Task:- Differences b/w Classification and Clustering, Classification and Regression

Name:- Kush Milkesh Mistry

ID:- MST03-0049

Q1) Classification VS Clustering:- Both classification and clustering involve assigning input instance to a class based on its properties, but there are some differences:

- Classification is a type of supervised learning, whereas clustering is a type of unsupervised learning.
- Classification is the process of grouping the input instances into categories based on their corresponding labels, whereas clustering is the process of grouping the unlabeled input data based on the similarities, differences, and patterns discovered by the model on its own.
- Training and Testing dataset is required in classification, to verify the model created, since it uses labeled data. But in clustering, no training and testing dataset is required, as it uses unlabeled data.
- Classification is more complex than clustering, since there are many levels in the classification phase, but in clustering, only grouping is done.
- Example algorithms in classification include Naïve Bayes Classifier, Support Vector Machines, Logistic Regression, etc., whereas, example algorithms in clustering are:- K-Means Clustering, Gaussian(EM) clustering, Fuzzy C-Means Clustering, etc.
- Example of classification:- If we supply images of cats, dogs, and lions along with their labels as “Cats”, “Dogs”, and “Lions”, respectively to the machine learning model, then the machine learns this mapping, and when we supply a new image of a cat, then the machine predicts the correct label: “Cat”.

  Example of Clustering:- If we supply images of cats, dogs, and lions, but this time, without any corresponding labels to the ML model, and the model has never seen this images nor it has any prior training, then the model is allowed to look for similarities and patterns in these images, and it will create three groups of the images:- one group of all cats, the other of all dogs, and the third of all lions.

- Common applications of classification include email spam detection , medical diagnosis, and image recognition, whereas clustering is used in market segmentation, social network analysis, and organizing large datasets.
- Classification metrics include Accuracy, Precision, Recall and F1 Score, whereas Clustering models are evaluated using metrics such as silhouette score, Davies- Bouldin Index, and within-cluster sum of squares.

Q2) Regression VS Classification:- Both regression and classification are types of supervised learning,

In which the aim is to approximate a function f(x) based on the input training dataset (x1,y1),(x2,y2),(x3,y3),……(xn,yn), where x_i are the features, and y_i are the labels.

- If f(x) is real- valued, then the type of learning is regression, but if f(x) is discrete/categorical, then it is classification.

- Regression models are evaluated using metrics such as Mean Squared Error(MSE), Root Mean Squared Error(RMSE), and R-Squared, whereas metrics for evaluating classification models include Accuracy, Precision, Recall and F1 Score.
- Classification is generally more complex than regression, as classification involves distinguishing between multiple classes, which can require more sophisticated algorithms, but regression typically involves predicting a continuous numerical value, which is usually simpler, especially in cases where the relationship between the variables is more straightforward and linear.
- Common applications of regression include predicting stock prices and house prices, whereas applications of classification include email spam detection and image recognition.
- Algorithms in regression include Linear regression, Ridge regression and Polynomial Regression. Classification algorithms include Logistic Regression and Decision Trees.
- Example of regression:- Suppose we have a dataset of no. of hours that students have studied before exam, and their corresponding final scores. We wish that the model can predict the final score of a student based on the given no. of study hours. Firstly, the input training data points (xi,yi), where xi = no. of study hours of a student, and yi = their final score, are plotted on a graph by the model, then the model has to approximate a function based on the training dataset. Then , when the model is given a test data(no. of study hours), it looks up the function that it has approximated, and outputs the predicted final score.
  The function approximation is done by searching through a hypothesis space for the hypothesis(function) that best fits the training data.

  Example of classification:- If we supply images of cats, dogs, and lions along with their labels as “Cats”, “Dogs”, and “Lions”, respectively to the machine learning model, then the machine learns this mapping, and when we supply a new image of a cat, then the machine predicts the correct label: “Cat”.