|  |  |  |  |
| --- | --- | --- | --- |
| A picture containing drawing, stop, room  Description automatically generated | Machine Learning  Practical # 3 | | |
|  |  |  |  |
|  |  |  |  |
| **Name** | Ninad Karlekar | **Roll Number** | 22306A1012 |
| **Subject/Course:** | Machine Learning | **Class** | M.Sc. IT – Sem III |
| **Topic** | Concept Learning | **Batch** | 1 |
|  |  |  |  |
| **Topic: Multiclass classification (Problem based Learning)** | | | |
| **AIM: Support vector machine (SVM) algorithm for multiclass classification using Iris.csv and wine dataset from sklearn.**  **Description: Calculate the TP, TN, FP, FN values for the class Setosa using the confusion matrix / contingency table and also calculate precision and recall for data file ‘wine’ from sklearn dataset:**  **TP (True Positives):** The number of correctly predicted positive instances in a binary classification problem.  **TN (True Negatives):** The number of correctly predicted negative instances in a binary classification problem.  **FP (False Positives):** The number of instances that were predicted as positive but are actually negative in a binary classification problem.  **FN (False Negatives):** The number of instances that were predicted as negative but are actually positive in a binary classification problem.  **Support Vector Machine (SVM):** A supervised machine learning algorithm that finds a hyperplane to maximize the margin between different classes in a dataset, making it effective for classification and regression tasks.  **Code and output for Iris.csv:**  import pandas as pd  import numpy as np  import seaborn as sns  import matplotlib.pyplot as plt  from sklearn import svm, datasets  import sklearn.model\_selection as model\_selection  from sklearn.metrics import accuracy\_score  from sklearn.metrics import f1\_score  from sklearn.svm import SVC  from sklearn.metrics import confusion\_matrix  iris = datasets.load\_iris()  #print(iris.data)  X = iris.data[:, :2]  y = iris.target  X\_train, X\_test, y\_train, y\_test = model\_selection.train\_test\_split(X, y, train\_size=0.80, test\_size=0.20, random\_state=101)  X\_train.shape, X\_test.shape, y\_train.shape, y\_test.shape  rbf = svm.SVC(kernel='rbf', gamma=0.5, C=0.1).fit(X\_train, y\_train)  poly = svm.SVC(kernel='poly', degree=3, C=1).fit(X\_train, y\_train)  poly\_pred = poly.predict(X\_test)  rbf\_pred = rbf.predict(X\_test)  poly\_accuracy = accuracy\_score(y\_test, poly\_pred)  poly\_f1 = f1\_score(y\_test, poly\_pred, average='weighted')  print('Accuracy (Polynomial Kernel): ', "%.2f" % (poly\_accuracy\*100))  print('F1 (Polynomial Kernel): ', "%.2f" % (poly\_f1\*100))  rbf\_accuracy = accuracy\_score(y\_test, rbf\_pred)  rbf\_f1 = f1\_score(y\_test, rbf\_pred, average='weighted')  print('Accuracy (RBF Kernel): ', "%.2f" % (rbf\_accuracy\*100))  print('F1 (RBF Kernel): ', "%.2f" % (rbf\_f1\*100))  cm = confusion\_matrix(y\_test,poly\_pred)  cm\_df = pd.DataFrame(cm,  index = ['SETOSA','VERSICOLR','VIRGINICA'],  columns = ['SETOSA','VERSICOLR','VIRGINICA'])  plt.figure(figsize=(5,4))  #print(cm\_df)  sns.heatmap(cm\_df, annot=True)  plt.title('Confusion Matrix')  plt.ylabel('Actual Values')  plt.xlabel('Predicted Values')  plt.show() | | | |
| **Code and output for wine dataset from sklearn:**  import pandas as pd  import numpy as np  import seaborn as sns  import matplotlib.pyplot as plt  from sklearn import svm, datasets  import sklearn.model\_selection as model\_selection  from sklearn.metrics import accuracy\_score  from sklearn.metrics import f1\_score  from sklearn.svm import SVC  from sklearn.metrics import confusion\_matrix  wine = datasets.load\_wine()  X = wine.data[:, :2]  y = wine.target  X\_train, X\_test, y\_train, y\_test = model\_selection.train\_test\_split(X, y, train\_size=0.80, test\_size=0.20, random\_state=101)  rbf = svm.SVC(kernel='rbf', gamma=0.5, C=0.1).fit(X\_train, y\_train)  poly = svm.SVC(kernel='poly', degree=3, C=1).fit(X\_train, y\_train)  poly\_pred = poly.predict(X\_test)  rbf\_pred = rbf.predict(X\_test)  poly\_accuracy = accuracy\_score(y\_test, poly\_pred)  poly\_f1 = f1\_score(y\_test, poly\_pred, average='weighted')  print('Accuracy (Polynomial Kernel): ', "%.2f" % (poly\_accuracy\*100))  print('F1 (Polynomial Kernel): ', "%.2f" % (poly\_f1\*100))  rbf\_accuracy = accuracy\_score(y\_test, rbf\_pred)  rbf\_f1 = f1\_score(y\_test, rbf\_pred, average='weighted')  print('Accuracy (RBF Kernel): ', "%.2f" % (rbf\_accuracy\*100))  print('F1 (RBF Kernel): ', "%.2f" % (rbf\_f1\*100))  cm = confusion\_matrix(y\_test,poly\_pred)  cm\_df = pd.DataFrame(cm)  plt.figure(figsize=(5,4))  #print(cm\_df)  sns.heatmap(cm\_df, annot=True)  plt.title('Confusion Matrix')  plt.ylabel('Actual Values')  plt.xlabel('Predicted Values')  plt.show() | | | |
|  | | | |

**Learnings:**

The provided code performs machine learning classification using two SVM kernels (Polynomial and RBF) on the wine dataset.

The code also generates and visualizes a confusion matrix, which is a valuable tool for evaluating the classification model's performance. The confusion matrix displays the number of true positive (correctly predicted positive class), true negative (correctly predicted negative class), false positive (predicted positive but actual negative), and false negative (predicted negative but actual positive) instances. It provides a clear summary of the model's classification accuracy and potential errors, helping to assess its strengths and weaknesses in differentiating between the wine categories.