MLA0413-DEEP LEARNING FOR COMPLEX DATA MINING

NAME: SAANIYA SHADAAP

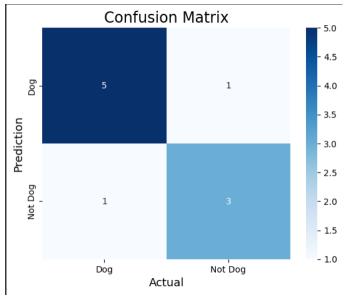
REG NO: 192225004

EXPERIMENT:1(A)

AIM: To demonstrate confusion matrix using python

PROGRAM:

```
import numpy as np
from sklearn.metrics import confusion matrix
import seaborn as sns
import matplotlib.pyplot as plt
        = np.array(
Dog','Not Dog'])
predicted = np.array(
Dog', 'Not Dog'])
cm = confusion matrix(actual, predicted)
sns.heatmap(cm,
            xticklabels=['Dog','Not Dog'],
            yticklabels=['Dog','Not Dog'])
plt.ylabel('Prediction', fontsize=13)
plt.xlabel('Actual', fontsize=13)
plt.title('Confusion Matrix',fontsize=17)
plt.show()
```

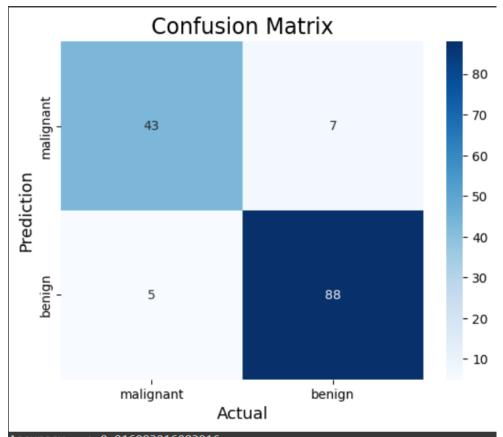


EXPERIMENT:1(B)

AIM: To demonstrate 2 class confusion matrix using python

```
from sklearn.datasets import load breast cancer
from sklearn.model selection import train test split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import confusion matrix
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.metrics import accuracy score, precision score,
recall score, f1 score
X, y= load breast cancer(return X y=True)
X train, X test, y train, y test = train test split(X,
y, test size=0.25)
# Train the model
tree = DecisionTreeClassifier(random state=23)
tree.fit(X_train, y_train)
y pred = tree.predict(X test)
cm = confusion matrix(y test, y pred)
#Plot the confusion matrix.
sns.heatmap(cm,
            annot=True,
            xticklabels=['malignant', 'benign'],
            yticklabels=['malignant', 'benign'])
plt.ylabel('Prediction', fontsize=13)
plt.xlabel('Actual', fontsize=13)
plt.title('Confusion Matrix', fontsize=17)
plt.show()
# Finding precision and recall
accuracy = accuracy score(y test, y pred)
print("Accuracy :", accuracy)
```

```
precision = precision_score(y_test, y_pred)
print("Precision :", precision)
recall = recall_score(y_test, y_pred)
print("Recall :", recall)
F1_score = f1_score(y_test, y_pred)
print("F1-score :", F1_score)
```

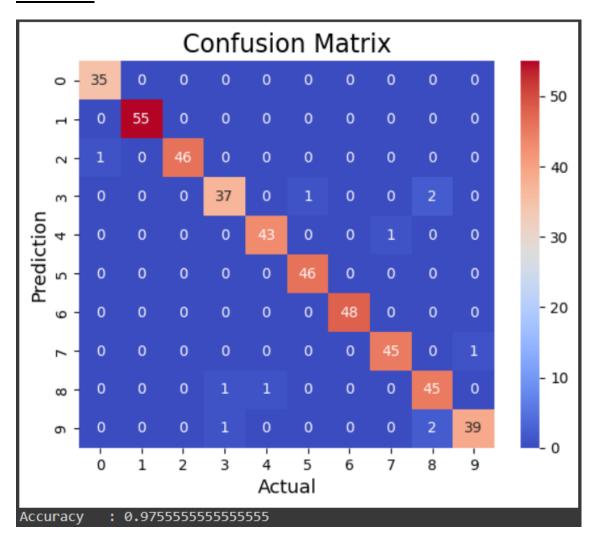


Accuracy : 0.916083916083916
Precision : 0.9263157894736842
Recall : 0.946236559139785
F1-score : 0.9361702127659575

EXPERIMENT: 2

<u>AIM:</u> Verifying the performance of a multi class confusion matrix by using chosen database with python code

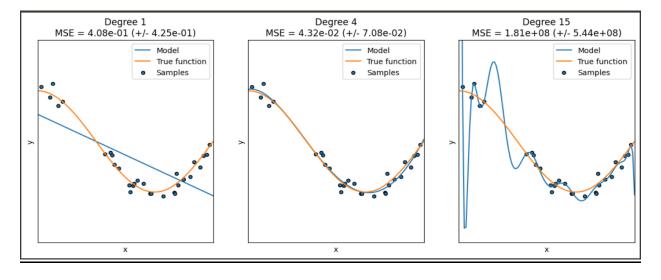
```
from sklearn.datasets import load digits
from sklearn.model selection import train test split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import confusion matrix
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.metrics import accuracy score, precision score,
recall score, f1 score
X, y= load digits(return X y=True)
X train, X test, y train, y test = train test split(X,
y, test size=0.25)
clf = RandomForestClassifier(random state=23)
y pred = clf.predict(X test)
cm = confusion matrix(y test, y pred)
sns.heatmap(cm,
            annot=True,
plt.ylabel('Prediction', fontsize=13)
plt.xlabel('Actual', fontsize=13)
plt.title('Confusion Matrix', fontsize=17)
plt.show()
# Finding precision and recall
accuracy = accuracy score(y test, y pred)
print("Accuracy :", accuracy)
```



EXPERIMENT:3

<u>AIM:</u>: Verifying the performance of an overfitting by using chosen database with python code

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import PolynomialFeatures
from sklearn.linear model import LinearRegression
from sklearn.model selection import cross val score
def true fun(X):
    return np.cos(1.5 * np.pi * X)
np.random.seed(0)
n \text{ samples} = 30
degrees = [1, 4, 15]
X = np.sort(np.random.rand(n samples))
y = true fun(X) + np.random.randn(n samples) * 0.1
plt.figure(figsize=(14, 5))
for i in range(len(degrees)):
    ax = plt.subplot(1, len(degrees), i + 1)
    plt.setp(ax, xticks=(), yticks=())
    polynomial features = PolynomialFeatures(degree=degrees[i],
include bias=False)
    linear_regression = LinearRegression()
    pipeline = Pipeline(
            ("polynomial features", polynomial features),
            ("linear regression", linear regression),
    pipeline.fit(X[:, np.newaxis], y)
    scores = cross val score(
        pipeline, X[:, np.newaxis], y,
scoring="neg mean squared error", cv=10
```

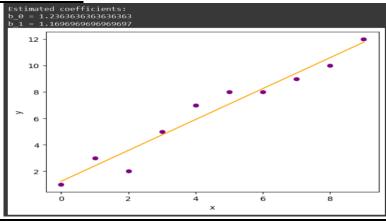


EXPERIMENT:4

<u>AIM:</u> To demonstrate the performance of a linear regression by using chosen database with python code

PROGRAM: LINEAR REGRESSION

```
import numpy as np
import matplotlib.pyplot as plt
def estimate_coef(x, y):
    n = np.size(x)
    m_x = np.mean(x)
    m_y = np.mean(y)
    SS_xy = np.sum(y*x) - n*m_y*m_x
    SS xx = np.sum(x*x) - n*m x*m x
```

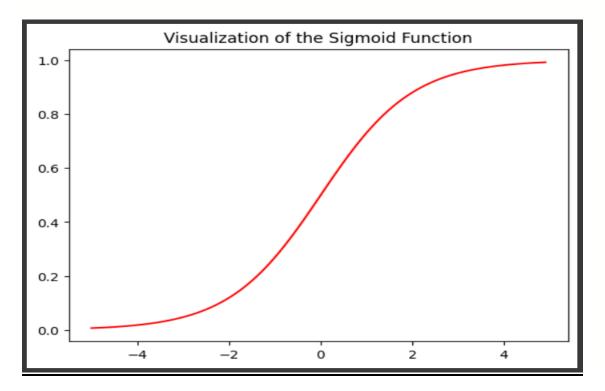


EXPERIMENT:5

<u>AIM:</u> To demonstrate the performance of a logistic regression by using chosen database with python code.

PROGRAM:

```
import numpy as np
import matplotlib.pyplot as plt
def sigmoid(z):
    return 1 / (1 + np.exp( - z))
plt.plot(np.arange(-5, 5, 0.1), sigmoid(np.arange(-5, 5, 0.1)))
plt.title('Visualization of the Sigmoid Function')
plt.show()
```



EXPERIMENT:6(a)KNN

AIM: Finding accuracy value of iris data set using KNN algorithm

PROGRAM:

```
import numpy as np
import pandas as pd
dataset = pd.read csv("/Iris.csv")
X = dataset.iloc[:, :-1].values
y = dataset.iloc[:, -1].values
dataset.shape
from sklearn.model selection import train test split
X train, X test, y train, y test = train test split(X, y, test size =
0.20, random state = 42)
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X train = sc.fit transform(X train)
X test = sc.transform(X test)
from sklearn.neighbors import KNeighborsClassifier
classifier = KNeighborsClassifier(n neighbors = 5, metric =
from sklearn.metrics import confusion matrix, accuracy score
y pred = classifier.predict(X test)
cm = confusion matrix(y test, y pred)
print(cm)
accuracy score(y test, y pred)
```

```
[[85 0]
[ 2 50]]
0.9854014598540146
```

EXPERIMENT:6(B)NAVIE

AIM: : finding accuracy value of iris data set using NAVIE BAYES algorithm

PROGRAM:

```
import numpy as np
import pandas as pd
dataset = pd.read csv("/Iris.csv")
X = dataset.iloc[:, :-1].values
y = dataset.iloc[:, -1].values
from sklearn.model selection import train test split
X train, X test, y train, y test = train test split(X, y, test size =
from sklearn.preprocessing import StandardScaler
X train = sc.fit transform(X train)
X test = sc.transform(X test)
from sklearn.naive bayes import GaussianNB
classifier = GaussianNB()
classifier.fit(X train, y train)
GaussianNB(priors=None, var smoothing=1e-09)
from sklearn.metrics import confusion matrix, accuracy score
y pred = classifier.predict(X test)
cm = confusion matrix(y test, y pred)
print(cm)
accuracy score(y test, y pred)
```

```
[[114 2]
[ 2 53]]
0.9766081871345029
```

EXPERIMENT:6(C)LOGISTIC

<u>AIM:</u>: finding accuracy value of iris data set using LOGISTIC REGRESSION algorithm

PROGRAM:

```
import numpy as np
import pandas as pd
dataset = pd.read csv("/Iris.csv")
X = dataset.iloc[:, :-1].values
from sklearn.model selection import train test split
X train, X test, y train, y test = train test split(X, y, test size =
0.30, random state = 2)
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X train = sc.fit transform(X train)
X test = sc.transform(X test)
from sklearn.linear model import LogisticRegression
classifier = LogisticRegression(random state = 0)
classifier.fit(X train, y train)
LogisticRegression (C=1.0, class weight=None, dual=False,
fit intercept=True,
                   intercept scaling=1, 11 ratio=None, max iter=100,
                   multi class='warn', n jobs=None, penalty='12',
verbose=0,
                   warm start=False)
from sklearn.metrics import confusion matrix, accuracy score
y pred = classifier.predict(X test)
cm = confusion matrix(y test, y pred)
print(cm)
accuracy score(y test, y pred)
```

```
[[117 8]
[ 6 74]]
0.9317073170731708
```

EXPERIMENT:6(D)DECISION

<u>AIM: : finding accuracy value of iris data set using DECISION TREE algorithm</u>

```
import numpy as np
import pandas as pd
dataset = pd.read csv("/Iris.csv")
X = dataset.iloc[:, :-1].values
y = dataset.iloc[:, -1].values
from sklearn.model selection import train test split
X train, X test, y train, y test = train test split(X, y, test size =
0.25, random state = 8)
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X train = sc.fit transform(X train)
X test = sc.transform(X test)
from sklearn.tree import DecisionTreeClassifier
= 5)
classifier.fit(X train, y train)
from sklearn.tree import plot tree
import matplotlib.pyplot as plt
plt.figure(figsize=(20,10))
plot tree(classifier, filled=True, rounded=True,
feature names=dataset.columns[:-1])
plt.show()
y pred = classifier.predict(X test)
from sklearn.metrics import confusion matrix, accuracy score
cm = confusion matrix(y test, y pred)
print(cm)
accuracy score(y test, y pred)
```

Output:

```
petal_width <= -0.533
                           entropy = 1.585
samples = 112
value = [37, 38, 37]
                                                         petal_width <= 0.71
  entropy = 0.0
                                                              entropy = 1.0
samples = 37
value = [37, 0, 0]
                                                          samples = 75
value = [0, 38, 37]
                          petal_length <= 0.732
                                                                                            entropy = 0.0
samples = 36
                             entropy = 0.172
samples = 39
value = [0, 38, 1]
                                                                                          value = [0, 0, 36]
                                                        sepal_width <= -0.658
  entropy = 0.0
                                                              entropy = 1.0
samples = 37
value = [0, 37, 0]
                                                            samples = 2
value = [0, 1, 1]
                                                                                           entropy = 0.0
samples = 1
value = [0, 0, 1]
                                entropy = 0.0
                               samples = 1
value = [0, 1, 0]
```

```
[[13 0 0]
[ 0 12 0]
[ 0 0 13]]
1.0
```

EXPERIMENT:6(E)SVM

AIM: : finding accuracy value of iris data set using SVM algorithm

PROGRAM:

```
import numpy as np
import pandas as pd
dataset = pd.read csv("/Iris.csv")
X = dataset.iloc[:, :-1].values
y = dataset.iloc[:, -1].values
from sklearn.model selection import train test split
X train, X test, y train, y test = train test split(X, y,
test size=0.25, random state=32)
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X train = sc.fit transform(X train)
X test = sc.transform(X test)
from sklearn.svm import SVC
classifier = SVC(kernel='linear', random state=0)
classifier.fit(X train, y train)
y pred = classifier.predict(X test)
cm = confusion matrix(y test, y pred)
print(cm)
print('Accuracy: {:.2f}%'.format(accuracy score(y test, y pred) * 100))
```

```
[[108 1]
[ 5 57]]
Accuracy: 96.49%
```

EXPERIMENT:6(F)RANDOM

AIM:: finding accuracy value of iris data set using RANDOM FOREST algorithm

PROGRAM:

```
import numpy as np
import pandas as pd
dataset = pd.read csv("/Iris.csv")
X = dataset.iloc[:, :-1].values
y = dataset.iloc[:, -1].values
from sklearn.model selection import train test split
X train, X test, y train, y test = train test split(X, y,
test size=0.25, random state=39)
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X train = sc.fit transform(X train)
X test = sc.transform(X test)
from sklearn.ensemble import RandomForestClassifier
classifier.fit(X train, y train)
y pred = classifier.predict(X test)
cm = confusion matrix(y test, y pred)
print(cm)
print('Accuracy:', accuracy score(y test, y pred))
```

```
[[111 1]
[ 2 57]]
Accuracy: 0.98245614
```

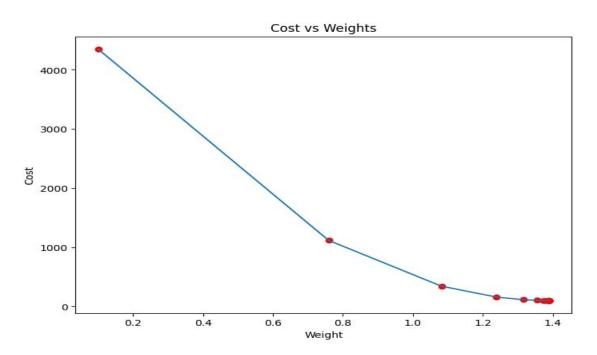
EXPERIMENT:7(A)

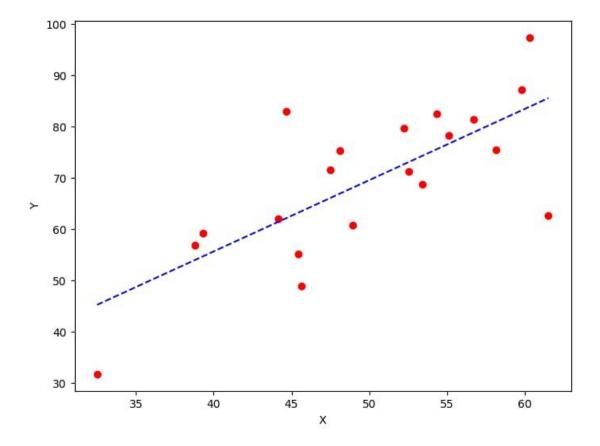
AIM: To demonstrate gradient descent using python(actual data)

```
import numpy as np
import matplotlib.pyplot as plt
def mean squared error(y true, y predicted):
    cost = np.sum((y true - y predicted)**2) / len(y true)
    return cost
def gradient descent(x, y, iterations=1000, learning rate=0.0001,
stopping threshold=1e-6):
    current weight = 0.1
    current bias = 0.01
    n = float(len(x))
    costs = []
    weights = []
    previous cost = None
    for i in range(iterations):
        y predicted = (current weight * x) + current bias
        current cost = mean squared error(y, y predicted)
        if previous cost and abs(previous cost - current cost) <=</pre>
stopping threshold:
        previous cost = current cost
        costs.append(current cost)
        weights.append(current weight)
        weight derivative = -(2/n) * sum(x * (y - y predicted))
        bias derivative = -(2/n) * sum(y - y predicted)
        current weight -= (learning rate * weight derivative)
        current bias -= (learning rate * bias derivative)
        if i % 100 == 0:
            print(f"Iteration {i+1}: Cost {current cost}, Weight
{current weight}, Bias {current bias}", end=", ")
    plt.figure(figsize=(8, 6))
    plt.plot(weights, costs)
    plt.scatter(weights, costs, marker='o', color='red')
    plt.title("Cost vs Weights")
    plt.ylabel("Cost")
    plt.xlabel("Weight")
    plt.show()
    return current weight, current bias
def main():
```

```
X = np.array([32.50234527, 53.42680403, 61.53035803, 47.47563963,
59.81320787, 55.14218841, 52.21179669, 39.29956669, 48.10504169,
52.55001444, 45.41973014, 54.35163488, 44.1640495, 58.16847072,
56.72720806, 48.95588857, 44.68719623, 60.29732685, 45.61864377,
38.81681754])
    Y = np.array([31.70700585, 68.77759598, 62.5623823, 71.54663223,
87.23092513, 78.21151827, 79.64197305, 59.17148932, 75.3312423,
81.43619216, 60.72360244, 82.89250373, 97.37989686, 48.84715332,
56.87721319])
    estimated weight, estimated bias = gradient descent(X, Y,
iterations=2000)
    print(f"Estimated Weight: {estimated weight}, Estimated Bias:
{estimated bias}")
    Y pred = estimated weight * X + estimated bias
    plt.figure(figsize=(8, 6))
    plt.plot([min(X), max(X)], [min(Y pred), max(Y pred)],
color='blue', markersize=10, linestyle='dashed')
    plt.xlabel("X")
    plt.ylabel("Y")
    plt.show()
    main()
```

output:



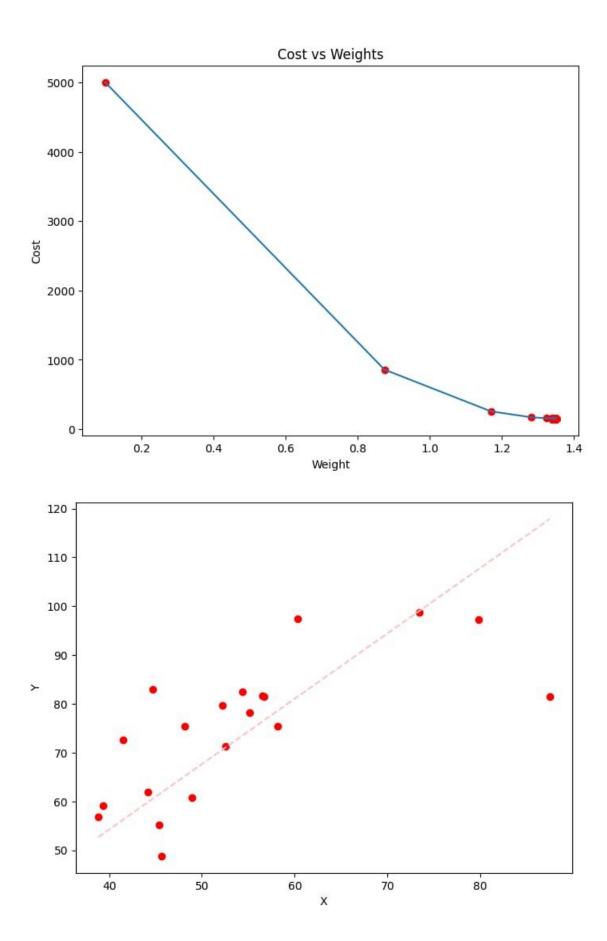


Experiment:7(b)

AIM: To demonstrate gradient descent using python (modified data)

```
import numpy as np
import matplotlib.pyplot as plt
def mean squared error(y true, y predicted):
    cost = np.sum((y true - y predicted)**2) / len(y true)
    return cost
def gradient descent(x, y, iterations=1000, learning rate=0.0001,
stopping threshold=1e-6):
    current weight = 0.1
    current bias = 0.01
    n = float(len(x))
    costs = []
    weights = []
    previous cost = None
    for i in range(iterations):
        y predicted = (current weight * x) + current bias
        current cost = mean squared error(y, y predicted)
        if previous cost and abs(previous cost - current cost) <=</pre>
stopping threshold:
        previous cost = current cost
        costs.append(current cost)
        weights.append(current weight)
        weight derivative = -(2/n) * sum(x * (y - y predicted))
        bias derivative = -(2/n) * sum(y - y predicted)
        current weight -= (learning rate * weight derivative)
        current bias -= (learning rate * bias derivative)
            print(f"Iteration {i+1}: Cost={current cost},
Weight={current weight}, Bias={current bias}")
    plt.figure(figsize=(8, 6))
    plt.plot(weights, costs)
    plt.scatter(weights, costs, marker='o', color='red')
    plt.title("Cost vs Weights")
    plt.ylabel("Cost")
    plt.xlabel("Weight")
    plt.show()
    return current weight, current bias
def main():
    X = \text{np.array}([52.50234527, 63.42680403, 81.53035803, 47.47563963,
89.81320787,
                  55.14218841, 52.21179669, 39.29956669, 48.10504169,
52.55001444,
```

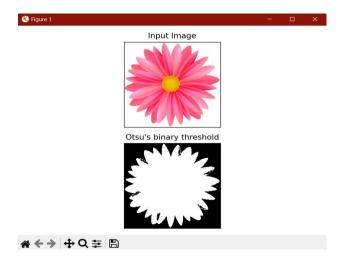
```
56.72720806,
                  48.95588857, 44.68719623, 60.29732685, 45.61864377,
38.81681754])
    Y = np.array([41.70700585, 78.77759598, 82.5623823, 91.54663223,
77.23092513,
71.30087989,
                  55.16567715, 82.47884676, 62.00892325, 75.39287043,
81.43619216,
                  60.72360244, 82.89250373, 97.37989686, 48.84715332,
56.87721319])
    estimated weight, estimated bias = gradient descent(X, Y,
iterations=2000)
    print(f"Estimated Weight: {estimated weight}\nEstimated Bias:
{estimated bias}")
   Y pred = estimated weight * X + estimated bias
   plt.figure(figsize=(8, 6))
   plt.scatter(X, Y, color='orange')
    plt.plot([min(X), max(X)], [min(Y_pred), max(Y_pred)],
color='blue', linestyle='dashed', linewidth=2)
    plt.xlabel("X")
   plt.ylabel("Y")
   plt.title("Linear Regression Line")
   plt.show()
```



EXPERIMENT:8(A)SEGMENTATION

<u>AIM:</u>: Verifying the performance of a image processing by using choosen database with phython code

```
import numpy as np
import cv2
from matplotlib import pyplot as plt
img = cv2.imread(r'C:\Users\Saaniya\Pictures\Screenshots\.4)
flower.jpg')
b,g,r = cv2.split(img)
rgb img = cv2.merge([r,g,b])
gray = cv2.cvtColor(img,cv2.COLOR BGR2GRAY)
ret, thresh =
cv2.threshold(gray,0,255,cv2.THRESH BINARY INV+cv2.THRESH OTSU)
kernel = np.ones((2,2),np.uint8)
closing = cv2.morphologyEx(thresh,cv2.MORPH CLOSE, kernel, iterations =
sure bg = cv2.dilate(closing, kernel, iterations=3)
dist transform = cv2.distanceTransform(sure bg,cv2.DIST L2,3)
ret, sure fq =
cv2.threshold(dist transform,0.1*dist transform.max(),255,0)
sure fg = np.uint8(sure fg)
unknown = cv2.subtract(sure bg, sure fg)
ret, markers = cv2.connectedComponents(sure fg)
markers = markers+1
markers[unknown==255] = 0
markers = cv2.watershed(img,markers)
img[markers == -1] = [255, 0, 0]
plt.subplot(211),plt.imshow(rgb img)
plt.title('Input Image'), plt.xticks([]), plt.yticks([])
plt.subplot(212),plt.imshow(thresh, 'gray')
plt.imsave(r'thresh.png',thresh)
plt.title("Otsu's binary threshold"), plt.xticks([]), plt.yticks([])
plt.tight layout()
plt.show()
```



EXPERIMENT:8(B)

<u>AIM:</u> Verifying the performance of an image processing by using water shed database with python code

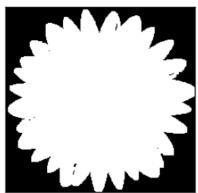
```
import numpy as np
import cv2
from matplotlib import pyplot as plt
img = cv2.imread(r'C:\Users\Saaniya\Pictures\Screenshots\.4)
b, g, r = cv2.split(img)
rgb img = cv2.merge([r, g, b])
gray = cv2.cvtColor(img, cv2.COLOR BGR2GRAY)
ret, thresh = cv2.threshold(gray, 0, 255, cv2.THRESH BINARY INV +
cv2.THRESH OTSU)
kernel = np.ones((2, 2), np.uint8)
closing = cv2.morphologyEx(thresh, cv2.MORPH CLOSE, kernel)
sure bg = cv2.dilate(closing, kernel, iterations=3)
plt.subplot(211), plt.imshow(closing, 'gray')
plt.title("MorphologyEx: Closing: 2x2"), plt.xticks([]), plt.yticks([])
plt.subplot(212), plt.imshow(sure bg, 'gray')
plt.title("Dilation"), plt.xticks([]), plt.yticks([])
plt.imsave(r'dilation.png', sure bg)
plt.tight layout()
plt.show()
```



MorphologyEx: Closing: 2x2



Dilation

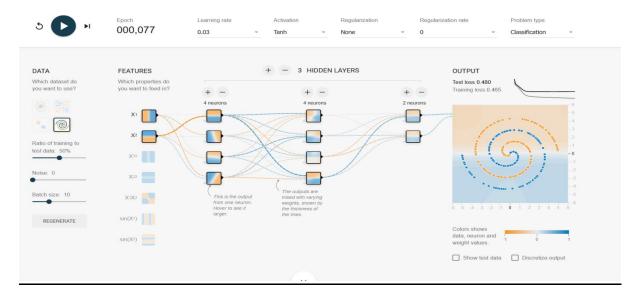




EXPERIMENT:9 (a) TANH

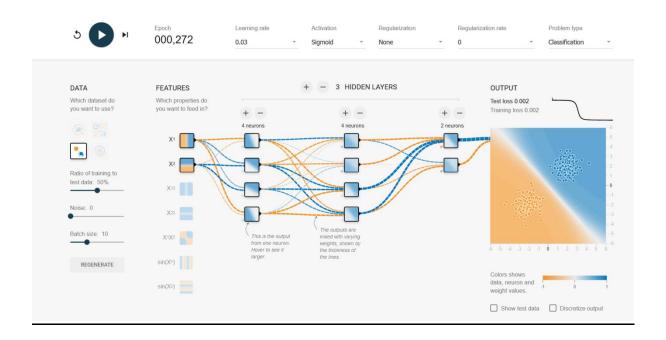
AIM: Neural network analysis using TANH activation

OUTPUT:



EXPERIMENT:9(B) SIGMIOD

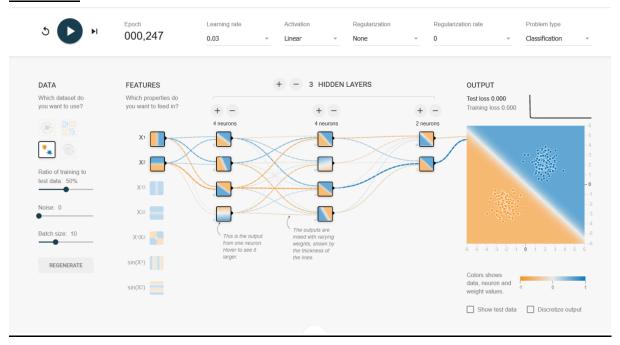
AIM: Neural network analysis using SIGMOID activation



EXPERIMENT:9(C) LINEAR

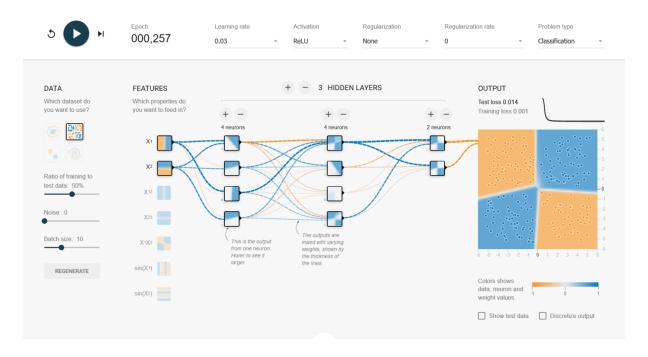
AIM: Neural network analysis using LINEAR activation

OUTPUT:



EXPERIMENT:9(D)RELU

AIM: Neural network analysis using ReLU activation

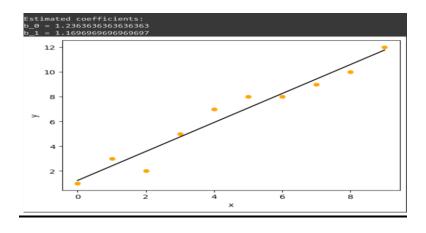


EXPERIMENT:10

<u>AIM:</u> To demonstrate linear separability using python code

PROGRAM:

```
import numpy as np
import matplotlib.pyplot as plt
def estimate coef(x, y):
   n = np.size(x)
   m x = np.mean(x)
    m y = np.mean(y)
    SS xy = np.sum(y*x) - n*m y*m x
    SS xx = np.sum(x*x) - n*m x*m x
def plot regression line(x, y, b):
   plt.scatter(x, y, color = "r",
    y \text{ pred} = b[0] + b[1] *x
    plt.plot(x, y pred, color = "b")
    plt.xlabel('x')
    plt.ylabel('y')
    plt.show()
def main():
    x = np.array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9])
    y = np.array([1, 3, 2, 5, 7, 8, 8, 9, 10, 12])
    b = estimate coef(x, y)
    print("Estimated coefficients:\nb 0 = {} \)
          \nb 1 = \{\}".format(b[0], b[1]))
   plot regression line(x, y, b)
   main()
```



```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import confusion_matrix, accuracy_score
# Load dataset
dataset = pd.read_csv("/Iris.csv")
X = dataset.iloc[:, :-1].values
y = dataset.iloc[:, -1].values
# Split dataset
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20, random_state=42)
# Standardize features
sc = StandardScaler()
X_{train} = sc.fit_{transform}(X_{train})
X_{test} = sc.transform(X_{test})
# Function to train models and evaluate performance
def evaluate_model(model_name, classifier):
  classifier.fit(X_train, y_train)
  y_pred = classifier.predict(X_test)
  cm = confusion_matrix(y_test, y_pred)
  accuracy = accuracy_score(y_test, y_pred)
  print(f"\{model\_name\}\ Confusion\ Matrix:\n\{cm\}")
  print(f"\{model\_name\}\ Accuracy: \{accuracy:.2f\} \backslash n")
# K-Nearest Neighbors (KNN)
from sklearn.neighbors import KNeighborsClassifier
knn_classifier = KNeighborsClassifier(n_neighbors=5, metric='minkowski', p=2)
evaluate_model("KNN", knn_classifier)
# Naive Bayes
from sklearn.naive_bayes import GaussianNB
nb_classifier = GaussianNB()
evaluate_model("Naive Bayes", nb_classifier)
```

Logistic Regression

evaluate_model("Random Forest", rf_classifier)

```
from sklearn.linear_model import LogisticRegression

lr_classifier = LogisticRegression(random_state=42)

evaluate_model("Logistic Regression", lr_classifier)

# Decision Tree

from sklearn.tree import DecisionTreeClassifier

dt_classifier = DecisionTreeClassifier(criterion='entropy', random_state=42)

evaluate_model("Decision Tree", dt_classifier)

# Support Vector Machine (SVM)

from sklearn.svm import SVC

svm_classifier = SVC(kernel='linear', random_state=42)

evaluate_model("SVM", svm_classifier)

# Random Forest

from sklearn.ensemble import RandomForestClassifier

rf_classifier = RandomForestClassifier(n_estimators=10, criterion='entropy', random_state=42)
```