CHRIST
UNIVERSITY
BANGALORE, INDIA

MACHINE LEARNING
CAT3 -COMPONENT 2

Submitted by
Name: Aleena Mary Varghese
Class: 4MCA – B
Register Number: 2147239

- **DATA-SET CHOSEN : HISTORY OF ROCK MUSIC 1950-2020**(Over 5400 rock songs spanning over 70 years**)**

- **OVERVIEW:**
  Rock is one of the most popular music genre's today but where did it begin? Rock music first emerged in the late 1940's.
  As a general description, the genre can be described as hard-edged music, performed with electric bass, electric guitar, drums and a vocalist and an injection of high volume and distortion and flamboyant performance. The first rock songs heavily leaned on classic blues structures, but had an aggressive edge to it which proved shocking at the time.

- **NEED : To study the evolution of rock music.**

LAB-1: DATA VISUALIZATION :

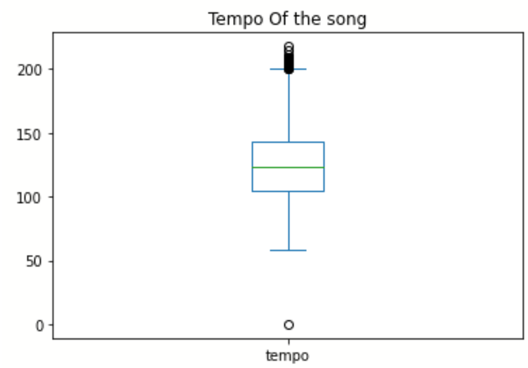
1. **BOXPLOT :**
   - A box plot provides a quartile-based view of the data.
   - A box plot is drawn using a box with boundaries of the box at lower quartile and upper quartile of the distribution. The median value is marked inside the box.

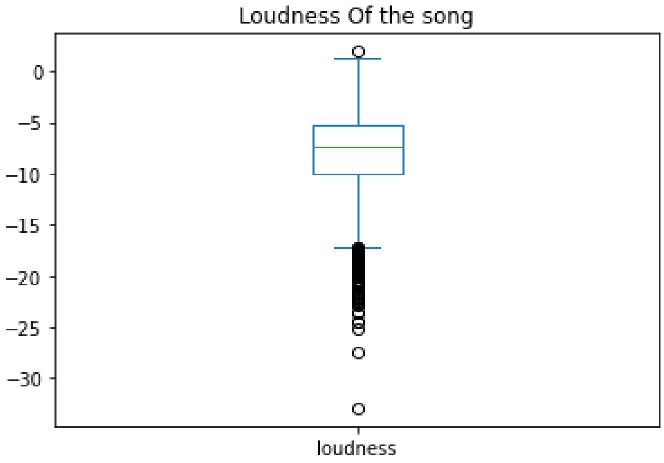
   I have used boxplot to plot the tempo of the song.

   Code :

```
df['tempo'].plot(kind='box', title='Tempo Of the song')

plt.show()
```





A box plot gives us a basic idea of the distribution of the data. If the box plot is relatively short, then the data is more compact. If the box plot is relatively tall, then the data is spread out.

2. **MAP FUNCTION:** is used to substitute each value in a series with another value that may be derived from a function.

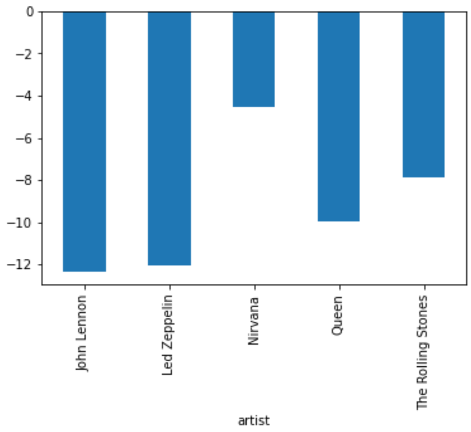
3. **RENAME:** To rename column names.

   I have renamed the column-" name" to "song_name" and column "artist" to "artist_name" inorder to make my dataset precise and clear.

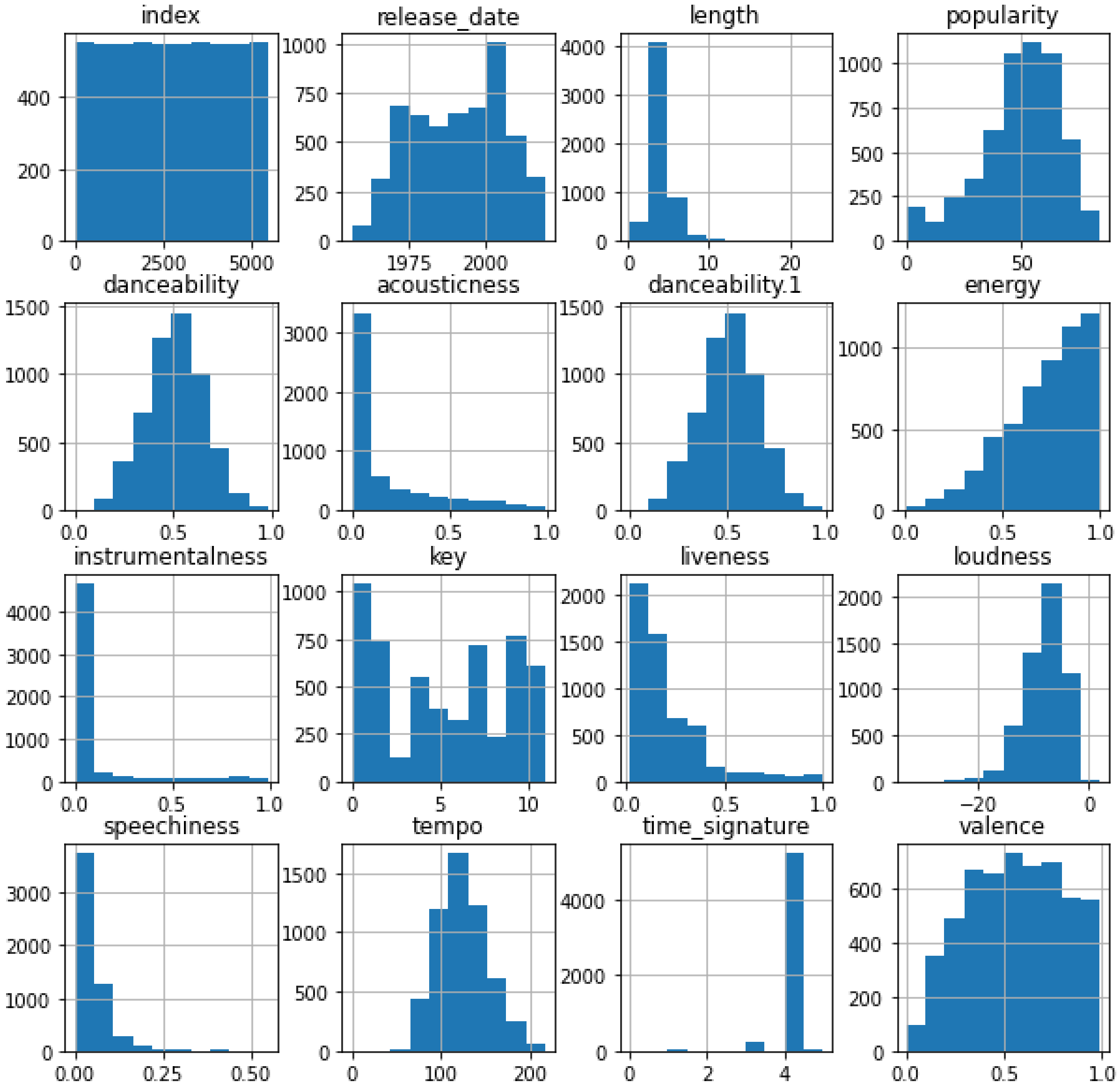
4. **GROUP BY:**

   I tried to group name of the song and artist

5. **BAR:** Used for comparisons among discrete values.



6. **HISTOGRAM:** A histogram though, even in this case, conveniently does the grouping for you. You get values that are close to each other counted and plotted as values of given ranges/bins:



7. **SCATTER() :** We use a scatter plot to determine whether or not two variables have a relationship or correlation.

8. **PAIRPLOT**: A pairs plot is a matrix of scatterplots that lets you understand the pairwise relationship between different variables in a dataset.

9.**DISTPLOT**: To visualize the parametric distribution of a dataset.

10. **CORR()** : It is used to find the pairwise correlation of all columns in the dataframe. Any na values are automatically excluded. For any non-numeric data type columns in the dataframe it is ignored.
The linear relationship between two variables. If one var is inc, other var is incc- +ve else -ve
Normalization : different units of measure- unified and consistent format, make it fall under a smaller range 0 to 1
11. **JOIN PLOT()**: A Jointplot comprises three plots. Out of the three, one plot displays a bivariate graph which shows how the dependent variable(Y) varies with the independent variable(X). Another plot is placed horizontally at the top of the bivariate graph and it shows the distribution of the independent variable(X). The third plot is placed on the right margin of the bivariate graph with the orientation set to vertical and it shows the distribution of the dependent variable(Y). It is very helpful to have univariate and bivariate plots together in one figure. This is because the univariate analysis focuses on one variable, it describes, summarizes and shows any patterns in your data and the bivariate analysis explores the relationship between two variables and also describes the strength of their relationship.

12. **STRIP PLOT()**: A strip plot is a single-axis scatter plot that is used to visualize the distribution of many individual one-dimensional values. The values are plotted as dots along one unique axis, and the dots with the same value can overlap. To show overlapping values, the opacity or color of the dots can be changed, or a jitter plot or counts plot can be used instead. Typically, several strip plots are placed side by side to compare the distributions of data points among several values, categories or ranges.

13.**DATA PLOT()**: This function allows you to plot different charts only by changing the parameters.

14. **VIOLIN PLOT :** A violin plot is a method of plotting numeric data. It is a box plot with a rotated kernel density plot on each side. The violin plot is similar to box plots, except that they also show the probability density of the data at different values.
Typically violin plots will include a marker for the median of the data and a box indicating the interquartile range, as in standard box plots.

LAB- 2 KNN- K-nEIGHBOUR
It is a supervised (It uses known and labeled data as input.) machine learning algorithm. The algorithm can be used to solve both classification and regression problem statements. The number of nearest neighbors to a new unknown variable that has to be predicted or classified is denoted by the symbol 'K'.
K is the number of nearest neighbors to use. For classification, a majority vote is used to determine which class a new observation should fall into. Larger values of *K* are often more

robust to outliers and produce more stable decision boundaries than very small values (*K=3* would be better than *K=1*, which might produce undesirable results.
The KNN algorithm assumes that similar things exist in close proximity. In other words, similar things are near to each other.
KNN captures the idea of similarity (sometimes called distance, proximity, or closeness) with some mathematics we might have learned in our childhood— calculating the distance between points on a graph.

Applying the KNN algorithm on the rock music dataset we notice that there are relationships among the songs that will not be accounted for (artists,year of release etc)when using the KNN algorithm simply because the data that captures those relationships are missing from the data set.
Consequently, when we run the KNN algorithm on our data, similarity will be based solely on the genres by attributes like strength, danceability etc.

The KNN Algorithm

1. Load the data
2. Initialize K to your chosen number of neighbors
3. For each example in the data
4. Calculate the distance between the query example and the current example from the data.
5. Add the distance and the index of the example to an ordered collection
6. Sort the ordered collection of distances and indices from smallest to largest (in ascending order) by the distances
7. Pick the first K entries from the sorted collection
8. Get the labels of the selected K entries
9. If regression, return the mean of the K labels
10. If classification, return the mode of the K labels

Choosing the right value for K

To select the K that's right for your data, we run the KNN algorithm several times with different values of K and choose the K that reduces the number of errors we encounter while maintaining the algorithm's ability to accurately make predictions when it's given data it hasn't seen before.

Advantages

1. The algorithm is simple and easy to implement.
2. There's no need to build a model, tune several parameters, or make additional assumptions.
3. The algorithm is versatile. It can be used for classification, regression, and search (as we will see in the next section).

Disadvantages

1. The algorithm gets significantly slower as the number of examples and/or predictors/independent variables increase.

KNN in practice

KNN's main disadvantage of becoming significantly slower as the volume of data increases makes it an impractical choice in environments where predictions need to be made rapidly. Moreover, there are faster algorithms that can produce more accurate classification and regression results.

However, provided you have sufficient computing resources to speedily handle the data you are using to make predictions, KNN can still be useful in solving problems that have solutions that depend on identifying similar objects. An example of this is using the KNN algorithm in recommender systems, an application of KNN-search.

LAB-3 NAIVE BAYES CLASSIFICATION

Naive Bayes models are probabilistic classifiers that use the Bayes theorem and make a strong assumption that the features of the data are independent. In our case, this means that each artist is independent of others.
The Bayes theorem is defined as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

where ***A*** and ***B*** are some events and ***P(.)*** is a probability.This equation gives us the conditional probability of event ***A*** occurring given ***B*** has happened. In order to find this, we need to calculate the probability of ***B*** happening given ***A*** has happened and multiply that by the probability of ***A*** (known as *Prior*) happening. All of this is divided by the probability of ***B*** happening on its own.

Naive Bayes classifier calculates the probabilities for every factor (here in case of loudness it would be danceability and accosticness for a given input feature). Then it selects the outcome with highest probability.

This classifier assumes the features (in this case we had words as input) are independent. Hence the word naive. Even with this it is powerful algorithm used for

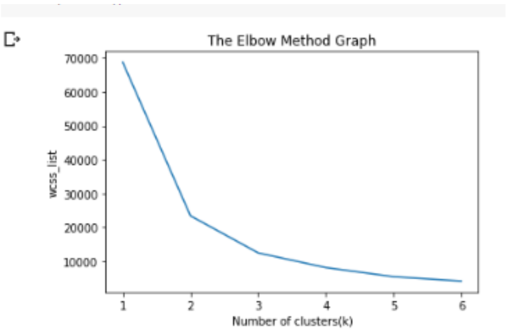
LAB -4 K-MEANS

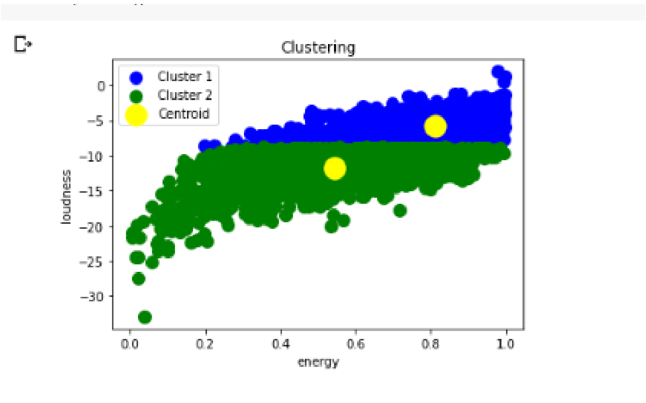
K means algorithm is an iterative algorithm that tries to partition the dataset into *K*pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group. It tries to make the intra-cluster data points as similar as possible while also keeping the clusters as different (far) as possible. It assigns data points to a cluster such that the sum of the squared distance between the data points and the cluster's centroid (arithmetic mean of all the data points that belong to that cluster) is at the minimum. The less variation we have within clusters, the more homogeneous (similar) the data points are within the same cluster.

1. Elbow method: edges- k-number of clusters
2. We distinguish between K-means and K-means++ algorithms in detail. Both K-means and K-means++ are clustering methods which come under unsupervised learning. The main difference between the two algorithms lies in:

- the selection of the centroids around which the clustering takes place
- k means++ removes the drawback of K means which is it is dependent on initialization of centroid
- centroids: A centroid is a point which we assume to be the center of the cluster.

3. It is used to solve unsupervised clustering problems.
4. The first step involves the random initialization of k data points which are called means.
    - In this step we cluster each data point to its nearest mean and after that we update the mean of the current clusters. mean: is the average of a group of values.





In our dataset we come to notice that k=2 is not a bad choice. Sometimes it's still hard to figure out a good number of clusters to use because the curve is monotonically decreasing and may not show any elbow or has an obvious point where the curve starts flattening out.

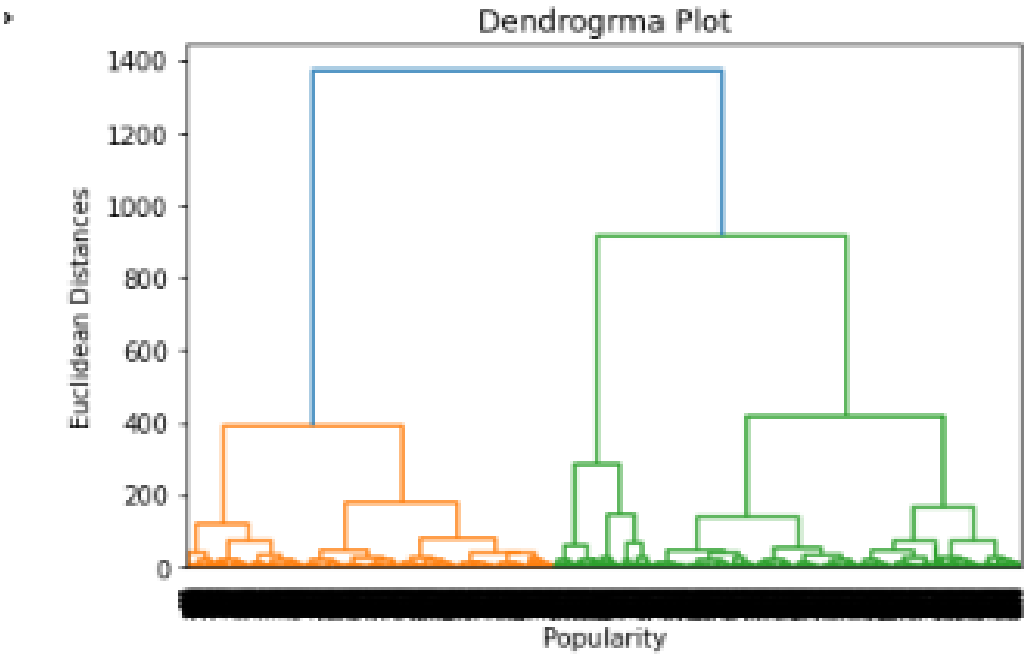
LAB 5 : HIERARCHICAL CLUSTERING

Merge the most similar points or clusters in hierarchical clustering.In hierarchical clustering, we have a concept called a proximity matrix. This stores the distances between each point. A dendrogram is a tree-like diagram that records the sequences of merges or splits.More the distance of the vertical lines in the dendrogram, more the distance between those clusters.

1. AGGLOMERATIVE:
    - Decide the clusters
    - Merge the closest pair of clusters and repeat this step until only a single cluster is left.
2. DIVISIVE
    - Divisive hierarchical clustering works in the opposite way. Instead of starting with n clusters (in case of n observations), we start with a single cluster and assign all the points to that cluster.

Interpretation of dendrogram

Each level of dendrogram has a subtle meaning to the relationship between its data members. In a regular relationship chart, one may interpret that at the top lies grandparents or the first generation, the next level corresponds to parents or second generation and the final level belongs to children or third generation. Likewise, in every branching procedure of dendrogram, all the data points having the membership at each level belongs to a certain class.However, to infer this class entity, one has to go through a few individual samples of each level within the formulated cluster and find out what feature is common in the resulting cluster. Also, these inferred classes need not be similar at the sister branches. For example, in our case , the songs have been clustered on popularity and length but artists have been clustered on a similar attribute of size.



HIERARCHICAL CLUSTERING VS K MEANS

As we know, clustering is a subjective statistical analysis, and there is more than one appropriate algorithm for every dataset and type of problem. So how to choose between K-means and hierarchical?

1. If there is a specific number of clusters in the dataset, but the group they belong to is unknown, choose K-means
2. If the distinguishes are based on prior beliefs, hierarchical clustering should be used to know the number of clusters
3. With a large number of variables, K-means compute faster
4. The result of K-means is unstructured, but that of hierarchal is more interpretable and informative
5. It is easier to determine the number of clusters by hierarchical clustering's dendrogram

LAB-6 DECISION TREES

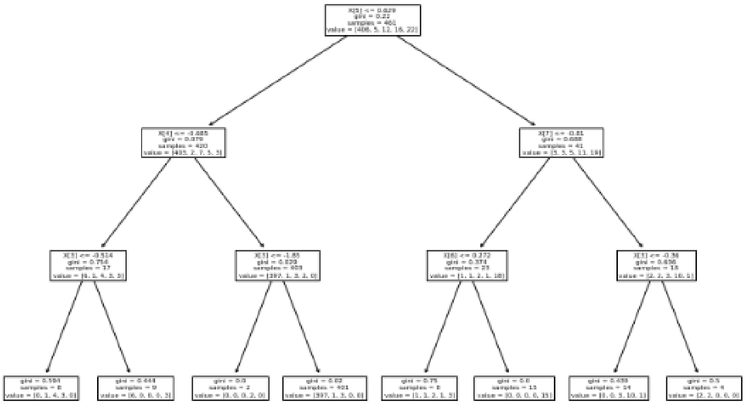
Decision tree is one of the most popular machine learning algorithms used all alongDecision trees are used for both classification and regression problems

Why Decision trees? We have a couple of other algorithms there, so why do we have to choose Decision trees??

1. Decision trees often mimic human level thinking so it's so simple to understand the data and make some good interpretations.

2. Decision trees actually make you see the logic for the data to interpret(not like black box algorithms like SVM,NN,etc..)

*So what is the decision tree??*A decision tree is a tree where each node represents a feature(attribute), each link(branch) represents a decision(rule) and each leaf represents an outcome(categorical or continuous value).The whole idea is to create a tree like this for the entire data and process a single outcome at every leaf(or minimize the error in every leaf).



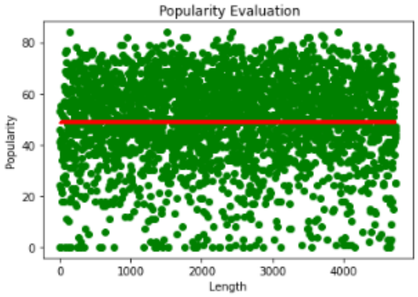
LAB-6.2- LINEAR REGRESSION

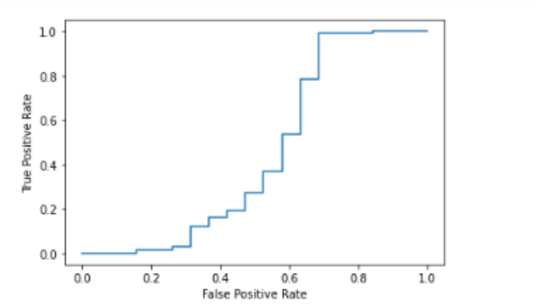
It is a statistical linear machine learning algorithm that is used for predictive analysis. Here, the predicted analysis is continuous and has a constant slope which is used to predict values within a continuous/real range such as salary, age, product, sales, price rather than trying to classify them into categories as petals, cat, dog.

Why switch to Linear Regression? Linear regression algorithms show a linear relationship between a dependent variable, y, and one or more independent variables,x i.e., how the value of the dependent variable, y changes according to the value of the independent variable.

In our case : Popularity is a dependent variable on length which is an explanatory variable, i.e

The popularity evaluation throughout the year depends on the length of the song.





Here are the assumptions that linear regression makes about the data sets it gets applied to:

1. **Autocorrelation:** This assumption takes place when residual errors are dependent on each other and indicates little to no autocorrelation in data.
2. **Multicollinearity:** This assumption happens when independent variables show some dependency. It states that data multicollinearity either doesn't exist at all or is present scarcely.
3. **Variable relationship:** This assumption shows that there is a linear relationship between feature variables and response variables.

LAB- 7 - LOGISTIC REGRESSION

It's a classification algorithm that is used where the response variable is *categorical*. The idea of Logistic Regression is to find a relationship between features and probability of particular outcome.

E.g. When we have to predict if a song is a hit or flop in a billboard chart when we check the popularity rate of each song as a feature, the response variable has two values, hit or flop.

This type of a problem is referred to as **Binomial Logistic Regression**, where the response variable has two values 0 and 1 or pass and fail or true and false. **Multinomial Logistic Regression** deals with situations where the response variable can have three or more possible values.

Why Logistic, not Linear?

With binary classification, let ***'x'*** be some feature and ***'y'*** be the output which can be either 0 or 1. Also, Linear regression has a considerable effect on outliers

Pros

- Simple and efficient.
- Low variance.
- It provides a probability score for observations.

Cons:

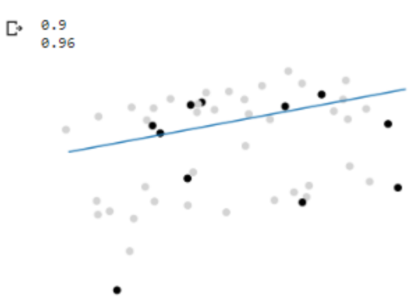
- Doesn't handle **large** numbers of categorical features/variables well.
- It requires transformation of non-linear features.

LAB - 8 - SVM CLASSIFICATION

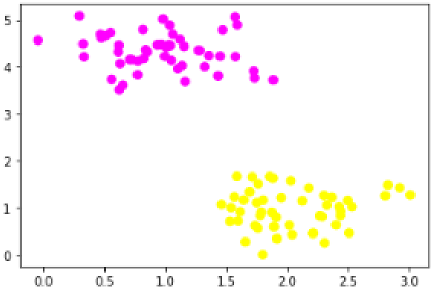
Support vector machines, so called as SVM is a *supervised learning algorithm* which can be used for classification and regression problems as support vector classification (SVC) and support vector regression (SVR). It is used for smaller datasets as it takes too long to process. In this set, we will be focusing on SVC.

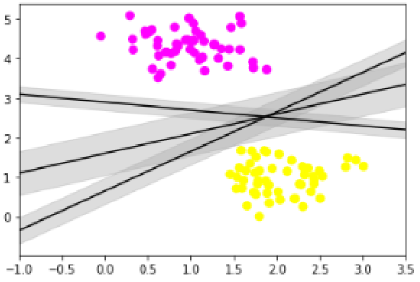
The ideology behind SVM:

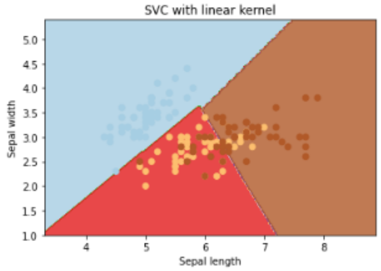
SVM is based on the idea of finding a hyperplane that best separates the features into different domains.

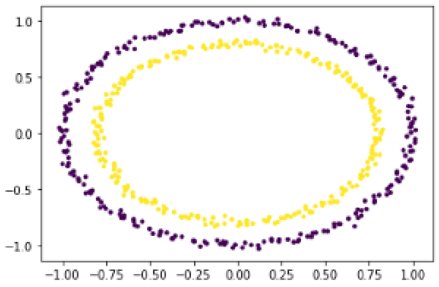


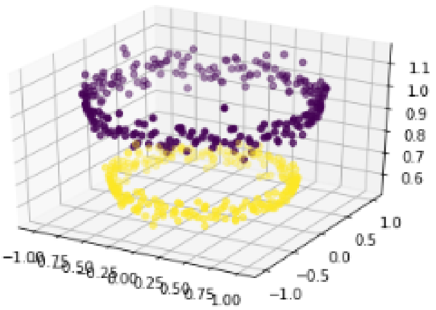
```
0.9
0.96
```

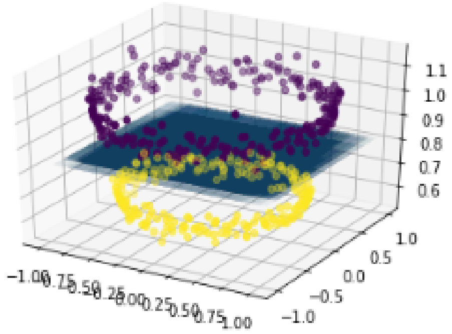












LAB-9 - MULTI LAYER FEED FORWARD NEURAL NETWORK

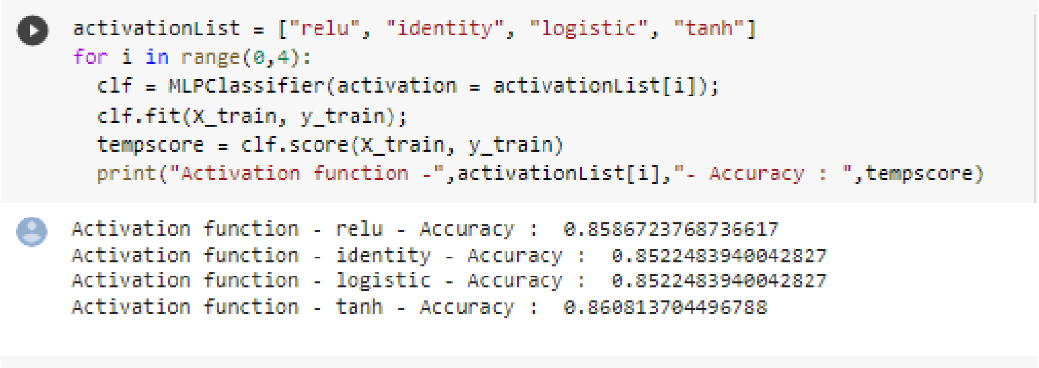
A multi layer perceptron (MLP) is a class of feed forward artificial neural network. MLP consists of at least three layers of nodes: an input layer, a hidden layer and an output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. MLP utilizes a supervised learning technique called back propagation for training.

What is Activation Function? *It's just a thing function that you use to get the output of node. It is also known as* ***Transfer Function****.*

Types of activation functions :

1. Sigmoid activation function
2. Hyperbolic Tangent Function (Tanh)
3. Rectified Linear Unit (ReLU) Function

86.7%which testifies the robustness of MLP classifier as one of the most preferred models for binary classification challenges.



```
activationList = ["relu", "identity", "logistic", "tanh"]
for i in range(0,4):
  clf = MLPClassifier(activation = activationList[i]);
  clf.fit(X_train, y_train);
  tempscore = clf.score(X_train, y_train)
  print("Activation function -",activationList[i],"- Accuracy : ",tempscore)
```

```
Activation function - relu - Accuracy :  0.8586723768736617
Activation function - identity - Accuracy :  0.8522483940042827
Activation function - logistic - Accuracy :  0.8522483940042827
Activation function - tanh - Accuracy :  0.860813704496788
```

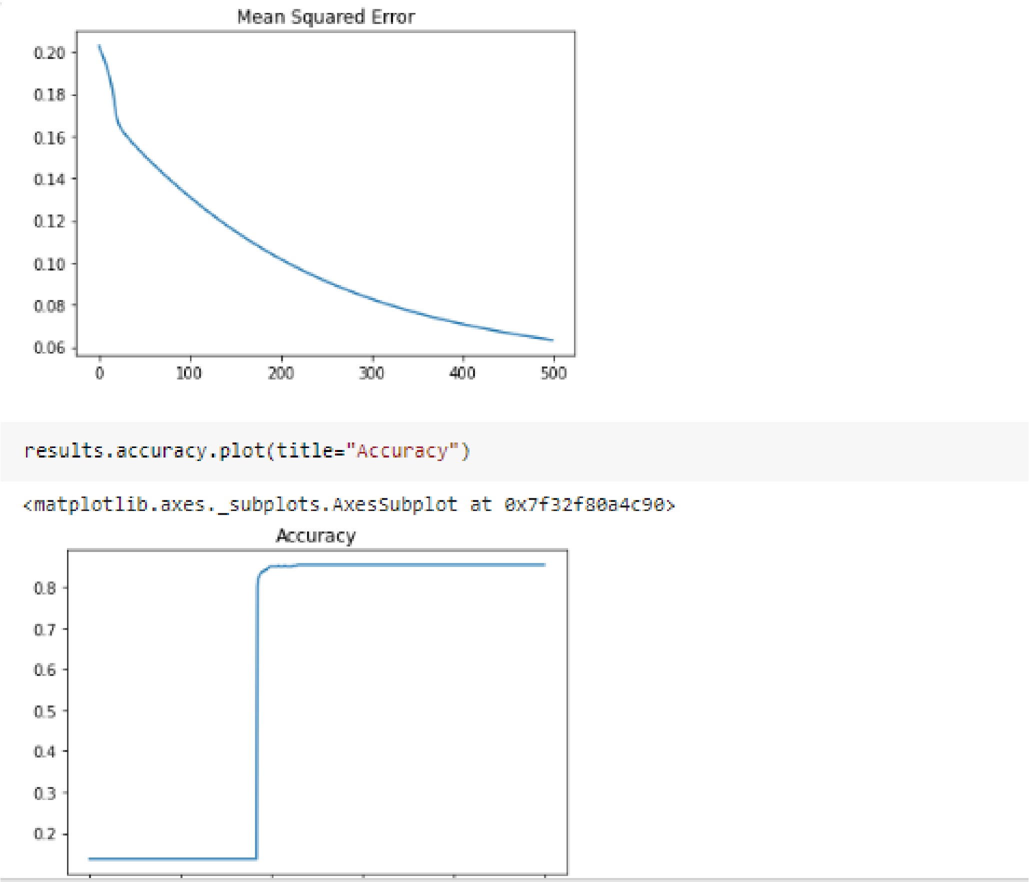
LAB-10- BPN

The Backpropagation algorithm is a supervised learning method for multilayer feed-forward networks from the field of Artificial Neural Networks.

Feed-forward neural networks are inspired by the information processing of one or more neural cells, called a neuron. A neuron accepts input signals via its dendrites, which pass the electrical signal down to the cell body. The axon carries the signal out to synapses, which are the connections of a cell's axon to another cell's dendrites.The principle of the backpropagation approach is to model a given function by modifying internal weightings of input signals to produce an expected output signal. The system is trained using a supervised learning method, where the error between the system's output and a known expected output is presented to the system and used to modify its internal state.

Technically, the backpropagation algorithm is a method for training the weights in a multilayer feed-forward neural network. As such, it requires a network structure to be defined of one or more layers where one layer is fully connected to the next layer. A standard network structure is one input layer, one hidden layer, and one output layer.Backpropagation can be used for both classification and regression problems, but we will focus on classification in this tutorial.

In classification problems, best results are achieved when the network has one neuron in the output layer for each class value.



```
results.accuracy.plot(title="Accuracy")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f32f80a4c90>
```