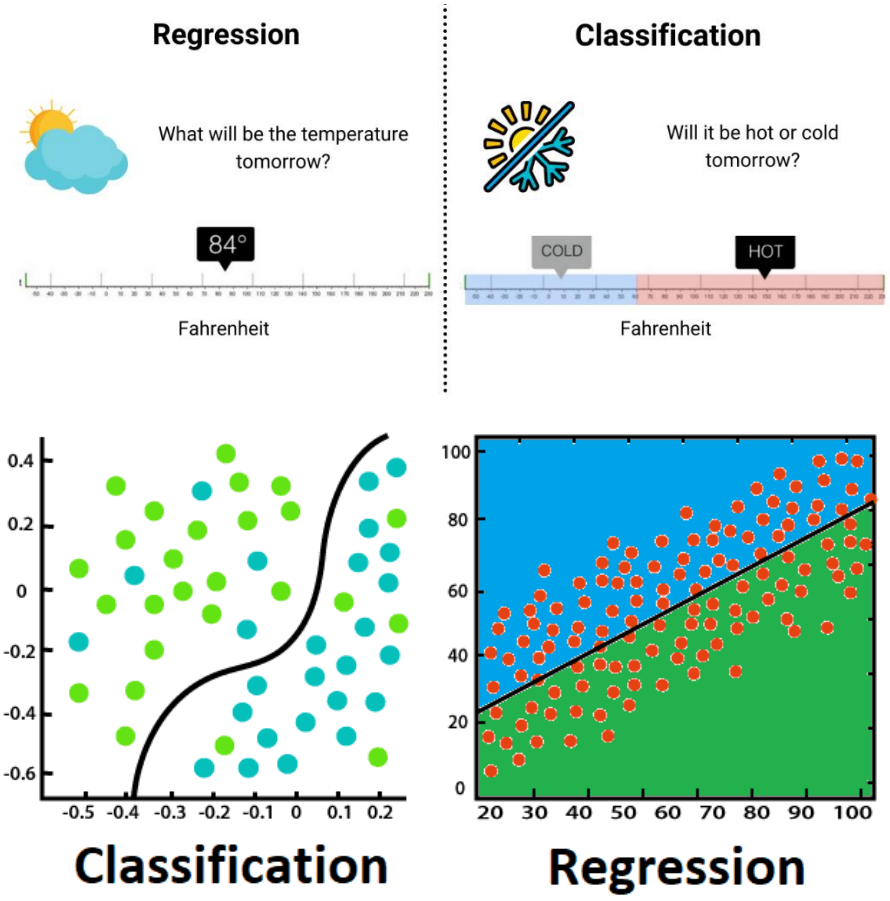
Name: Patricia Stanley

MACHINE LEARNING ASSIGNMENT - 2

Classification vs Regression in Machine Learning

Classification and **Regression** are two major prediction problems that are usually dealt with in Data Mining and Machine Learning.



Classification Algorithms:

Classification is the process of finding or discovering a model or function that helps in separating the data into multiple categorical classes i.e. discrete values. In classification, data is categorized under different labels according to some parameters given in the input and then the labels are predicted for the data.

- In a classification task, we are supposed to predict discrete target variables(class labels) using independent features.
- In the classification task, we are supposed to find a decision boundary that can separate the different classes in the target variable.

Regression Algorithms:

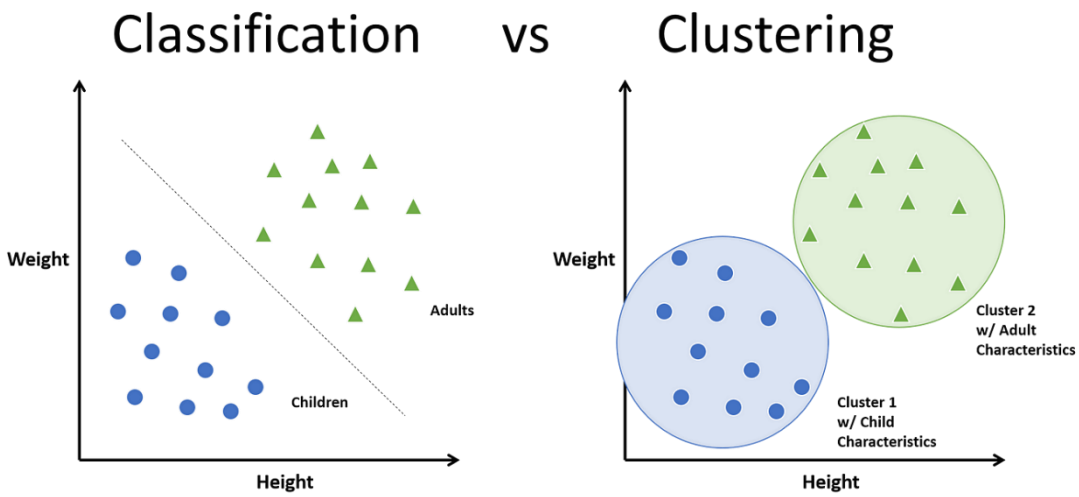
Regression is the process of finding a model or function for distinguishing the data into continuous real values instead of using classes or discrete values. It can also identify the distribution movement depending on the historical data. Because a regression predictive model predicts a quantity, therefore, the skill of the model must be reported as an error in those predictions.

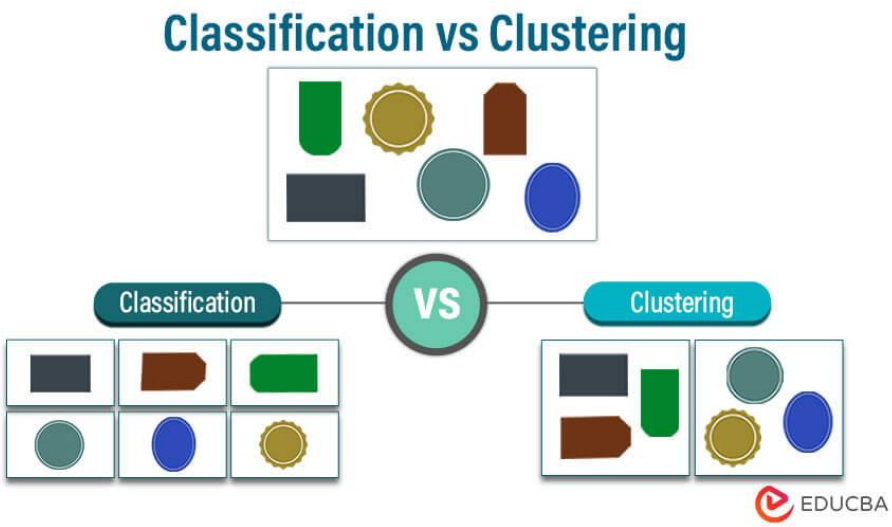
- In a regression task, we are supposed to predict a continuous target variable using independent features.
- In the regression tasks, we are faced with generally two types of problems linear and non-linear regression.

Comparison between Classification and Regression:

CLASSIFICATION	REGRESSION
In this problem statement, the target variables are discrete.	In this problem statement, the target variables are continuous.
In this algorithm, we try to find the best possible decision boundary which can separate the two classes with the maximum possible separation.	In this algorithm, we try to find the best-fit line which can represent the overall trend in the data.
Evaluation metrics like Precision, Recall, and F1-Score are used here to evaluate the performance of the classification algorithms.	Evaluation metrics like Mean Squared Error, R2-Score, and MAPE are used here to evaluate the performance of the regression algorithms.
Input Data are Independent variables and categorical dependent variable.	Input Data are Independent variables and continuous dependent variable.
Output is Categorical labels.	Output is Continuous numerical values.
Objective is to Predict categorical/class labels.	Objective is to Predicting continuous numerical values.
Example use cases are Spam detection, image recognition, sentiment analysis	Example use cases are Stock price prediction, house price prediction, demand forecasting.
Examples of classification algorithms are: Logistic Regression, Decision Trees, Random Forest, Support Vector Machines (SVM), K-Nearest Neighbors (K-NN), Naive Bayes, Neural Networks, K-Means Clustering, Multi-layer Perceptron (MLP), etc.	Examples of regression algorithms are: Linear Regression, Polynomial Regression, Ridge Regression, Lasso Regression, Support Vector Regression (SVR), Decision Trees for Regression, Random Forest Regression, K-Nearest Neighbors (K-NN) Regression, Neural Networks for Regression, etc.

Classification vs Clustering





Classification and **Clustering** is used for the categorization of objects into one or more classes based on the features. They appear to be a similar process as the basic difference is minute. In the case of Classification, there are predefined labels assigned to each input instance according to their properties whereas in clustering those labels are missing.

Comparison between Classification and Clustering:

CLASSIFICATION	CLUSTERING
It is used for supervised learning	It is used for unsupervised learning
The process of classifying the input instances based on their corresponding class labels	Grouping the instances based on their similarity without the help of class labels
It has labels so there is need of training and testing dataset for verifying the model created	There is no need of training and testing dataset
More complex as compared to clustering	Less complex as compared to classification
Example : Logistic regression, Naive Bayes classifier, Support vector machines, etc.	Example: k-means clustering algorithm, Fuzzy c-means clustering algorithm, Gaussian (EM) clustering algorithm, etc.

Examples of Clustering:

- Medical institutions can categorize patients with chronic conditions, such as diabetes, into distinct groups based on the severity of their illness and their response to treatment.

- Streaming services often use clustering analysis to identify viewers who have similar behavior:
  - Minutes watched per day
  - Total viewing sessions per week
  - Number of unique shows viewed per month