

**RAJALAKSHMI ENGINEERING COLLEGE**  
**RAJALAKSHMI NAGAR, THANDALAM – 602 105**



**RAJALAKSHMI**  
**ENGINEERING COLLEGE**

**A123331**

**FUNDAMENTALS OF MACHINE LEARNING**

**LABORATORY LAB MANUAL**

Name : **DONEESWARAN J** .....

Year / Branch / Section : **II YEAR / AIML / A** .....

Register No. : **231501502** .....

Semester : **III SEMESTER** .....

Academic Year : **2024-2025** .....

## INDEX

REG. NO : 231501502

NAME: DONEESWARAN J

YEAR : II YEAR

BRANCH: AIML      SEC: A

S.NO.	DATE	TITLE	PAGE NO.	TEACHER'S SIGNATURE REMARKS
1.	01/08/24	A PYTHON PROGRAM TO IMPLEMENT UNIVARIATE, BIVARIATE AND MULTIVARIATE REGRESSION	3	
2.	22/08/24	A PYTHON PROGRAM TO IMPLEMENT SIMPLE LINEAR REGRESSION USING LEAST SQUARE METHOD	13	
3.	29/08/24	A PYTHON PROGRAM TO IMPLEMENT LOGISTIC MODEL	20	
4.	05/10/24	A PYTHON PROGRAM TO IMPLEMENT SINGLE LAYER PERCEPTRON	29	
5.	12/09/24	A PYTHON PROGRAM TO IMPLEMENT MULTILAYER PERCEPTRON WITH BACK PROPOGATION	33	
6.	19/09/24	A PYTHON PROGRAM TO IMPLEMENT SVM CLASSIFIER MODEL	41	
7.	03/10/24	A PYTHON PROGRAM TO IMPLEMENT DECISION TREE	48	
8.	10/10/24	A PYTHON PROGRAM TO IMPLEMENT ADA BOOSTING	55	
9 A	17/10/24	A PYTHON PROGRAM TO IMPLEMENT KNN MODEL	66	
9 B	24/10/24	A PYTHON PROGRAM TO IMPLEMENT K-MEANS MODEL	73	
10.	07/11/24	A PYTHON PROGRAM TO IMPLEMENT DIMENSIONALITY REDUCTION USING PCA	78	

**EXPERIMENT NO : 1**

**DATE : 01/08/24**

**REGISTER NO : 231501502**

**NAME : DONEESWARAN J**

---

## **A PYTHON PROGRAM TO IMPLEMENT UNIVARIATE, BIVARIATE AND MULTIVARIATE REGRESION**

### **AIM:**

To implement a python program using univariate, bivariate and multivariate regression features for a given iris dataset.

### **ALGORITHM:**

#### **STEP 1: IMPORT NECESSARY LIBRARIES:**

- pandas for data manipulation, numpy for numerical operations, and matplotlib.pyplot for plotting.

#### **STEP 2: READ THE DATASET:**

- Use the pandas `read\_csv` function to read the dataset.
- Store the dataset in a variable (e.g., `data`).

#### **STEP 3: PREPARE THE DATA:**

- Extract the independent variable(s) (X) and dependent variable (y) from the dataset.
- Reshape X and y to be 2D arrays if needed.

#### **STEP 4:UNIVARIATE REGRESSION:**

- For univariate regression, use only one independent variable.
- Fit a linear regression model to the data using numpy's polyfit function or sklearn's LinearRegression class.
- Make predictions using the model.
- Calculate the R-squared value to evaluate the model's performance.

### **STEP 5: BIVARIATE REGRESSION:**

- For bivariate regression, use two independent variables.
- Fit a linear regression model to the data using numpy's `polyfit` function or sklearn's `LinearRegression` class.
- Make predictions using the model.
- Calculate the R-squared value to evaluate the model's performance.

### **STEP 6: MULTIVARIATE REGRESSION:**

- For multivariate regression, use more than two independent variables.
- Fit a linear regression model to the data using sklearn's `LinearRegression` class.
- Make predictions using the model.
- Calculate the R-squared value to evaluate the model's performance.

### **STEP 7: PLOT THE RESULTS:**

- For univariate regression, plot the original data points (X, y) as a scatter plot and the regression line as a line plot.
- For bivariate regression, plot the original data points (X1, X2, y) as a 3D scatter plot and the regression plane.
- For multivariate regression, plot the predicted values against the actual values.

### **STEP 8: DISPLAY THE RESULTS:**

- Print the coefficients (slope) and intercept for each regression model.
- Print the R-squared value for each regression model.

### **STEP 9: COMPLETE THE PROGRAM:**

- Combine all the steps into a Python program.
- Run the program to perform univariate, bivariate, and multivariate regression on the dataset.

## CODE

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np

df=pd.read_csv('/content/drive/MyDrive/Datasets/iris.csv')
df.head(150)
df.shape
df

df_Setosa=df.loc[df['species']=='setosa']
df_Virginica=df.loc[df['species']=='virginica']
df_Versicolor=df.loc[df['species']=='versicolor']
df_Setosa

#univariate for sepal width
plt.scatter(df_Setosa['sepal_width'],np.zeros_like(df_Setosa['sepal_width']))
plt.scatter(df_Virginica['sepal_width'],np.zeros_like(df_Virginica['sepal_width']
))
plt.scatter(df_Versicolor['sepal_width'],np.zeros_like(df_Versicolor['sepal_widt
h']))
plt.xlabel('sepal_width')
plt.show()

#univariate for sepal length
plt.scatter(df_Setosa['sepal_length'],np.zeros_like(df_Setosa['sepal_length']))
plt.scatter(df_Virginica['sepal_length'],np.zeros_like(df_Virginica['sepal_length
']))
```

```
plt.scatter(df_Versicolor['sepal_length'],np.zeros_like(df_Versicolor['sepal_length']))
```

```
plt.xlabel('sepal_length')
```

```
plt.show()
```

```
#univariate for sepal width
```

```
plt.scatter(df_Setosa['petal_width'],np.zeros_like(df_Setosa['petal_width']))
```

```
plt.scatter(df_Virginica['petal_width'],np.zeros_like(df_Virginica['petal_width']))
```

```
plt.scatter(df_Versicolor['petal_width'],np.zeros_like(df_Versicolor['petal_width']))
```

```
plt.xlabel('petal_width')
```

```
plt.show()
```

```
#univariate for sepal length
```

```
plt.scatter(df_Setosa['petal_length'],np.zeros_like(df_Setosa['petal_length']))
```

```
plt.scatter(df_Virginica['petal_length'],np.zeros_like(df_Virginica['petal_length']))
```

```
plt.scatter(df_Versicolor['petal_length'],np.zeros_like(df_Versicolor['petal_length']))
```

```
plt.xlabel('petal_length')
```

```
plt.show()
```

```
#bivariate sepal.width vs petal.width
```

```
sns.FacetGrid(df,hue='species',height=5).map(plt.scatter,"sepal_width","petal_width").add_legend();
```

```
plt.show()
```

```
#bivariate sepal.length vs petal.length
```

```
sns.FacetGrid(df,hue='species',height=5).map(plt.scatter,"sepal_length","petal_length").add_legend();
```

```
plt.show()
```

```
#multivariate all the features
```


```
sns.pairplot(df,hue='species',size=2)
```



## OUTPUT

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa
...	...	...	...	...	...
145	6.7	3.0	5.2	2.3	virginica
146	6.3	2.5	5.0	1.9	virginica
147	6.5	3.0	5.2	2.0	virginica
148	6.2	3.4	5.4	2.3	virginica
149	5.9	3.0	5.1	1.8	virginica

150 rows x 5 columns

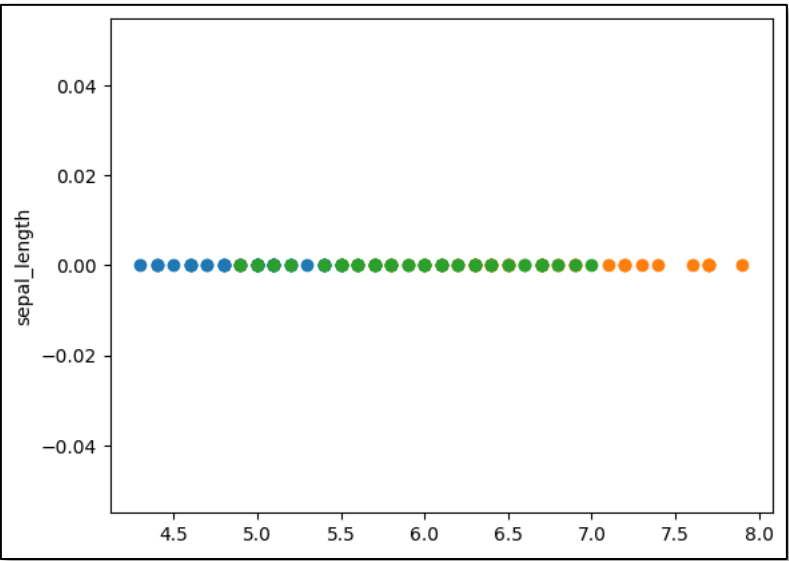
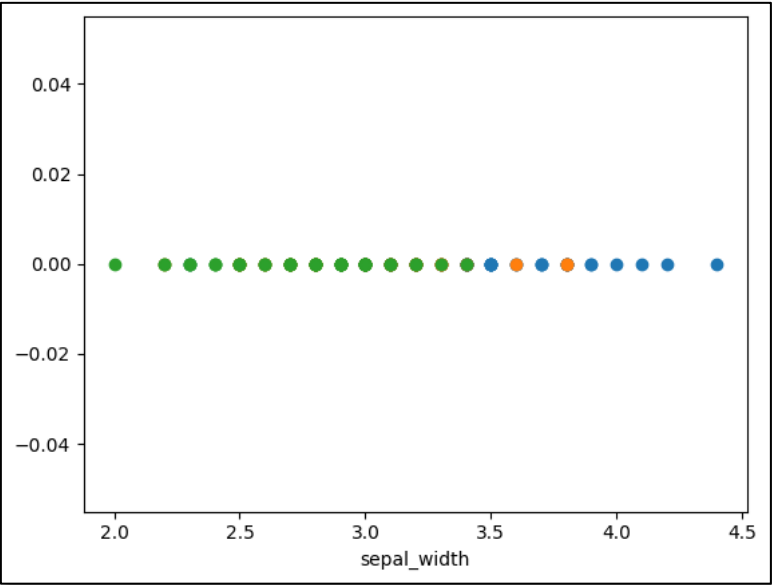
	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa
5	5.4	3.9	1.7	0.4	setosa
6	4.6	3.4	1.4	0.3	setosa
7	5.0	3.4	1.5	0.2	setosa
8	4.4	2.9	1.4	0.2	setosa
9	4.9	3.1	1.5	0.1	setosa
10	5.4	3.7	1.5	0.2	setosa
11	4.8	3.4	1.6	0.2	setosa

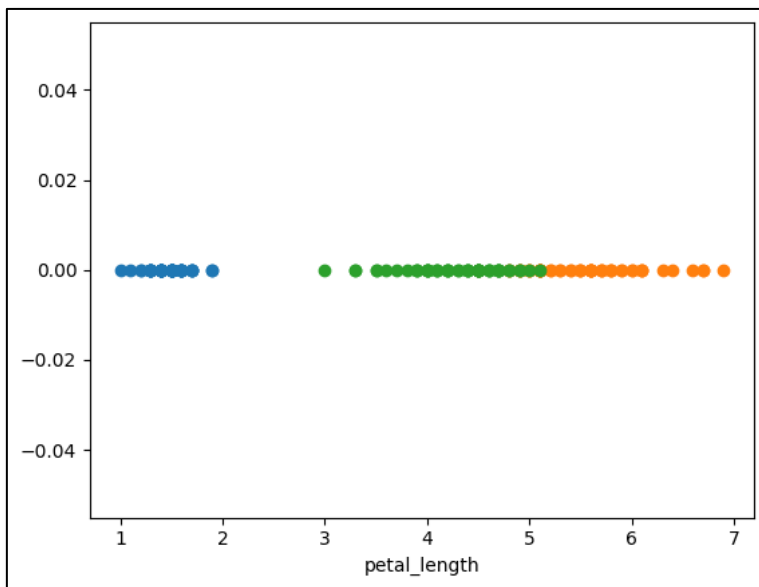
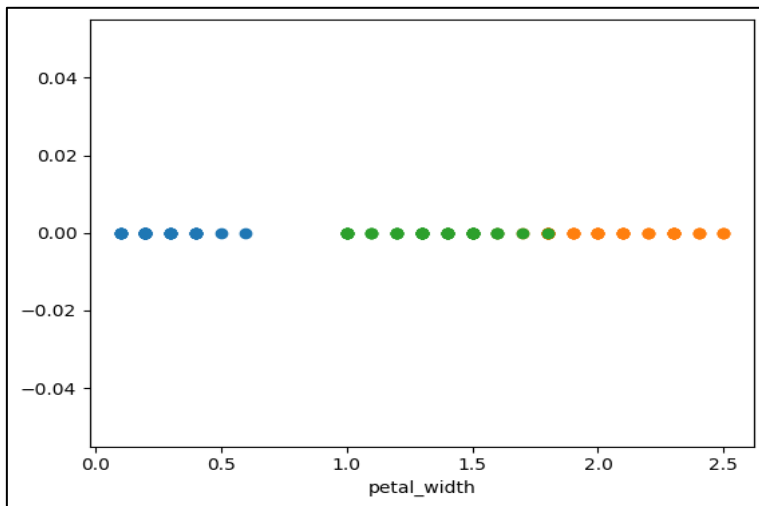
	12	4.8	3.0	1.4	0.1	setosa
	13	4.3	3.0	1.1	0.1	setosa
	14	5.8	4.0	1.2	0.2	setosa
	15	5.7	4.4	1.5	0.4	setosa
	16	5.4	3.9	1.3	0.4	setosa
	17	5.1	3.5	1.4	0.3	setosa
	18	5.7	3.8	1.7	0.3	setosa
	19	5.1	3.8	1.5	0.3	setosa
	20	5.4	3.4	1.7	0.2	setosa
	21	5.1	3.7	1.5	0.4	setosa
	22	4.6	3.6	1.0	0.2	setosa
	23	5.1	3.3	1.7	0.5	setosa
	24	4.8	3.4	1.9	0.2	setosa

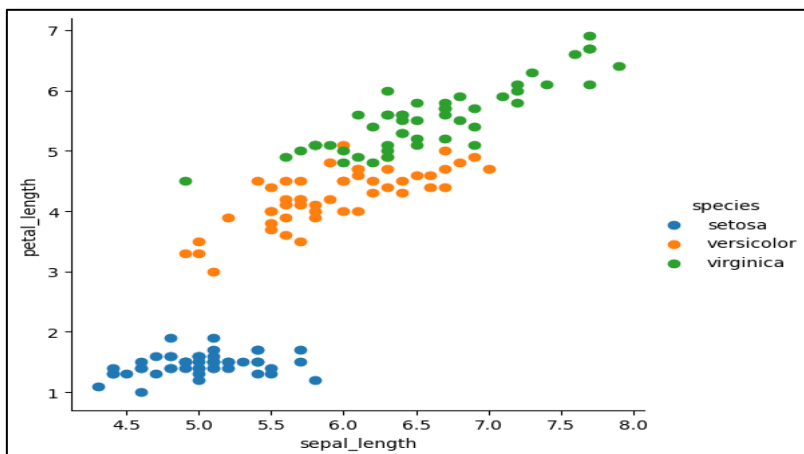
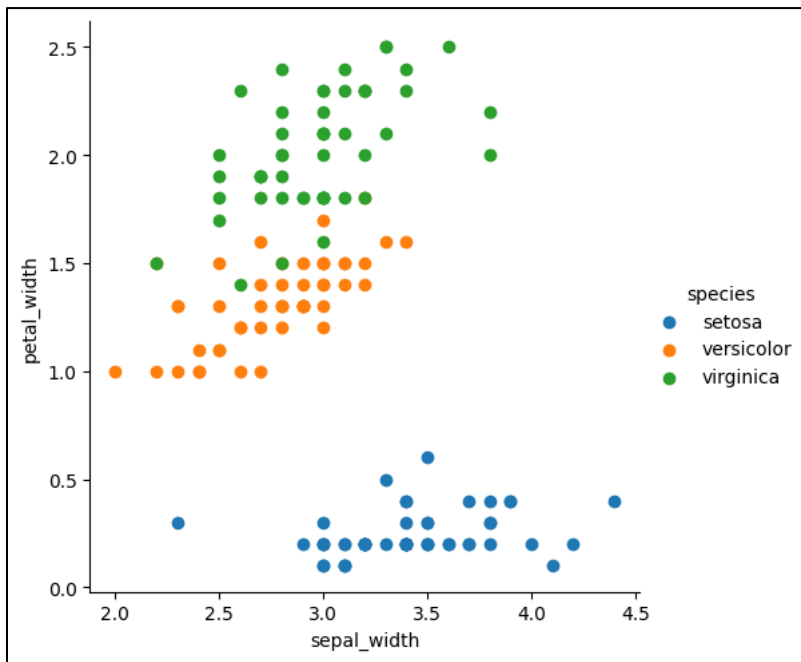
	25	5.0	3.0	1.6	0.2	setosa
	26	5.0	3.4	1.6	0.4	setosa
	27	5.2	3.5	1.5	0.2	setosa
	28	5.2	3.4	1.4	0.2	setosa
	29	4.7	3.2	1.6	0.2	setosa
	30	4.8	3.1	1.6	0.2	setosa
	31	5.4	3.4	1.5	0.4	setosa
	32	5.2	4.1	1.5	0.1	setosa
	33	5.5	4.2	1.4	0.2	setosa
	34	4.9	3.1	1.5	0.1	setosa
	35	5.0	3.2	1.2	0.2	setosa
	36	5.5	3.5	1.3	0.2	setosa
	37	4.9	3.1	1.5	0.1	setosa
	38	4.4	3.0	1.3	0.2	setosa
	39	5.1	3.4	1.5	0.2	setosa
	40	5.0	3.5	1.3	0.3	setosa
	41	4.5	2.3	1.3	0.3	setosa
	42	4.4	3.2	1.3	0.2	setosa
	43	5.0	3.5	1.6	0.6	setosa

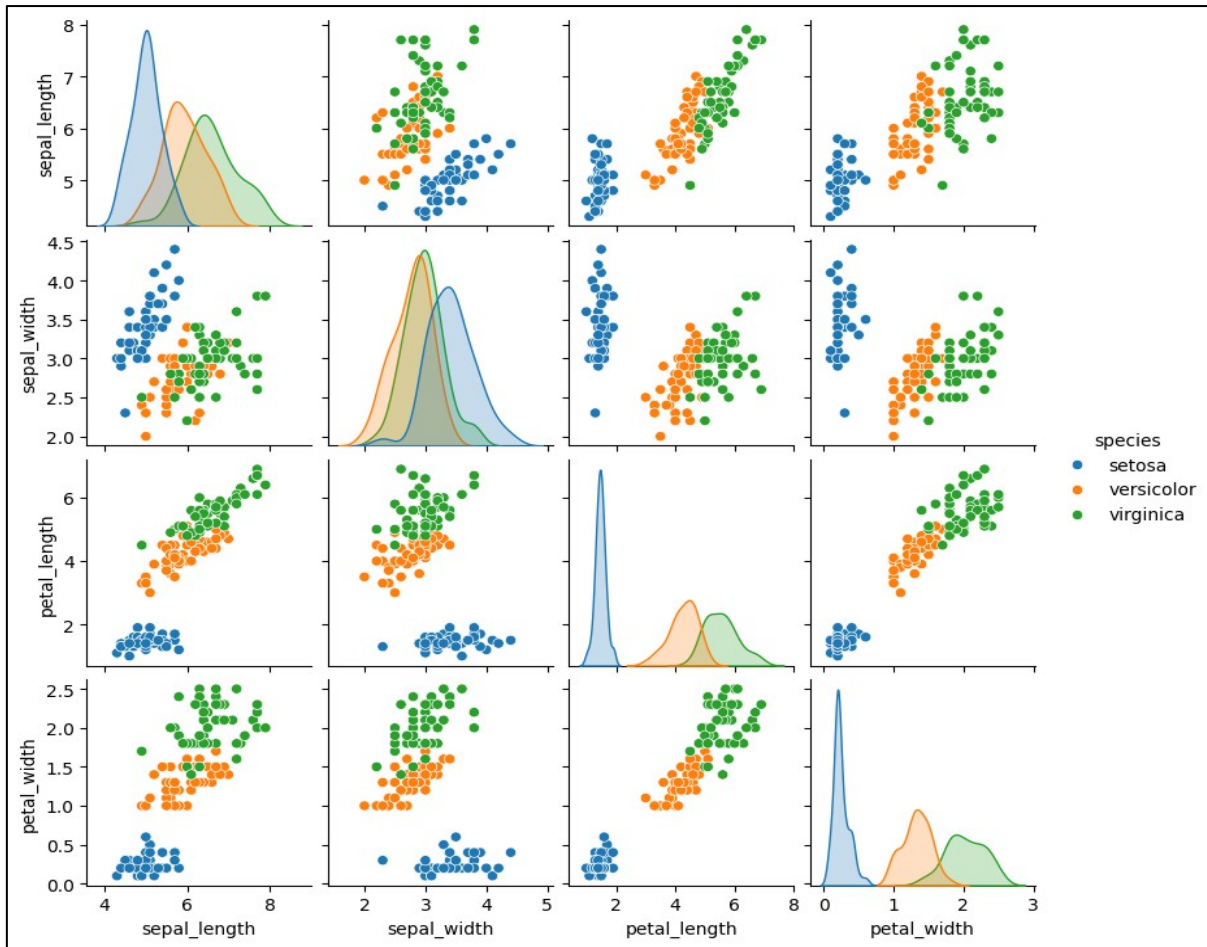


44	5.1	3.8	1.9	0.4	setosa
45	4.8	3.0	1.4	0.3	setosa
46	5.1	3.8	1.6	0.2	setosa
47	4.6	3.2	1.4	0.2	setosa
48	5.3	3.7	1.5	0.2	setosa
49	5.0	3.3	1.4	0.2	setosa









## **RESULT:**

Thus, the python program to implement univariate, bivariate and multivariate has been successfully implemented and the results have been verified and analysed.

**EXPERIMENT NO : 2**

**DATE : 22/08/24**

**REGISTER NO : 231501502**

**NAME : DONEESWARAN J**

---

## **A PYTHON PROGRAM TO IMPLEMENT SIMPLE LINEAR REGRESSION USING LEAST SQUARE METHOD**

### **AIM:**

To implement a python program for constructing a simple linear regression using least square method.

### **ALGORITHM:**

#### **STEP 1: IMPORT NECESSARY LIBRARIES:**

- pandas for data manipulation and matplotlib.pyplot for plotting.

#### **STEP 2: READ THE DATASET:**

- Use the pandas `read\_csv` function to read the dataset (e.g., headbrain.csv).
- Store the dataset in a variable (e.g., `data`).

#### **STEP 3: PREPARE THE DATA:**

- Extract the independent variable (X) and dependent variable (y) from the dataset.
- Reshape X and y to be 2D arrays if needed.

#### **STEP 4: CALCULATE THE MEAN:**

- Calculate the mean of X and y.

### STEP 5: CALCULATE THE COEFFICIENTS:

- Calculate the slope (m) using the formula:

$$m = \frac{\sum_{i=1}^n (X_i - \bar{X})(y_i - \bar{y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

- Calculate the intercept (b) using the formula:  $b = \bar{y} - m\bar{X}$

### STEP 6: MAKE PREDICTIONS:

- Use the calculated slope and intercept to make predictions for each X value:

$$\hat{y} = mx + b$$

### STEP 7: PLOT THE REGRESSION LINE:

- Plot the original data points (X, y) as a scatter plot.
- Plot the regression line (X, predicted\_y) as a line plot.

### STEP 8: CALCULATE THE R-SQUARED VALUE:

- Calculate the total sum of squares (TSS) using the formula:  
 $TSS = \sum_{i=1}^n (y_i - \bar{y})^2$
- Calculate the residual sum of squares (RSS) using the formula:  
 $RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2$
- Calculate the R-squared value using the formula:  $R^2 = 1 - \frac{RSS}{TSS}$

### STEP 9: DISPLAY THE RESULTS:

- Print the slope, intercept, and R-squared value.

### STEP 10: COMPLETE THE PROGRAM:

- Combine all the steps into a Python program.
- Run the program to perform simple linear regression on the dataset.

## CODE

```
from google.colab import drive
drive.mount('/content/drive')

import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
data = pd.read_csv('/content/drive/MyDrive/Datasets/headbrain.csv')

x, y = np.array(list(data['Head Size(cm^3)'])), np.array(list(data['Brain
Weight(grams)']))
print(x[:5], y[:5])

def get_line(x, y):
    x_m, y_m = np.mean(x), np.mean(y)
    print(x_m, y_m)
    x_d, y_d = x-x_m, y-y_m
    m = np.sum(x_d*y_d)/np.sum(x_d**2)
    c = y_m - (m*x_m)
    print(m, c)
    return lambda x : m*x+c
lin = get_line(x, y)

X = np.linspace(np.min(x)-100, np.max(x)+100, 1000)
Y = np.array([lin(x) for x in X])
plt.plot(X, Y, color='red', label='Regression line')
plt.scatter(x, y, color='green', label='Scatter plot')
plt.xlabel('Head Size(cm^3)')
```

```

plt.ylabel('Brain Weight(grams)')
plt.legend()
plt.show()

X = np.linspace(np.min(x)-100, np.max(x)+100, 1000)

Y = np.array([lin(x) for x in X])
plt.plot(X, Y, color='red', label='Regression line')
plt.scatter(x, y, color='green', label='Scatter plot')
plt.xlabel('Head Size(cm^3)')
plt.ylabel('Brain Weight(grams)')
plt.legend()
plt.show()

def get_error(line_fuc, x, y):
    y_m = np.mean(y)
    y_pred = np.array([line_fuc(_) for _ in x])
    ss_t = np.sum((y-y_m)**2)
    ss_r = np.sum((y-y_pred)**2)
    return 1-(ss_r/ss_t)

get_error(lin, x, y)

from sklearn.linear_model import LinearRegression
x = x.reshape((len(x),1))
reg=LinearRegression()
reg=reg.fit(x, y)
print(reg.score(x, y))

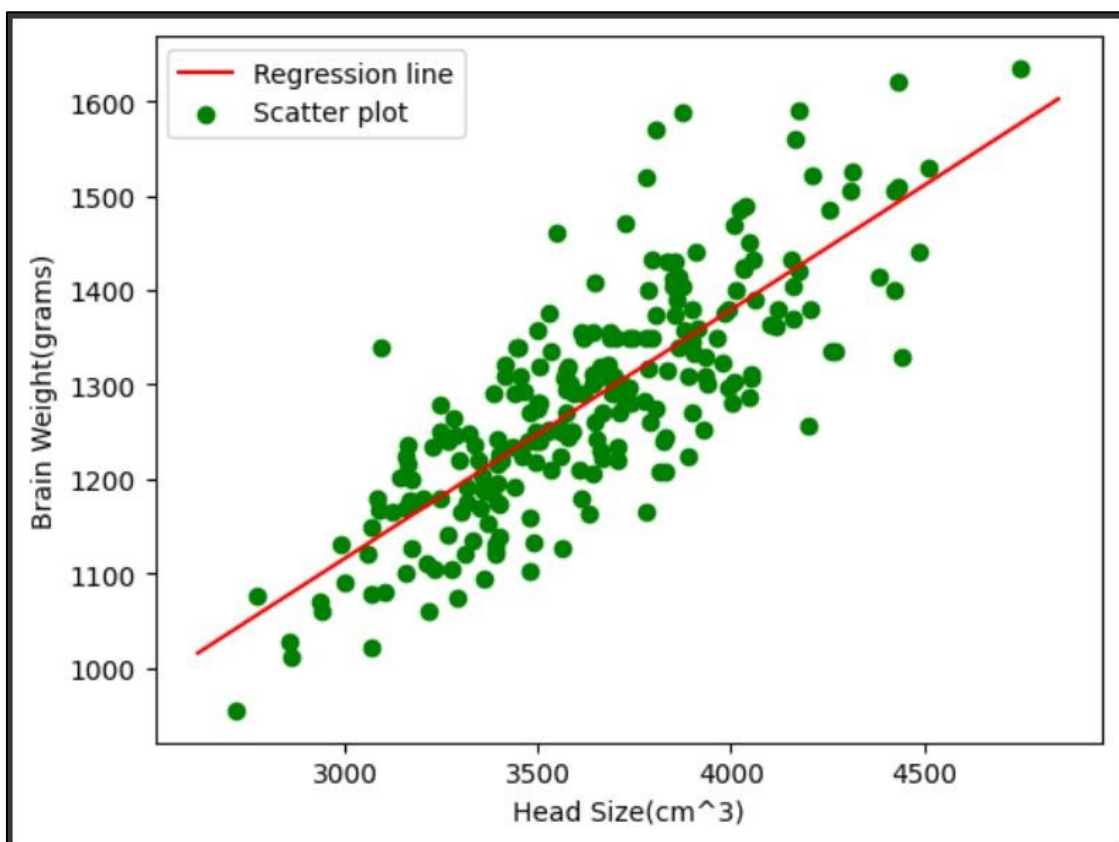
```

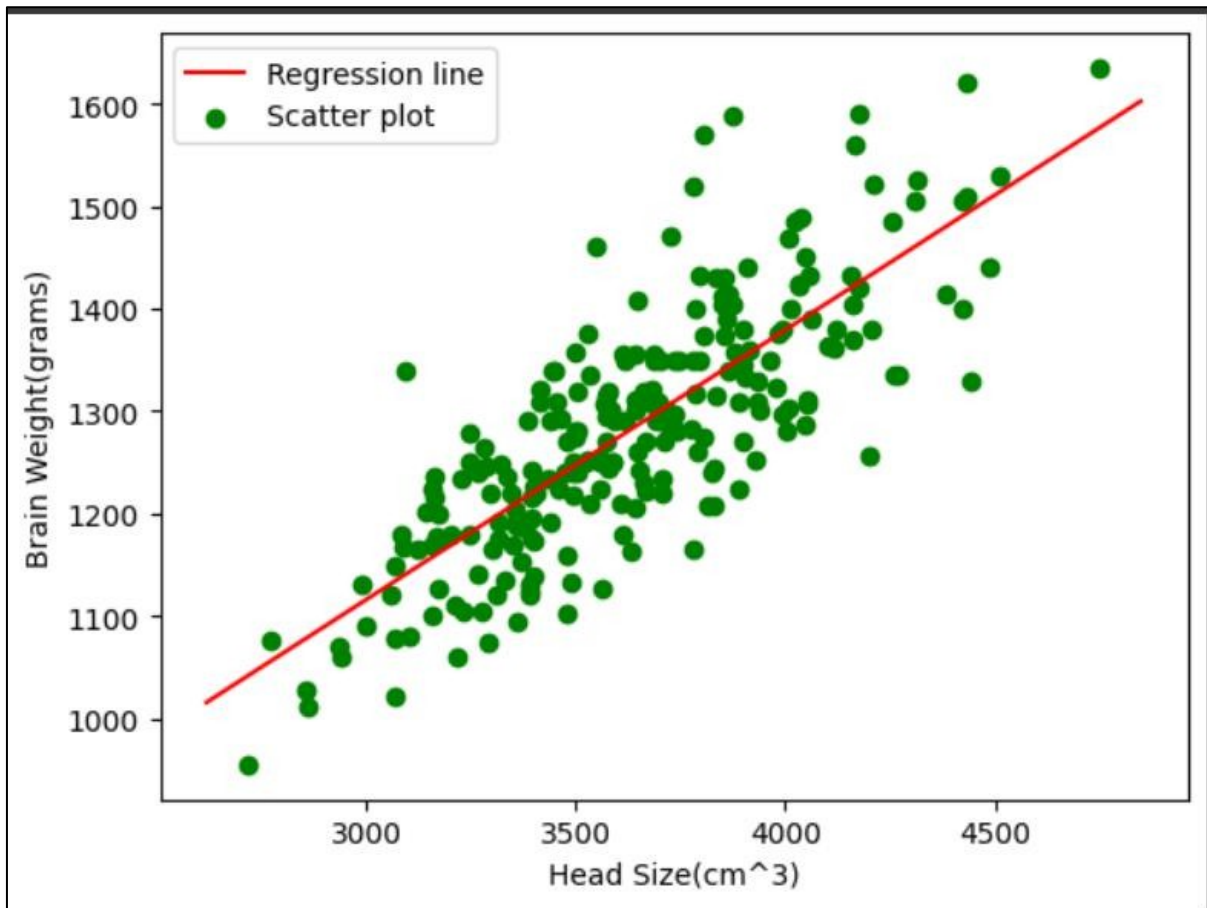


## OUTPUT

```
[4512 3738 4261 3777 4177] [1530 1297 1335 1282 1590]
```

```
3633.9915611814345 1282.873417721519  
0.2634293394893993 325.5734210494428
```





0.639311719957

0.639311719957

## **RESULT**

Thus, the python program to Simple Linear Regression using Least Square Method has been successfully implemented and the results have been verified and analysed.

**EXPERIMENT NO : 3**

**DATE : 29/08/24**

**REGISTER NO : 231501502**

**NAME : DONEESWARAN J**

---

## **A PYTHON PROGRAM TO IMPLEMENT LOGISTIC MODEL**

### **AIM:**

To implement python program for the logistic model using suv car dataset.

### **ALGORITHM:**

#### **STEP 1: IMPORT NECESSARY LIBRARIES:**

- pandas for data manipulation
- sklearn.model\_selection for train-test split
- sklearn.preprocessing for data preprocessing
- sklearn.linear\_model for logistic regression
- matplotlib.pyplot for plotting

#### **STEP 2: READ THE DATASET:**

- Use pandas to read the suv\_cars.csv dataset into a DataFrame.

#### **STEP 3: PREPROCESS THE DATA:**

- Select the relevant columns for the analysis (e.g., 'Age', 'EstimatedSalary', 'Purchased').
- Encode categorical variables if necessary (e.g., using LabelEncoder or OneHotEncoder).
- Split the data into features (X) and target variable (y).

#### **STEP 4: SPLIT THE DATA:**

- Split the dataset into training and testing sets using `train_test_split`.

#### **STEP 5: FEATURE SCALING:**

- Standardize the features using `StandardScaler` to ensure they have the same scale.

#### **STEP 6: CREATE AND TRAIN THE MODEL:**

- Create a logistic regression model using `LogisticRegression` from `sklearn.linear_model`.
- Train the model on the training data using the `fit` method.
  - Create a function named “Sigmoid ()” which will define the sigmoid values using the
  - formula  $(1/1+e^{-z})$  and return the computed value.
  - Create a function named “initialize()” which will initialize the values with zeroes and assign the value to “weights” variable, initializes with ones and assigns the value to variable “x” and returns both “x” and “weights”.
  - Create a function named “fit” which will be used to plot the graph according to the training data.
  - Create a predict function that will predict values according to the training model created using the fit function.
  - Invoke the `standardize()` function for “x-train” and “x-test”

#### **STEP 7: MAKE PREDICTIONS:**

- Use the trained model to make predictions on the test data using the predict method.
  - Use the “predict()” function to predict the values of the testing data and assign the value to “y\_pred” variable.
  - Use the “predict()” function to predict the values of the training data and assign the value to “y\_trainn” variable.
  - Compute f1\_score for both the training and testing data and assign the values to “f1\_score\_tr” and “f1\_score\_te” respectively

### **STEP 8: EVALUATE THE MODEL:**

- Calculate the accuracy of the model on the test data using the score method.

(Accuracy = (tp+tn)/(tp+tn+fp+fn)).

- Generate a confusion matrix and classification report to further evaluate the model's performance.

### **STEP 9: VISUALIZE THE RESULTS:**

- Plot the decision boundary of the logistic regression model (optional).

### **CODE**

```
import pandas as pd
import numpy as np
from numpy import log,dot,exp,shape
from sklearn.metrics import confusion_matrix
data = pd.read_csv('/content/drive/MyDrive/suv_data.csv')
```

```
print(data.head())
```

```
x = data.iloc[:, [2, 3]].values
```

```
y = data.iloc[:, 4].values
```

```
from sklearn.model_selection import train_test_split
```

```
x_train, x_test, y_train, y_test=train_test_split(x,y,test_size=0.10,  
random_state=0)
```

```
from sklearn.preprocessing import StandardScaler
```

```
sc=StandardScaler()
```

```
x_train=sc.fit_transform(x_train)
```

```
x_test=sc.transform(x_test)
```

```
print (x_train[0:10,:])
```

```
from sklearn.linear_model import LogisticRegression
```

```
classifier=LogisticRegression(random_state=0)
```

```
classifier.fit(x_train,y_train)
```

```
LogisticRegression (random_state=0)
```

```
y_pred = classifier.predict(x_test)
```

```
print(y_pred)
```

```
from sklearn.metrics import confusion_matrix
```

```
cm = confusion_matrix(y_test, y_pred)
```

```
print ("Confusion Matrix : \n", cm)
```

```

from sklearn.metrics import accuracy_score

print ("Accuracy : ", accuracy_score(y_test, y_pred))

from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test=train_test_split(x,y,test_size=0.10,
random_state=0)

def Std(input_data):
    mean0 = np.mean(input_data[:, 0])
    sd0 = np.std(input_data[:, 0])
    mean1 = np.mean(input_data[:, 1])
    sd1 = np.std(input_data[:, 1])
    return lambda x:((x[0]-mean0)/sd0, (x[1]-mean1)/sd1)

my_std = Std(x)
my_std(x_train[0])

def standardize(X_tr):
    for i in range(shape(X_tr)[1]):
        X_tr[:,i] = (X_tr[:,i] - np.mean(X_tr[:,i]))/np.std(X_tr[:,i])

def F1_score(y,y_hat):
    tp,tn,fp,fn = 0,0,0,0
    for i in range(len(y)):
        if y[i] == 1 and y_hat[i] == 1:
            tp += 1
        elif y[i] == 1 and y_hat[i] == 0:
            fn += 1

```



```

        elif y[i] == 0 and y_hat[i] == 1:
            fp += 1
        elif y[i] == 0 and y_hat[i] == 0:
            tn += 1
    precision = tp/(tp+fp)
    recall = tp/(tp+fn)
    f1_score = 2*precision*recall/(precision+recall)
    return f1_score

class LogisticRegression:
    def sigmoid(self, z):
        sig = 1 / (1 + exp(-z))
        return sig

    def initialize(self, X):
        weights = np.zeros((shape(X)[1] + 1, 1))
        X = np.c_[np.ones((shape(X)[0], 1)), X]
        return weights, X

    def fit(self, X, y, alpha=0.001, iter=400):
        weights, X = self.initialize(X)

    def cost(theta):
        z = dot(X, theta)

```

```

cost0 = y.T.dot(log(self.sigmoid(z)))
cost1 = (1 - y).T.dot(log(1 - self.sigmoid(z)))
cost = -((cost1 + cost0)) / len(y)

return cost

cost_list = np.zeros(iter,)
for i in range(iter):
    weights = weights - alpha * dot(X.T, self.sigmoid(dot(X,
weights)) - np.reshape(y, (len(y), 1)))
    cost_list[i] = cost(weights).item()
self.weights = weights
return cost_list

def predict(self, X):
    z = dot(self.initialize(X)[1], self.weights)
    lis = []
    for i in self.sigmoid(z):
        if i > 0.5:
            lis.append(1)
        else:
            lis.append(0)
    return lis

standardize(x_train)

```

```

standardize(x_test)
obj1 = LogisticRegression()
model = obj1.fit(x_train, y_train)
y_pred = obj1.predict(x_test)
y_trainn = obj1.predict(x_train)
f1_score_tr = F1_score(y_train, y_trainn)
f1_score_te = F1_score(y_test, y_pred)
print(f1_score_tr)
print(f1_score_te)
conf_mat = confusion_matrix(y_test, y_pred)
accuracy = (conf_mat[0, 0] + conf_mat[1, 1]) / sum(sum(conf_mat))
print("Accuracy is : ", accuracy)

```

## OUTPUT

	User ID	Gender	Age	EstimatedSalary	Purchased
0	15624510	Male	19	19000	0
1	15810944	Male	35	20000	0
2	15668575	Female	26	43000	0
3	15603246	Female	27	57000	0
4	15804002	Male	19	76000	0

```

[[-1.05714987  0.53420426]
 [ 0.2798728  -0.51764734]
 [-1.05714987  0.41733186]
 [-0.29313691 -1.45262654]
 [ 0.47087604  1.23543867]
 [-1.05714987 -0.34233874]
 [-0.10213368  0.30045946]
 [ 1.33039061  0.59264046]
 [-1.15265148 -1.16044554]
 [ 1.04388575  0.47576806]]
[0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 1 0 0 1 0 1 0 1 0 0 0 0 0 0 1 0 0 0 0
 0 0 1]

```

Confusion Matrix :

```
[[31  1]
```

```
[ 1  7]]
```

Accuracy : 0.95

(-1.017692393473028, 0.5361288690822568)

0.7583333333333334

0.823529411764706

Accuracy is : 0.925

## **RESULT**

Thus, the python program to implement logistic model has been successfully implemented and the results have been verified and analyzed.

EXPERIMENT NO : 4

DATE : 05/10/24

REGISTER NO : 231501502

NAME : DONEESWARAN J

---

## **A PYTHON PROGRAM TO IMPLEMENT SINGLE LAYER PERCEPTRON**

### **AIM:**

To implement python program for the single layer perceptron.

### **ALGORITHM:**

#### **STEP 1: IMPORT NECESSARY LIBRARIES:**

- Import numpy for numerical operations.

#### **STEP 2: INITIALIZE THE PERCEPTRON:**

- Define the number of input features (input\_dim).
- Initialize weights (W) and bias (b) to zero or small random values.

#### **STEP 3: DEFINE ACTIVATION FUNCTION:**

- Choose an activation function (e.g., step function, sigmoid, or ReLU).
- User Defined function - sigmoid\_func(x):
  - Compute  $1/(1+\text{np.exp}(-x))$  and return the value.
- User Defined function - der(x):
  - Compute the product of value of sigmoid\_func(x) and  $(1 - \text{sigmoid\_func}(x))$  and return the value.

#### **STEP 4; DEFINE TRAINING DATA:**

- Define input features (X) and corresponding target labels (y).

### **STEP 5: DEFINE LEARNING RATE AND NUMBER OF EPOCHS:**

- Choose a learning rate ( $\alpha$ ) and the number of training epochs.

### **STEP 6: TRAINING THE PERCEPTRON:**

- For each epoch:
  - For each input sample in the training data:
  - Compute the weighted sum of inputs ( $z$ ) as the dot product of input features and weights plus bias ( $z = \text{np.dot}(X[i], W) + b$ ).
  - Apply the activation function to get the predicted output ( $y_{\text{pred}}$ ).
  - Compute the error ( $\text{error} = y[i] - y_{\text{pred}}$ ).
  - Update the weights and bias using the learning rate and error ( $W += \alpha * \text{error} * X[i]$ ;  $b += \alpha * \text{error}$ ).

### **STEP 7: PREDICTION:**

- Use the trained perceptron to predict the output for new input DATA.

### **STEP 8: EVALUATE THE MODEL:**

- Measure the performance of the model using metrics such as accuracy, precision, recall, etc.

## CODE

```
import numpy as np
from sklearn.metrics import accuracy_score, precision_score,
recall_score, f1_score

input_dim=2
W=np.zeros(input_dim)
b=0.0

def sigmoid_func(x):
    return 1 / (1 + np.exp(-x))
def der(x):
    sigmoid = sigmoid_func(x)
    return sigmoid * (1 - sigmoid)

np.random.seed(42)
x = np.array([[150,8],
               [130,7],
               [180,6],
               [170,5]])
y = np.array([0,0,1,1])

alpha = 0.1
epochs = 10000

for epoch in range(epochs):
    for i in range(len(x)):
        z = np.dot(x[i], W) + b
        y_pred = sigmoid_func(z)
        error = y[i] - y_pred
        W += alpha * error * x[i]
```

```
b += alpha * error

def predict(X):
    z = np.dot(X, W) + b
    return (sigmoid_func(z) > 0.5).astype(int)
y_pred = predict(x)
accuracy = accuracy_score(y, y_pred)
precision = precision_score(y, y_pred)
recall = recall_score(y, y_pred)

F1_score = f1_score(y, y_pred)

print("Prediction:", y_pred)
print("Accuracy:", accuracy)
print("Precision:", precision)
print("Recall:", recall)
print("F1 Score:", F1_score)
```

## **OUTPUT**

**Prediction: [0 0 1 1]**  
**Accuracy: 1.0**  
**Precision: 1.0**  
**Recall: 1.0**  
**F1 Score: 1.0**

---

## **RESULT**

Thus, the python program to implement single layer perceptron has been successfully implemented and the results have been verified and analysed.



**EXPERIMENT NO : 5**

**DATE : 19/09/24**

**REGISTER NO : 231501502**

**NAME : DONEESWARAN J**

---

## **A PYTHON PROGRAM TO IMPLEMENT MULTI LAYER PERCEPTRON WITH BACK PROPOGATION**

### **AIM:**

To implement multilayer perceptron with back propagation using python.

### **ALGORITHM:**

#### **STEP 1: IMPORT THE NECESSARY LIBRARIES**

- Import pandas as pd.
- Import numpy as np.

#### **STEP 2: READ AND DISPLAY THE DATASET**

- Use ``pd.read_csv("banknotes.csv")`` to read the dataset.
- Assign the result to a variable (e.g., ``data``).
- Display the first ten rows using ``data.head(10)``.

#### **STEP 3: DISPLAY DATASET DIMENSIONS**

- Use the ``.shape`` attribute on the dataset (e.g., ``data.shape``).

#### **STEP 4: DISPLAY DESCRIPTIVE STATISTICS**

- Use the ``.describe()`` function on the dataset (e.g., ``data.describe()``).

#### **STEP 5: IMPORT TRAIN-TEST SPLIT MODULE**

- Import ``train_test_split`` from ``sklearn.model_selection``.

## **STEP 6: SPLIT DATASET WITH 80-20 RATIO**

- Assign the features to a variable (e.g., `X = data.drop(columns='target')`).
- Assign the target variable to another variable (e.g., `y = data['target']`).
- Use `train_test_split` to split the dataset into training and testing sets with a ratio of 0.2.
- Assign the results to `x_train`, `x_test`, `y_train`, and `y_test`.

## **STEP 7: IMPORT MLPCLASSIFIER MODULE**

- Import `MLPClassifier` from `sklearn.neural_network`.

## **STEP 8: INITIALIZE MLPCLASSIFIER**

- Create an instance of `MLPClassifier` with `max_iter=500` and `activation='relu'`.
- Assign the instance to a variable (e.g., `clf`).

## **STEP 9: FIT THE CLASSIFIER**

- Fit the model using `clf.fit(x_train, y_train)`.

## **STEP 10: MAKE PREDICTIONS**

- Use the `.predict()` function on `x_test` (e.g., `pred = clf.predict(x_test)`).
- Display the predictions.

## **STEP 11: IMPORT METRICS MODULES**

- Import `confusion_matrix` from `sklearn.metrics`.
- Import `classification_report` from `sklearn.metrics`.

## **STEP 12: DISPLAY CONFUSION MATRIX**

- Use `confusion_matrix(y_test, pred)` to generate the confusion matrix.
- Display the confusion matrix.

### **STEP 13: DISPLAY CLASSIFICATION REPORT**

- Use ``classification_report(y_test, pred)`` to generate the classification report.
- Display the classification report.

### **STEP 14: REPEAT STEPS 9-13 WITH DIFFERENT ACTIVATION FUNCTIONS**

- Initialize ``MLPClassifier`` with ``activation='logistic'``.
- Fit the model and make predictions.
- Display the confusion matrix and classification report.
- Repeat for ``activation='tanh'``.
- Repeat for ``activation='identity'``.

### **STEP 15: REPEAT STEPS 7-14 WITH 70-30 RATIO**

- Use ``train_test_split`` to split the dataset into training and testing sets with a ratio of 0.3.
- Assign the results to ``x_train``, ``x_test``, ``y_train``, and ``y_test``.
- Repeat Steps 7-14 with the new training and testing sets.

## **CODE**

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.neural_network import MLPClassifier
from sklearn.metrics import classification_report, confusion_matrix

bnotes = pd.read_csv('../content/drive/MyDrive/bank_note_data.csv')
print(bnotes.head(10))

x = bnotes.drop('Class', axis=1)
y = bnotes['Class']
print(x.head(2))
```

```

print(y.head(2))

def train_and_evaluate(activation, x_train, y_train, x_test, y_test):
    mlp = MLPClassifier(max_iter=500, activation=activation)
    mlp.fit(x_train, y_train)

    pred = mlp.predict(x_test)
    print(f'Predictions using activation function '{activation}':\n{pred}\n")

    cm = confusion_matrix(y_test, pred)
    print(f'Confusion Matrix for '{activation}':\n{cm}\n")

    report = classification_report(y_test, pred)
    print(f'Classification Report for '{activation}':\n{report}\n")

x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2)

for activation in ['relu', 'logistic', 'tanh', 'identity']:
    train_and_evaluate(activation, x_train, y_train, x_test, y_test)

x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.3)
for activation in ['relu', 'logistic', 'tanh', 'identity']:
    train_and_evaluate(activation, x_train, y_train, x_test, y_test)

```

## OUTPUT

```
➡ Image.Var Image.Skew Image.Curt Entropy Class
0 3.62160 8.6661 -2.80730 -0.44699 0
1 4.54590 8.1674 -2.45860 -1.46210 0
2 3.86600 -2.6383 1.92420 0.10645 0
3 3.45660 9.5228 -4.01120 -3.59440 0
4 0.32924 -4.4552 4.57180 -0.98880 0
5 4.36840 9.6718 -3.96060 -3.16250 0
6 3.59120 3.0129 0.72888 0.56421 0
7 2.09220 -6.8100 8.46360 -0.60216 0
8 3.20320 5.7588 -0.75345 -0.61251 0
9 1.53560 9.1772 -2.27180 -0.73535 0
Image.Var Image.Skew Image.Curt Entropy
0 3.6216 8.6661 -2.8073 -0.44699
1 4.5459 8.1674 -2.4586 -1.46210
0 0
1 0
Name: Class, dtype: int64
Predictions using activation function 'relu':
[1 1 1 1 0 0 1 1 1 0 0 0 1 1 0 0 0 1 1 1 1 0 0 0 1 0 0 0 1 0 0 1 0 0 0 0
 0 0 0 1 0 0 0 1 1 1 0 0 1 1 0 0 0 1 0 1 1 1 1 0 0 0 0 1 1 0 0 1 0 1 0 1 1
 1 1 0 1 1 1 0 0 0 0 0 1 1 0 0 0 0 0 1 1 0 1 1 0 0 0 1 1 1 1 1 0 1 0
 1 1 1 0 1 1 1 0 0 0 1 1 1 0 1 0 1 0 1 1 0 0 1 0 1 0 0 0 0 0 1 0 0 0
 1 0 1 1 1 0 1 1 1 0 1 1 0 0 1 0 1 1 0 0 0 0 1 0 1 1 1 1 1 1 1 0 1 0 0 1 0
 0 1 1 0 0 1 0 0 1 1 0 1 0 0 0 1 1 1 1 0 1 1 1 0 0 1 1 0 1 0 0 0 1 0 0 1 1
 0 0 0 1 1 1 1 1 0 0 1 0 1 0 0 0 0 1 0 0 1 0 1 1 1 0 1 1 0 0 0 0 1 0 0 0
 0 0 1 1 0 0 0 0 0 1 1 1 1 1 1 1 0]
```

Confusion Matrix for 'relu':

```
[[143  0]
 [  0 132]]
```

```
▶ Classification Report for 'relu':
➡ precision recall f1-score support
0 1.00 1.00 1.00 143
1 1.00 1.00 1.00 132

accuracy 1.00 275
macro avg 1.00 1.00 1.00 275
weighted avg 1.00 1.00 1.00 275
```

Predictions using activation function 'logistic':

```
[1 1 1 1 0 0 1 1 1 0 0 0 1 1 0 0 0 1 1 1 1 0 0 1 0 0 0 0 0 1 0 0 1 0 0 0 0
 0 0 0 1 0 0 0 1 1 1 0 0 1 1 0 0 0 1 0 1 1 1 1 0 0 0 0 1 1 0 0 1 0 1 0 1 1
 1 1 0 1 1 1 0 0 0 0 1 1 0 0 0 0 0 1 1 0 1 1 0 0 0 1 0 0 1 1 1 1 0 1 0
 1 1 1 0 1 1 1 0 0 0 1 1 1 0 1 1 0 1 0 1 1 0 0 1 0 1 1 0 0 0 0 0 0 1 0 0 0
 1 0 1 1 1 0 1 1 1 0 1 1 0 0 1 0 1 1 0 0 0 0 1 0 1 1 1 1 1 1 1 0 1 0 0 1 0
 0 1 1 0 0 1 0 0 1 1 0 1 0 0 0 1 1 1 1 0 1 1 1 0 0 1 1 0 1 0 0 0 1 0 0 1 1
 0 0 0 1 1 1 1 1 0 0 1 0 1 0 0 0 0 1 0 0 1 0 1 1 1 0 1 1 0 0 0 0 1 0 0 0
 0 0 1 1 0 0 0 0 0 1 1 1 1 1 1 1 0]
```

Confusion Matrix for 'logistic':

```
[[143  0]
 [  0 132]]
```

Classification Report for 'logistic':

```
precision recall f1-score support
0 1.00 1.00 1.00 143
1 1.00 1.00 1.00 132

accuracy 1.00 275
macro avg 1.00 1.00 1.00 275
weighted avg 1.00 1.00 1.00 275
```



Predictions using activation function 'tanh':

```
[1 1 1 1 0 0 1 1 1 0 0 0 1 1 0 0 0 1 1 1 0 0 1 0 0 0 0 0 1 0 0 1 0 0 0 0
0 0 0 1 0 0 0 1 1 1 0 0 1 1 0 0 0 1 0 1 1 1 1 0 0 0 0 1 1 0 0 1 0 1 0 1 1
1 1 0 1 1 1 0 0 0 0 0 1 1 0 0 0 0 0 0 1 1 0 1 1 0 0 0 1 0 0 1 1 1 1 0 1 0
1 1 1 0 1 1 1 0 0 0 1 1 1 0 1 1 0 1 0 1 1 0 0 1 0 1 1 0 0 0 0 0 0 1 0 0 0
1 0 1 1 1 0 1 1 1 0 1 1 0 0 1 0 1 1 0 0 0 0 1 0 1 1 1 1 1 1 1 0 1 0 0 1 0
0 1 1 0 0 1 0 0 1 1 0 1 0 0 0 1 1 1 1 0 1 1 1 0 0 1 1 0 1 0 0 0 1 0 0 1 1
0 0 0 1 1 1 1 1 0 0 1 0 1 0 0 0 0 1 0 0 1 0 1 1 1 0 1 1 0 0 0 0 1 0 0 0
0 0 1 1 0 0 0 0 0 1 1 1 1 1 1 0]
```

Confusion Matrix for 'tanh':

```
[[143  0]
 [  0 132]]
```

Classification Report for 'tanh':

	precision	recall	f1-score	support
0	1.00	1.00	1.00	143
1	1.00	1.00	1.00	132
accuracy			1.00	275
macro avg	1.00	1.00	1.00	275
weighted avg	1.00	1.00	1.00	275

Predictions using activation function 'identity':

```
[1 1 1 1 0 0 1 1 1 0 0 0 1 1 1 0 0 0 1 1 1 0 0 1 0 0 0 0 0 1 0 0 1 0 0 0 1
0 0 0 1 0 0 0 1 1 1 0 0 1 1 0 0 0 1 0 1 1 1 1 0 0 0 0 1 1 0 0 1 0 1 0 1 1
1 1 0 1 1 1 0 0 0 0 0 1 1 0 0 0 0 0 0 1 1 0 1 1 0 0 0 1 0 0 1 1 1 1 0 1 0
1 1 1 0 1 1 1 0 0 0 1 1 1 0 1 1 0 1 0 1 1 0 0 1 0 1 1 0 0 0 0 0 0 1 0 1 0
1 0 1 1 1 0 1 1 1 0 1 1 0 0 1 0 1 1 0 0 0 0 1 0 1 1 1 1 1 1 1 1 0 1 0 0 1 0
0 1 1 0 0 1 0 0 1 1 0 1 0 0 0 1 1 1 1 0 1 1 1 0 0 1 1 0 1 0 0 0 1 0 0 1 1
0 0 0 1 1 1 1 1 0 0 1 0 1 0 0 0 0 1 0 0 1 0 1 1 1 0 1 1 0 0 0 0 1 0 0 0
0 0 1 1 0 0 0 0 0 1 1 1 1 1 1 0]
```



Confusion Matrix for 'identity':

```
[[141  2]
 [  0 132]]
```

Classification Report for 'identity':

	precision	recall	f1-score	support
0	1.00	0.99	0.99	143
1	0.99	1.00	0.99	132
accuracy			0.99	275
macro avg	0.99	0.99	0.99	275
weighted avg	0.99	0.99	0.99	275

Predictions using activation function 'relu':

```
[0 0 0 0 0 1 0 0 1 0 1 1 1 1 1 0 0 1 0 0 0 1 1 0 1 0 0 1 1 0 0 1 0 1 0 0 0
0 1 0 0 1 1 0 1 1 1 0 1 1 0 0 1 1 0 1 1 1 1 1 0 1 0 1 1 0 0 0 1 1 0 1 0
1 0 0 0 0 0 0 0 0 1 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 1 0 1 0 0 0 1 0 1 0 1 0
0 0 0 0 0 1 0 0 0 0 1 1 1 1 0 1 1 1 0 1 1 1 1 0 1 0 0 0 1 0 0 0 1 0 0 0 0
0 0 0 1 1 1 1 0 0 0 0 0 1 0 1 0 0 1 0 1 1 0 1 1 0 1 1 0 0 1 0 0 1 1 1 1 1
0 0 1 1 0 0 1 0 0 0 0 0 1 0 1 1 0 0 0 1 0 1 1 0 1 0 0 0 0 1 1 0 0 0 0 0 0
1 1 1 0 0 0 1 0 0 0 0 1 1 0 0 1 1 1 0 0 1 1 1 1 0 0 0 1 0 1 0 0 1 1 1 1
0 0 1 0 1 0 0 1 0 1 0 0 0 0 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 0 0 0 1 0 0 0
0 0 0 1 1 1 0 0 1 0 1 0 1 0 0 1 0 0 0 0 1 0 1 1 0 1 0 0 1 1 1 1 0 0 0 1 1
0 0 0 0 1 0 1 0 0 0 0 0 1 1 1 1 0 0 0 1 0 1 0 1 0 1 0 1 1 1 0 0 0 1 0 0 0
0 1 0 1 0 1 0 0 1 0 0 0 0 0 0 1 1 1 1 0 0 0 1 1 1 0 0 0 0 0 1 0 0 0 0
0 0 1 0 1]
```

Confusion Matrix for 'relu':

```
[[239  0]
 [  0 173]]
```



## **RESULT**

Thus, the python program to implement multi-layer perceptron has been successfully implemented and the results have been verified and analysed.



**EXPERIMENT NO : 6**

**DATE : 19/09/24**

**REGISTER NO : 231501502**

**NAME : DONEESWARAN J**

---

## **A PYTHON PROGRAM TO IMPLEMENT SVM CLASSIFIER MODEL**

### **AIM:**

To implement a SVM classifier model using python and determine its accuracy.

### **ALGORITHM:**

#### **STEP 1: IMPORT NECESSARY LIBRARIES**

- Import numpy as np.
- Import pandas as pd.
- Import SVM from sklearn.
- Import matplotlib.pyplot as plt.
- Import seaborn as sns.
- Set the font\_scale attribute to 1.2 in seaborn.

#### **STEP 2: LOAD AND DISPLAY DATASET**

- Read the dataset (muffins.csv) using `pd.read\_csv()`.
- Display the first five instances using the `head()` function.

#### **STEP 3: PLOT INITIAL DATA**

- Use the `sns.lmplot()` function.
- Set the x and y axes to "Sugar" and "Flour".
- Assign "recipes" to the data parameter.
- Assign "Type" to the hue parameter.
- Set the palette to "Set1".
- Set fit\_reg to False.
- Set scatter\_kws to {"s": 70}.
- Plot the graph.

#### **STEP 4: PREPARE DATA FOR SVM**

- Extract "Sugar" and "Butter" columns from the recipes dataset and assign to variable `sugar\_butter`.
- Create a new variable `type\_label`.
- For each value in the "Type" column, assign 0 if it is "Muffin" and 1 otherwise.

## STEP 5: TRAIN SVM MODEL

- Import the SVC module from the svm library.
- Create an SVC model with kernel type set to linear.
- Fit the model using `sugar\_butter` and `type\_label` as the parameters.

## STEP 6: CALCULATE DECISION BOUNDARY

- Use the `model.coef\_` function to get the coefficients of the linear model.
- Assign the coefficients to a list named `w`.
- Calculate the slope `a` as  $w[0] / w[1]$ .
- Use `np.linspace()` to generate values from 5 to 30 and assign to variable `xx`.
- Calculate the intercept using the first value of the model intercept and divide by `w[1].
- Calculate the decision boundary line `y` as  $a * xx - (model.intercept_[0] / w[1])$ .

## STEP 7: CALCULATE SUPPORT VECTOR BOUNDARIES

- Assign the first support vector to variable `b`.
- Calculate `yy\_down` as  $a * xx + (b[1] - a * b[0])$ .
- Assign the last support vector to variable `b`.
- Calculate `yy\_up` using the same method.

## STEP 8: PLOT DECISION BOUNDARY

- Use the `sns.lmplot()` function again with the same parameters as in Step 3.
- Plot the decision boundary line `xx` and `yy`.

## STEP 9: PLOT SUPPORT VECTOR BOUNDARIES

- Plot the decision boundary with `xx`, `yy\_down`, and `k--`.
- Plot the support vector boundaries with `xx`, `yy\_up`, and `k--`.
- Scatter plot the first and last support vectors.

## STEP 10: IMPORT ADDITIONAL LIBRARIES

- Import `confusion\_matrix` from `sklearn.metrics`.
- Import `classification\_report` from `sklearn.metrics`.
- Import `train\_test\_split` from `sklearn.model\_selection`.

## STEP 11: SPLIT DATASET

- Assign `x\_train`, `x\_test`, `y\_train`, and `y\_test` using `train\_test\_split`.
- Set the test size to 0.2.

## STEP 12: TRAIN NEW MODEL

- Create a new SVC model named `model1`.
- Fit the model using the training data (`x\_train` and `y\_train`).

## STEP 13: MAKE PREDICTIONS

- Use the `predict()` function on `model1` with `x\_test` as the parameter.
- Assign the predictions to variable `pred`.

## STEP 14: EVALUATE MODEL

- Display the confusion matrix.
- Display the classification report.

## CODE

```
import numpy as np
import pandas as pd
from sklearn import svm
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.metrics import confusion_matrix, classification_report
from sklearn.model_selection import train_test_split
```

```

sns.set(font_scale=1.2)

recipes = pd.read_csv('recipes_muffins_cupcakes.csv')
print(recipes.head())
print(recipes.shape)

sns.lmplot(x='Sugar', y='Flour', data=recipes, hue='Type', palette='Set1',
fit_reg=False, scatter_kws={"s": 70})

sugar_butter = recipes[['Sugar', 'Flour']].values
type_label = np.where(recipes['Type'] == 'Muffin', 0, 1)

model = svm.SVC(kernel='linear')
model.fit(sugar_butter, type_label)

w = model.coef_[0]
a = -w[0] / w[1]
xx = np.linspace(5, 30)
yy = a * xx - (model.intercept_[0] / w[1])

b = model.support_vectors_[0]
yy_down = a * xx + (b[1] - a * b[0])

b = model.support_vectors_[-1]
yy_up = a * xx + (b[1] - a * b[0])

sns.lmplot(x='Sugar', y='Flour', data=recipes, hue='Type', palette='Set1',
fit_reg=False, scatter_kws={"s": 70})
plt.plot(xx, yy, linewidth=2, color='black')
plt.plot(xx, yy_down, 'k--')
plt.plot(xx, yy_up, 'k--')
plt.scatter(model.support_vectors_[0], model.support_vectors_[0], s=80,
facecolors='none')

x_train, x_test, y_train, y_test = train_test_split(sugar_butter, type_label,
test_size=0.2)

```

```

model1 = svm.SVC(kernel='linear')
model1.fit(x_train, y_train)
pred = model1.predict(x_test)

print(pred)
print(confusion_matrix(y_test, pred))
print(classification_report(y_test, pred, zero_division=1))

plt.show()

```

## OUTPUT

```

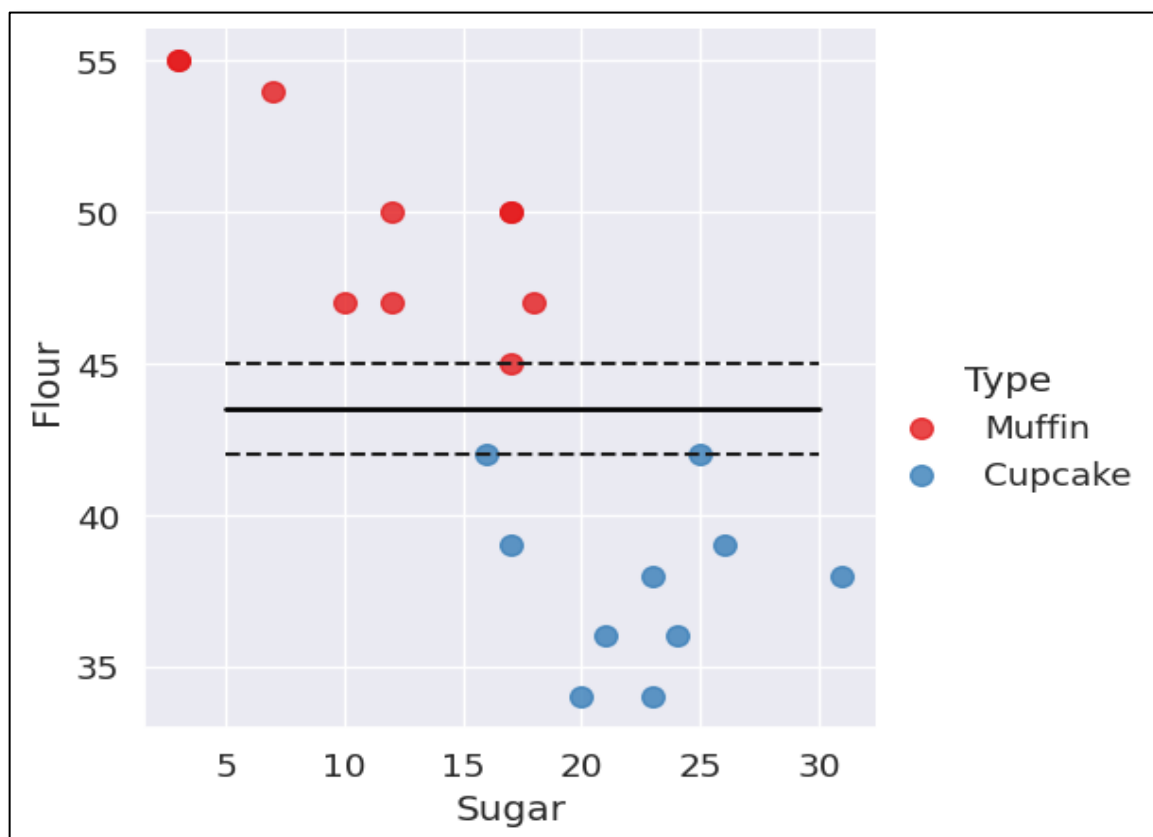
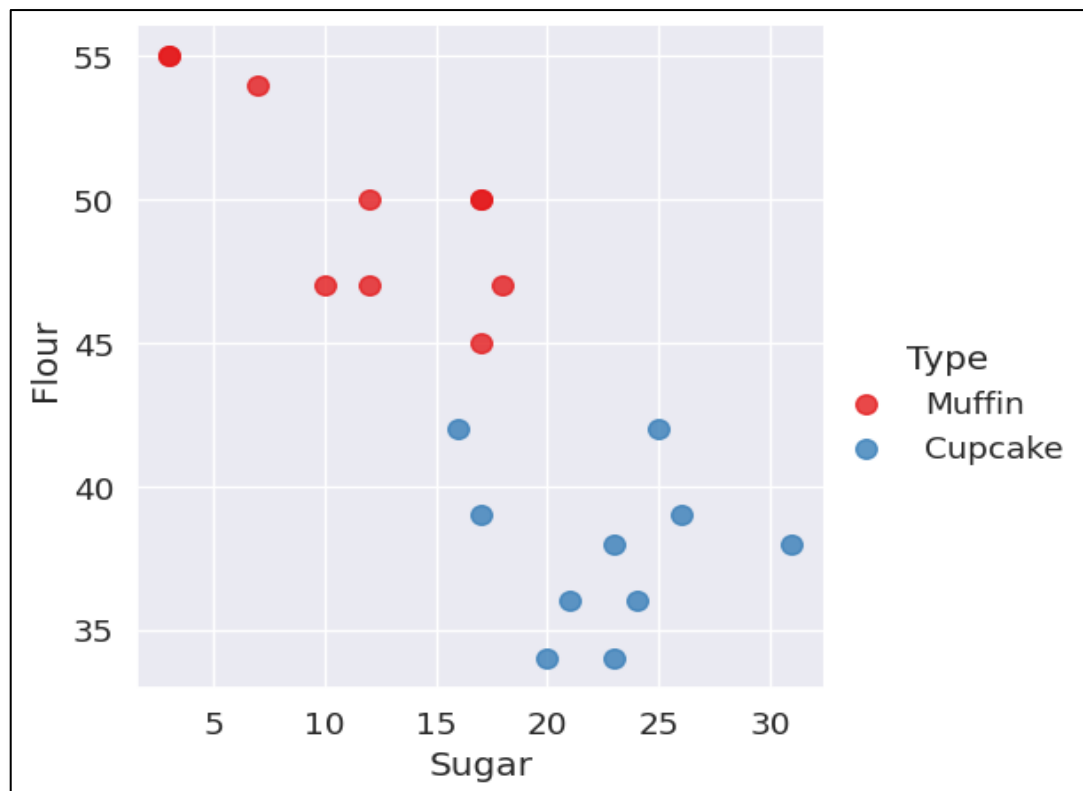
Type  Flour  Milk  Sugar  Butter  Egg  Baking Powder  Vanilla  Salt
0 Muffin   55   28    3     7    5             2        0    0
1 Muffin   47   24   12     6    9             1        0    0
2 Muffin   47   23   18     6    4             1        0    0
3 Muffin   45   11   17    17    8             1        0    0
4 Muffin   50   25   12     6    5             2        1    0
(20, 9)
[1 0 1 0]
[[2 0]
 [0 2]]

      precision    recall  f1-score   support

      0       1.00      1.00      1.00         2
      1       1.00      1.00      1.00         2

 accuracy          1.00          1.00          1.00          4
 macro avg          1.00          1.00          1.00          4
weighted avg          1.00          1.00          1.00          4

```



## **RESULT:**

Thus, the python program to implement SVM classifier model has been successfully implemented and the results have been verified and analysed.

EXPERIMENT NO : 7

DATE : 03/10/24

REGISTER NO : 231501502

NAME : DONEESWARAN J

---

## **A PYTHON PROGRAM TO IMPLEMENT DECISION TREE**

### **AIM:**

To implement a decision tree using a python program for the given dataset and plot the trained decision tree.

### **ALGORITHM:**

#### **STEP 1: IMPORT THE IRIS DATASET**

1. Import `'load_iris'` from `'sklearn.datasets'`.

#### **STEP 2: IMPORT NECESSARY LIBRARIES**

1. Import numpy as np.
2. Import matplotlib.pyplot as plt.
3. Import `'DecisionTreeClassifier'` from `'sklearn.tree'`.

#### **STEP 3: DECLARE AND INITIALIZE PARAMETERS**

1. Declare and initialize `'n_classes = 3'`.
2. Declare and initialize `'plot_colors = "ryb"'`.
3. Declare and initialize `'plot_step = 0.02'`.

#### **STEP 4: PREPARE DATA FOR MODEL TRAINING**

1. Load the iris dataset using `'load_iris()'`.
2. Assign the dataset's data to variable `'X'`.
3. Assign the dataset's target to variable `'Y'`.



## **STEP 5: TRAIN THE MODEL**

1. Create an instance of `'DecisionTreeClassifier'`.
2. Fit the classifier using `'clf.fit(X, Y)'`.

## **STEP 6: INITIALIZE PAIR INDEX AND PLOT GRAPH**

1. Loop through each pair of features using `'for pairidx, pair in enumerate(combinations(range(X.shape[1]), 2)):'`
2. Inside the loop, assign `'X'` with the selected pair of features (e.g., `'X = iris.data[:, pair]'`).
3. Assign `'Y'` with the target list (e.g., `'Y = iris.target'`).

## **STEP 7: ASSIGN AXIS LIMITS**

1. Inside the loop, assign `'x_min'` with the minimum value of the selected feature minus 1 (e.g., `'x_min, x_max = X[:, 0].min() - 1, X[:, 0].max() + 1'`).
2. Assign `'x_max'` with the maximum value of the selected feature plus 1.
3. Assign `'y_min'` with the minimum value of the second selected feature minus 1 (e.g., `'y_min, y_max = X[:, 1].min() - 1, X[:, 1].max() + 1'`).
4. Assign `'y_max'` with the maximum value of the second selected feature plus 1.

## **STEP 8: CREATE MESHGRID**

1. Use `'np.meshgrid'` to create a grid of values from `'x_min'` to `'x_max'` and `'y_min'` to `'y_max'` with steps of `'plot_step'`.
2. Assign the results to variables `'xx'` and `'yy'`.

## **STEP 9: PLOT GRAPH WITH TIGHT LAYOUT**

1. Use ``plt.tight_layout()`` to adjust the layout of the plots.
2. Set ``h_pad=0.5``, ``w_pad=0.5``, and ``pad=2.5``.

## **STEP 10: PREDICT AND RESHAPE**

1. Use the classifier to predict on the meshgrid (e.g., ``Z = clf.predict(np.c_[xx.ravel(), yy.ravel()])``).
2. Reshape ``Z`` to the shape of ``xx``.

## **STEP 11: PLOT DECISION BOUNDARY**

1. Use ``plt.contourf(xx, yy, Z, cmap=plt.cm.RdYlBu)`` to plot the decision boundary with the "RdYlBu" color scheme.

## **STEP 12: PLOT FEATURE PAIRS**

1. Inside the loop, label the x-axis and y-axis with the feature names (e.g., ``plt.xlabel(iris.feature_names[pair[0]])`` and ``plt.ylabel(iris.feature_names[pair[1]])``).

## **STEP 13: PLOT TRAINING POINTS**

1. Use ``plt.scatter(X[:, 0], X[:, 1], c=Y, cmap=plt.cm.RdYlBu, edgecolor='k', s=15)`` to plot the training points with the "RdYlBu" color scheme, black edge color, and size 15.

## **STEP 14: PLOT FINAL DECISION TREE**

1. Set the title of the plot to "Decision tree trained on all the iris features" (e.g., `plt.title("Decision tree trained on all the iris features")`).
2. Display the plot using `plt.show()`.

### CODE

```
from sklearn.datasets import load_iris
iris = load_iris()
import numpy as np
import matplotlib.pyplot as plt
from sklearn.tree import DecisionTreeClassifier

# Parameters
n_classes = 3
plot_colors = "ryb"
plot_step = 0.02
for pairidx, pair in enumerate([[0, 1], [0, 2], [0, 3], [1, 2], [1, 3], [2,
3]]):
    # We only take the two corresponding features
    X = iris.data[:, pair]
    y = iris.target

# Train
clf = DecisionTreeClassifier().fit(X, y)

# Plot the decision boundary
plt.subplot(2, 3, pairidx + 1)
x_min, x_max = X[:, 0].min() - 1, X[:, 0].max() + 1
y_min, y_max = X[:, 1].min() - 1, X[:, 1].max() + 1
xx, yy = np.meshgrid(np.arange(x_min, x_max, plot_step),
np.arange(y_min, y_max, plot_step))
plt.tight_layout(h_pad=0.5, w_pad=0.5, pad=2.5)
```

```

Z = clf.predict(np.c_[xx.ravel(), yy.ravel()])
Z = Z.reshape(xx.shape)
cs = plt.contourf(xx, yy, Z, cmap=plt.cm.RdYlBu)
plt.xlabel(iris.feature_names[pair[0]])
plt.ylabel(iris.feature_names[pair[1]])

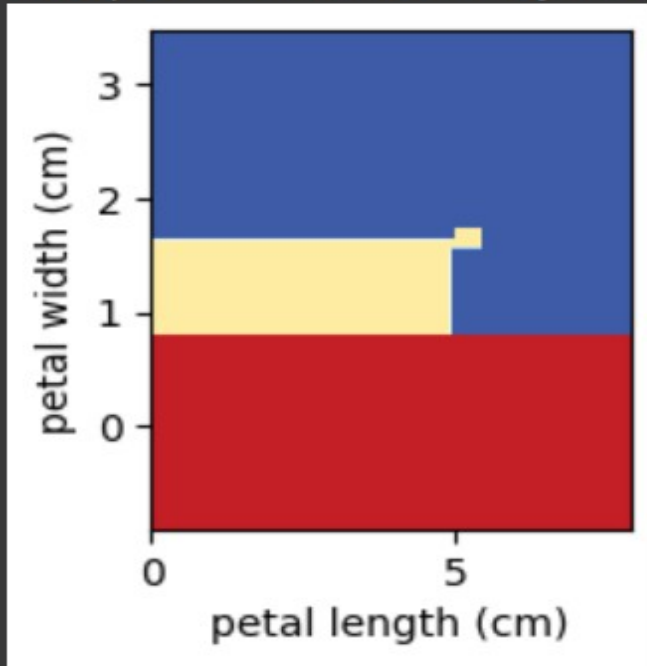
# Plot the training points
for i, color in zip(range(n_classes), plot_colors):
    idx = np.where(y == i)
    plt.scatter(X[idx, 0], X[idx,
1], c=color, label=iris.target_names[i], cmap=plt.cm.RdYlBu, edgecolor
="black", s=15)
plt.suptitle("Decision surface of decision trees trained on pairs of
features")
plt.legend(loc="lower right", borderpad=0, handletextpad=0)
plt.axis("tight")
plt.show()

from sklearn.tree import plot_tree
plt.figure()
clf = DecisionTreeClassifier().fit(iris.data, iris.target)
plot_tree(clf, filled=True)
plt.title("Decision tree trained on all the iris features")
plt.show()

```

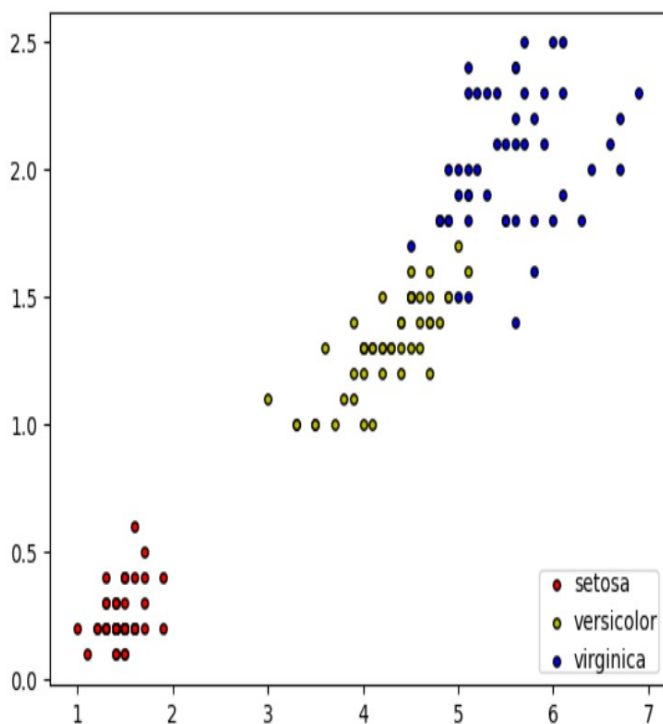
## OUTPUT

```
Text(392.83986928104576, 0.5, 'petal width (cm)')
```



```
<ipython-input-14-332623287b3a>:4: UserWarning: No data for colormapping provided via 'c'. Parameters 'cmap' will be ignored  
plt.scatter(X[idx, 0],X[idx, 1],c=color,label=iris.target_names[i],cmap=plt.cm.RdYlBu,edgecolor="black",s=15)
```

Decision surface of decision trees trained on pairs of features



## Decision tree trained on all the iris features



## RESULT

Thus, the python program to implement Decision Tree has been successfully implemented and the results have been verified and analysed.

EXPERIMENT NO : 8

DATE : 10/10/24

REGISTER NO : 231501502

NAME : DONEESWARAN J

---

## **A PYTHON PROGRAM TO IMPLEMENT ADA BOOSTING**

### **AIM:**

To implement a python program for Ada Boosting.

### **ALGORITHM:**

#### **STEP 1: IMPORT NECESSARY LIBRARIES**

Import numpy as np.

Import pandas as pd.

Import DecisionTreeClassifier from sklearn.tree.

Import train\_test\_split from sklearn.model\_selection.

Import accuracy\_score from sklearn.metrics.

#### **STEP 2: LOAD AND PREPARE DATA**

Load your dataset using pd.read\_csv() (e.g., df = pd.read\_csv('data.csv')).

Separate features (X) and target (y).

Split the dataset into training and testing sets using train\_test\_split().

#### **STEP 3: INITIALIZE PARAMETERS**

Set the number of weak classifiers n\_estimators.

Initialize an array weights for instance weights, setting each weight to 1 / number\_of\_samples.

#### **STEP 4: TRAIN WEAK CLASSIFIERS**

Loop for `n_estimators` iterations:

Train a weak classifier using `DecisionTreeClassifier(max_depth=1)` on the training data weighted by weights.

Predict the target values using the trained weak classifier.

Calculate the error rate `err` as the sum of weights of misclassified samples divided by the sum of all weights.

Compute the classifier's weight `alpha` using  $0.5 * \text{np.log}((1 - \text{err}) / \text{err})$ .

Update the weights: multiply the weights of misclassified samples by `np.exp(alpha)` and the weights of correctly classified samples by `np.exp(-alpha)`.

Normalize the weights so that they sum to 1.

Append the trained classifier and its weight to lists `classifiers` and `alphas`.

## **STEP 5: MAKE PREDICTIONS**

For each sample in the testing set:

Initialize a prediction score to 0.

For each trained classifier and its weight:

Add the classifier's prediction (multiplied by its weight) to the prediction score.

Take the sign of the prediction score as the final prediction.



## STEP 6: EVALUATE THE MODEL

Compute the accuracy of the AdaBoost model on the testing set using `accuracy_score()`.

## STEP 7: OUTPUT RESULTS

Print or plot the final accuracy and possibly other evaluation metrics.

### CODE

```
import pandas as pd
import numpy as np
from mlxtend.plotting import plot_decision_regions
df = pd.DataFrame()
df['X1']=[1,2,3,4,5,6,6,7,9,9]
df['X2']=[5,3,6,8,1,9,5,8,9,2]
df['label']=[1,1,0,1,0,1,0,1,0,0]
import seaborn as sns
sns.scatterplot(x=df['X1'],y=df['X2'],hue=df['label'])
df['weights']=1/df.shape[0]

from sklearn.tree import DecisionTreeClassifier

dt1 = DecisionTreeClassifier(max_depth=1)
x = df.iloc[:,0:2].values
y = df.iloc[:,2].values
# Step 2 - Train 1st Model

dt1.fit(x,y)
from sklearn.tree import plot_tree

plot_decision_regions(x,y,clf=dt1, legend=2)
df['y pred'] = dt1.predict(x)
```

```

def calculate_model_weight(error):
    return 0.5*np.log((1-error)/(error))

# Step - 3 Calculate model weight
alpha1 = calculate_model_weight(0.3)
alpha1

# Step -4 Update weights
def update_row_weights(row,alpha=0.423):
    if row['label'] == row['y_pred']:
        return row['weights']* np.exp(-alpha)
    else:
        return row['weights']* np.exp(alpha)
df['updated_weights'] = df.apply(update_row_weights,axis=1)

df['normalized_weights'] = df['updated_weights'] /
df['updated_weights'].sum() # Calculating normalized weights by
dividing updated weights by sum of all updated weights

df['normalized_weights'].sum()

df['cumsum_upper'] = np.cumsum(df['normalized_weights'])
df['cumsum_lower']=df['cumsum_upper'] - df['normalized_weights']
df[['X1','X2','label','weights','y_pred','updated_weights','cumsum_low
er','cumsum_upper']]

def create_new_dataset(df):
    indices= []
    for i in range(df.shape[0]):
        a = np.random.random()
        for index,row in df.iterrows():
            if row['cumsum_upper']>a and a>row['cumsum_lower']:
                indices.append(index)
    return indices
index_values = create_new_dataset(df)
index_values

```

```

second_df = df.iloc[index_values,[0,1,2,3]]
second_df

dt2 = DecisionTreeClassifier(max_depth=1)

x = second_df.iloc[:,0:2].values
y = second_df.iloc[:,2].values

dt2.fit(x,y)

plot_tree(dt2)

plot_decision_regions(x, y, clf=dt2, legend=2)

second_df['y_pred'] = dt2.predict(x)
second_df
alpha2 = calculate_model_weight(0.1)
alpha2

# Step 4 - Update weights
def update_row_weights(row,alpha=1.09):
    if row['label'] == row['y_pred']:
        return row['weights'] * np.exp(-alpha)
    else:
        return row['weights'] * np.exp(alpha)

second_df['updated_weights'] =
second_df.apply(update_row_weights,axis=1)
second_df
second_df['normalized_weights'] = second_df['updated_weights'] /
second_df['updated_weights'].sum()

second_df['normalized_weights'].sum()
second_df['cumsum_upper'] =
np.cumsum(second_df['normalized_weights'])

```

```
second_df['cumsum_lower'] = second_df['cumsum_upper'] -  
second_df['normalized_weights']  
second_df[['X1','X2','label','weights','y_pred','normalized_weights','cu  
msum_lower','cumsum_upper']]
```

```
alpha3 = calculate_model_weight(0.7)  
alpha3
```

```
from sklearn.tree import DecisionTreeClassifier
```

```
print(alpha1,alpha2,alpha3)
```

```
dt3 = DecisionTreeClassifier(max_depth=2)
```

```
# Fit dt3 before making predictions  
dt3.fit(x, y) # Assuming 'x' and 'y' are your training data from  
previous cells.
```

```
query = np.array([1,5]).reshape(1,2)  
dt1.predict(query)  
dt2.predict(query)  
dt3.predict(query)
```

```
alpha1*1 + alpha2*(1) + alpha3*(1)
```

```
np.sign(1.09)
```

```
query = np.array([9,9]).reshape(1,2)  
dt1.predict(query)
```

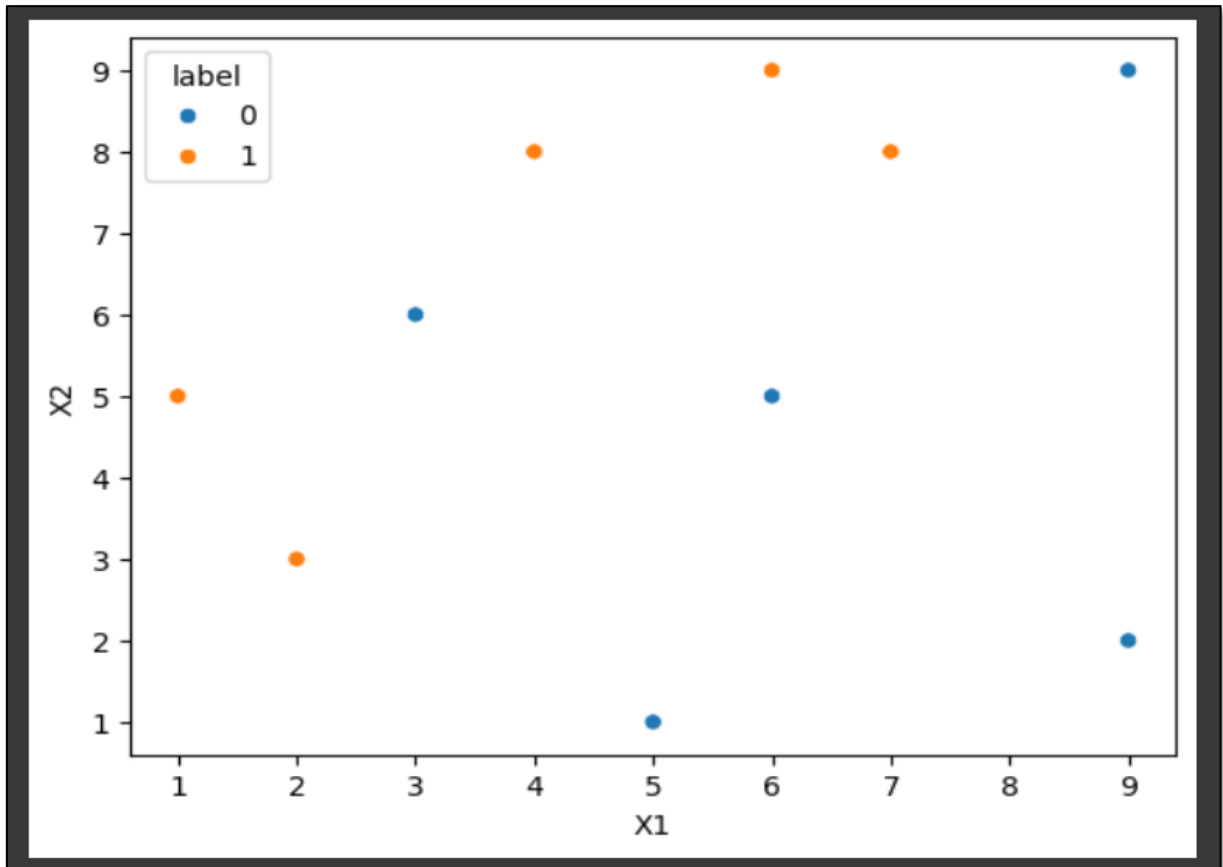
```
dt2.predict(query)
```

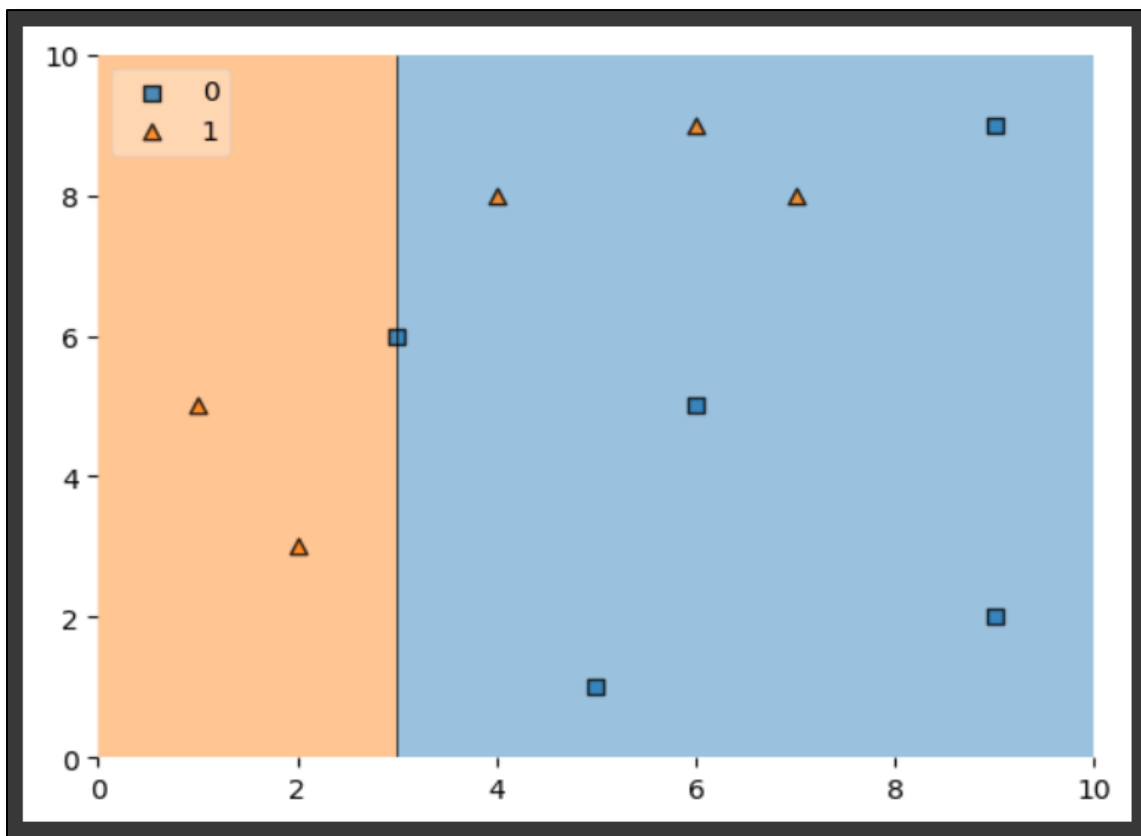
```
dt3.predict(query)
```

```
alpha1*(1) + alpha2*(-1) + alpha3*(-1)
```

```
np.sign(-0.25)
```

## OUTPUT

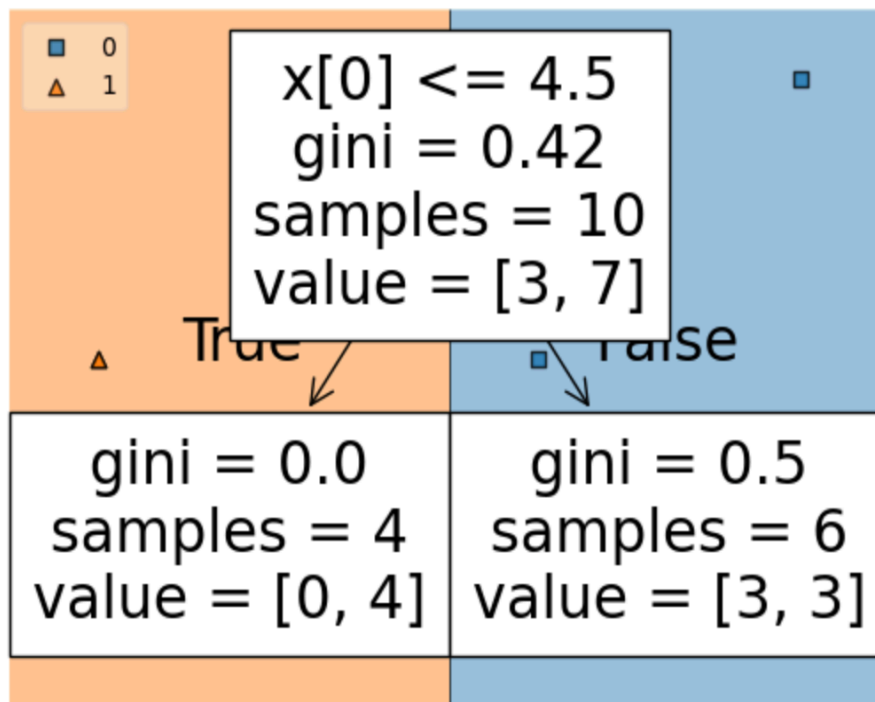




	X1	X2	label	weights	y_pred	updated_weights	cumsum_lower	cumsum_upper
0	1	5	1	0.1	1	0.065508	0.000000	0.071475
1	2	3	1	0.1	1	0.065508	0.071475	0.142950
2	3	6	0	0.1	0	0.065508	0.142950	0.214425
3	4	8	1	0.1	0	0.152653	0.214425	0.380983
4	5	1	0	0.1	0	0.065508	0.380983	0.452458
5	6	9	1	0.1	0	0.152653	0.452458	0.619017
6	6	5	0	0.1	0	0.065508	0.619017	0.690492
7	7	8	1	0.1	0	0.152653	0.690492	0.857050
8	9	9	0	0.1	0	0.065508	0.857050	0.928525
9	9	2	0	0.1	0	0.065508	0.928525	1.000000

	X1	X2	label	weights
5	6	9	1	0.1
0	1	5	1	0.1
0	1	5	1	0.1
9	9	2	0	0.1
5	6	9	1	0.1
9	9	2	0	0.1
3	4	8	1	0.1
0	1	5	1	0.1
3	4	8	1	0.1
0	1	5	1	0.1

0.42364893019360184 1.0986122886681098 -0.4236489301936017  
-1.0



0.9999999999999999

-0.4236489301936017

0.42364893019360184 1.0986122886681098 -0.4236489301936017  
-1.0



## **RESULT**

Thus the python program to implement ADA Boosting has been successfully implemented and the results have been verified and analyzed.

**EXPERIMENT NO : 9 A**

**DATE : 17/10/24**

**REGISTER NO : 231501502**

**NAME : DONEESWARAN J**

---

## **A PYTHON PROGRAM TO IMPLEMENT KNN MODEL**

### **AIM:**

To implement a python program using a KNN Algorithm in a model.

### **ALGORITHM:**

#### **STEP 1 : IMPORT NECESSARY LIBRARIES**

- Import necessary libraries: pandas, numpy, train\_test\_split from sklearn.model\_selection,
- StandardScaler from sklearn.preprocessing,
- KNeighborsClassifier from sklearn.neighbors,  
and classification\_report and confusion\_matrix from sklearn.metrics.

#### **STEP 2 : LOAD AND EXPLORE THE DATASET**

- Load the dataset using pandas.
- Display the first few rows of the dataset using df.head().
- Display the dimensions of the dataset using df.shape().
- Display the descriptive statistics of the dataset using df.describe().

### **STEP 3 : PREPROCESS THE DATA**

- Separate the features (X) and the target variable (y).
- Split the data into training and testing sets using `train_test_split`.
- Standardize the features using `StandardScaler`.

### **STEP 4 : TRAIN THE KNN MODEL**

- Create an instance of `KNeighborsClassifier` with a specified number of neighbors (k).
- For each data point, calculate the Euclidean distance to all other data points.
- Select the K nearest neighbors based on the calculated Euclidean distances.
- Among the K nearest neighbors, count the number of data points in each category.
- Assign the new data point to the category for which the number of neighbors is maximum.

### **STEP 5 : MAKE PREDICTIONS**

- Use the trained model to make predictions on the test data.
- Evaluate the Model
- Generate the confusion matrix and classification report using the actual and predicted values.
- Print the confusion matrix and classification report.

## CODE

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

dataset = pd.read_csv('../input/mall-customers/Mall_Customers.csv')
X = dataset.iloc[:,[3,4]].values
print(dataset)

from sklearn.cluster import KMeans
wcss = []
for i in range(1,11):
    kmeans = KMeans(n_clusters = i, init = 'k-means++', max_iter = 300,
n_init = 10, random_state = 0)
    kmeans.fit(X)
    wcss.append(kmeans.inertia_)

# Plot the graph to visualize the Elbow Method to find the optimal
number of cluster
plt.plot(range(1,11),wcss)
plt.title('The Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.show()
```

```
kmeans=KMeans(n_clusters= 5, init = 'k-means++', max_iter = 300,  
n_init = 10, random_state = 0)  
y_kmeans = kmeans.fit_predict(X)  
y_kmeans
```

```
type(y_kmeans)
```

```
y_kmeans
```

```
plt.scatter(X[y_kmeans == 0, 0], X[y_kmeans == 0, 1], s = 100, c =  
'red', label = 'Cluster 1')
```

```
plt.scatter(X[y_kmeans == 1, 0], X[y_kmeans == 1, 1], s = 100, c =  
'blue', label = 'Cluster 2')
```

```
plt.scatter(X[y_kmeans == 2, 0], X[y_kmeans == 2, 1], s = 100, c =  
'green', label = 'Cluster 3')
```

```
plt.scatter(X[y_kmeans == 3, 0], X[y_kmeans == 3, 1], s = 100, c =  
'cyan', label = 'Cluster 4')
```

```
plt.scatter(X[y_kmeans == 4, 0], X[y_kmeans == 4, 1], s = 100, c =  
'magenta', label = 'Cluster 5')
```

```
plt.scatter(kmeans.cluster_centers_[:, 0], kmeans.cluster_centers_[:, 1], s = 300, c = 'yellow', label = 'Centroids')
```

```
plt.title('Clusters of customers')
```

```
plt.xlabel('Annual Income (k$)')
```

```
plt.ylabel('Spending Score (1-100)')
```

```
plt.legend()
```

```
plt.show()
```

```
plt.scatter(X[y_kmeans == 0, 0], X[y_kmeans == 0, 1], s = 100, c = 'red', label = 'Cluster 1')
```

```
plt.scatter(X[y_kmeans == 1, 0], X[y_kmeans == 1, 1], s = 100, c = 'blue', label = 'Cluster 2')
```

```
plt.scatter(X[y_kmeans == 2, 0], X[y_kmeans == 2, 1], s = 100, c = 'green', label = 'Cluster 3')
```

```
plt.scatter(X[y_kmeans == 3, 0], X[y_kmeans == 3, 1], s = 100, c = 'cyan', label = 'Cluster 4')
```

```
plt.scatter(X[y_kmeans == 4, 0], X[y_kmeans == 4, 1], s = 100, c = 'magenta', label = 'Cluster 5')
```

```
plt.scatter(kmeans.cluster_centers_[:, 0], kmeans.cluster_centers_[:, 1], s = 300, c = 'yellow', label = 'Centroids')
```

```
plt.title('Clusters of customers')
```

```
plt.xlabel('Annual Income (k$)')
```

```
plt.ylabel('Spending Score (1-100)')
```

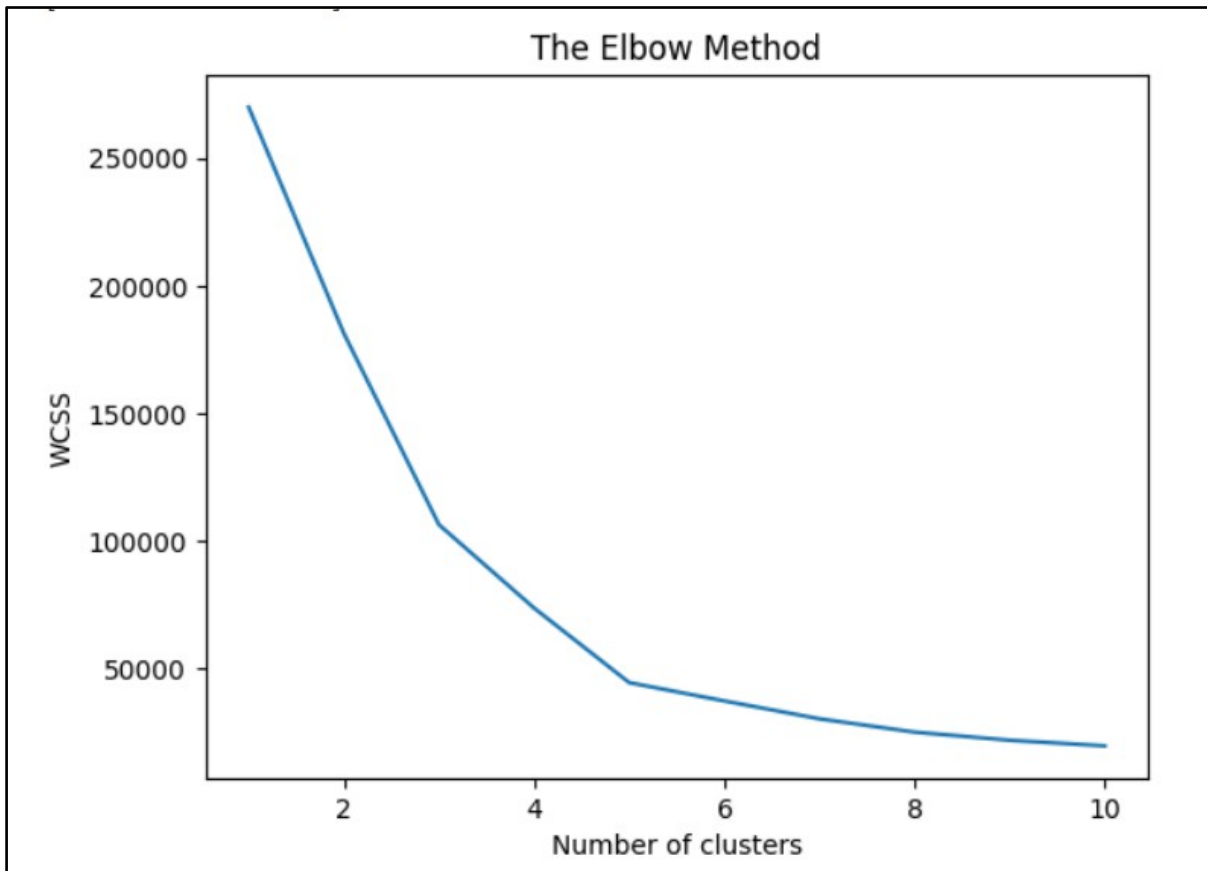
```
plt.legend()
```

```
plt.show()
```

## OUTPUT

	CustomerID	Gender	Age	Annual Income (k\$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40
..	...	...	...	...	...
195	196	Female	35	120	79
196	197	Female	45	126	28
197	198	Male	32	126	74
198	199	Male	32	137	18
199	200	Male	30	137	83

[200 rows x 5 columns]





## **RESULT**

Thus the python program to implement KNN model has been successfully implemented and the results have been verified and analyzed.



**EXPERIMENT NO : 9 B**

**DATE : 24/10/24**

**REGISTER NO : 231501502**

**NAME : DONEESWARAN J**

---

### **A PYTHON PROGRAM TO IMPLEMENT K-MEANS MODEL**

#### **AIM:**

To implement a python program using a K-Means Algorithm in a model.

#### **ALGORITHM:**

##### **STEP 1 : IMPORT NECESSARY LIBRARIES.**

Import required libraries like numpy, matplotlib.pyplot, and sklearn.cluster.

##### **STEP 2 : LOAD AND PREPROCESS DATA.**

Load the dataset.

Preprocess the data if needed (e.g., scaling).

##### **STEP 3 : INITIALIZE CLUSTER CENTERS.**

Choose the number of clusters (K).

Initialize K cluster centers randomly.

##### **STEP 4 : ASSIGN DATA POINTS TO CLUSTERS.**

For each data point, calculate the distance to each cluster center.  
Assign the data point to the cluster with the nearest center.

## **STEP 5 : UPDATE CLUSTER CENTERS**

Calculate the mean of the data points in each cluster.

Update the cluster centers to the calculated means.

## **STEP 6 : REPEAT STEPS 4 AND 5**

Repeat the assignment of data points to clusters and updating of cluster centers until convergence (i.e., when the cluster assignments do not change much between iterations).

## **STEP 7 :.PLOT THE CLUSTERS**

Plot the data points and the cluster centers to visualize the clustering result.

### **CODE**

```
import numpy as np
import pandas as pd

# Load the dataset
data = pd.read_csv('/content/drive/MyDrive/knn.csv')
data.head(5)

# Prepare the data
req_data = data.iloc[:, 1:]
req_data.head(5)
```

```

# Shuffle the data
shuffle_index = np.random.permutation(req_data.shape[0])
req_data = req_data.iloc[shuffle_index]
req_data.head(5)

# Split into training and testing sets
train_size = int(req_data.shape[0] * 0.7)
train_df = req_data.iloc[:train_size, :]
test_df = req_data.iloc[train_size:, :]
train = train_df.values
test = test_df.values
y_true = test[:, -1]

print('Train_Shape: ', train_df.shape)
print('Test_Shape: ', test_df.shape)

# Define Euclidean distance function
from math import sqrt
def euclidean_distance(x_test, x_train):
    distance = 0
    for i in range(len(x_test) - 1):
        distance += (x_test[i] - x_train[i]) ** 2
    return sqrt(distance)

```

```

# Define function to get nearest neighbors
def get_neighbors(x_test, x_train, num_neighbors):
    distances = []
    data = []
    for i in x_train:
        distances.append(euclidean_distance(x_test, i))
        data.append(i)
    distances = np.array(distances)
    data = np.array(data)
    sort_indexes = distances.argsort() # argsort() returns sorted indices
    data = data[sort_indexes]        # Sort data by distances
    return data[:num_neighbors]

```

```

# Define prediction function
def prediction(x_test, x_train, num_neighbors):
    classes = []
    neighbors = get_neighbors(x_test, x_train, num_neighbors)
    for i in neighbors:
        classes.append(i[-1])
    predicted = max(classes, key=classes.count) # Most common class
    return predicted

```

```

# Define function to predict classes
def predict_classifier(x_test):
    classes = []

```

```

neighbors = get_neighbors(x_test, req_data.values, 5)
for i in neighbors:
    classes.append(i[-1])
predicted = max(classes, key=classes.count)
print(predicted)
return predicted

# Define accuracy calculation
def accuracy(y_true, y_pred):
    num_correct = 0
    for i in range(len(y_true)):
        if y_true[i] == y_pred[i]:
            num_correct += 1
    accuracy = num_correct / len(y_true)
    return accuracy

# Predict on test data
y_pred = []
for i in test:
    y_pred.append(prediction(i, train, 5))

# Display predictions and accuracy
y_pred
accuracy = accuracy(y_true, y_pred)
accuracy

```

## **OUTPUT**

```
Train_Shape: (17, 2)  
Test_Shape: (8, 2)
```

---

## **RESULT**

Thus the python program to implement the K-Means model has been successfully implemented and the results have been verified and analyzed

**EXPERIMENT NO : 10**

**DATE : 07/11/24**

**REGISTER NO : 231501502**

**NAME : DONEESWARAN J**

---

## **A PYTHON PROGRAM TO IMPLEMENT DIMENSIONALITY REDUCTION USING PCA**

### **AIM:**

To implement Dimensionality Reduction using PCA in a python program.

### **ALGORITHM:**

#### **STEP 1: IMPORT LIBRARIES**

Import necessary libraries, including pandas, numpy, matplotlib.pyplot, and sklearn.decomposition.PCA.

#### **STEP 2: LOAD THE DATASET (IRIS DATASET)**

Load your dataset into a pandas DataFrame.

#### **STEP 3: STANDARDIZE THE DATA**

Standardize the features of the dataset using StandardScaler from sklearn.preprocessing.

#### **STEP 4: APPLY PCA**

- Create an instance of PCA with the desired number of components.
- Fit PCA to the standardized data.
- Transform the data to its principal components using transform.

## **STEP 5: EXPLAINED VARIANCE RATIO**

- Calculate the explained variance ratio for each principal component.
- Plot a scree plot to visualize the explained variance ratio.

## **STEP 6: CHOOSE THE NUMBER OF COMPONENTS**

Based on the scree plot, choose the number of principal components that explain a significant amount of variance.

## **STEP 7: APPLY PCA WITH CHOSEN COMPONENTS**

Apply PCA again with the chosen number of components.

## **STEP 8: VISUALIZE THE REDUCED DATA**

- Transform the original data to the reduced dimension using the fitted PCA.
- Visualize the reduced data using a scatter plot.

## **STEP 9: INTERPRETATION**

Interpret the results, considering the trade-offs between dimensionality reduction and information loss.



## CODE

```
from sklearn import datasets
import pandas as pd
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
import seaborn as sns

iris = datasets.load_iris()
df = pd.DataFrame(iris['data'], columns = iris['feature_names'])
df.head()

scalar = StandardScaler()
scaled_data = pd.DataFrame(scalar.fit_transform(df)) #scaling the data
scaled_data

sns.heatmap(scaled_data.corr())

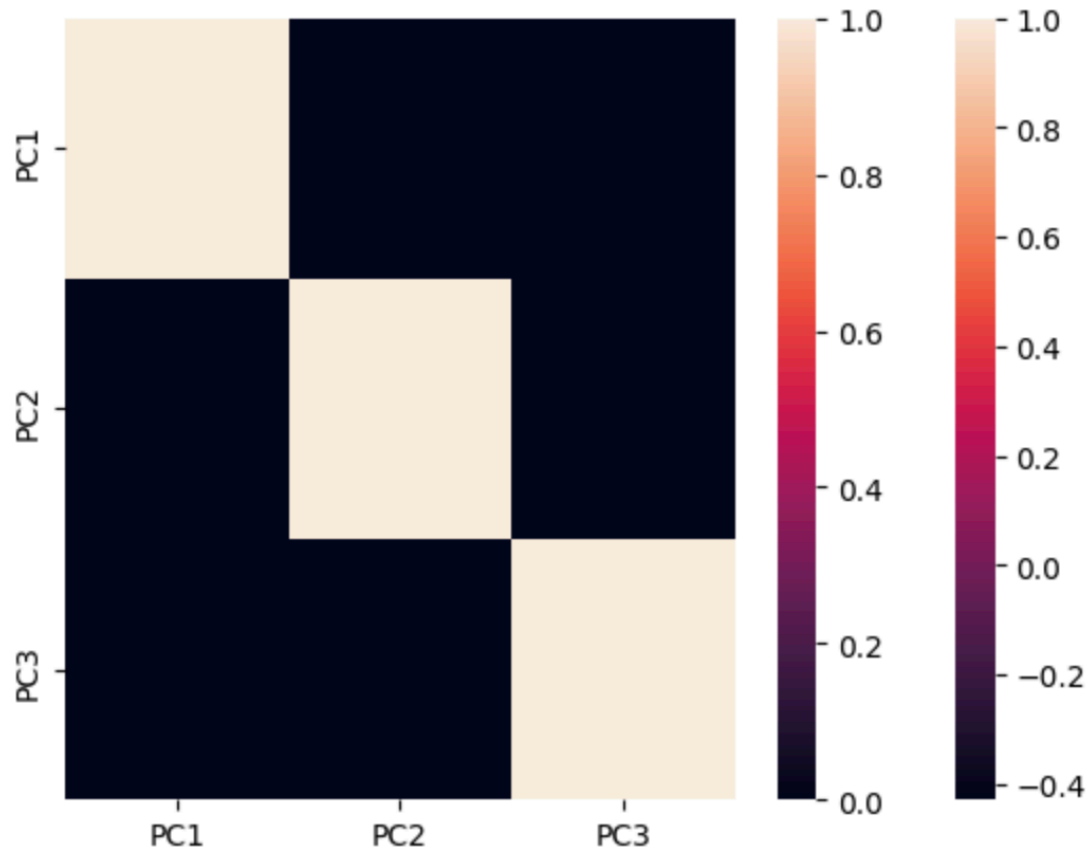
pca = PCA(n_components = 3)
pca.fit(scaled_data)
data_pca = pca.transform(scaled_data)
data_pca = pd.DataFrame(data_pca, columns=['PC1','PC2','PC3'])
data_pca.head()
sns.heatmap(data_pca.corr())
```

## OUTPUT

	CustomerID	Gender	Age	Annual Income (k\$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40
..	...	...	...	...	...
195	196	Female	35	120	79
196	197	Female	45	126	28
197	198	Male	32	126	74
198	199	Male	32	137	18
199	200	Male	30	137	83

[200 rows x 5 columns]

<Axes: >



**RESULT**

Thus the python program to implement Dimensionality Reduction using PCA has been successfully implemented and the results have been verified and analyzed