HELP MANUAL

DDCN-2019@CSED, MNNIT Allahabad

Classification using Decision Tree Algorithm

Decision Tree based Classification

- **Required packages to implement Decision Tree Classification algorithm**
  - import numpy as np
  - import pandas as pd
  - from sklearn.cross_validation import train_test_split
  - from sklearn.tree import DecisionTreeClassifier
  - from sklearn.metrics import accuracy_score
  - from sklearn import tree

Decision Tree based Classification Contd...

- **Import packages to load the datasets**
  - import pandas as pd
- **Load the dataset "balance.csv"**
  - balance_data=pd.read_csv('C:\\Users\\Dwivedi\\Desktop\\MLA-2017\\balance.csv',header=None,sep=',')
- **Printing Dataset characterstics and five records**
  - print("Dataset Length:: ", len(balance_data))
  - print ("Dataset Shape:: ", balance_data.shape)
  - print ("Dataset:: ", balance_data.head())

Decision Tree based Classification Contd...

- **Slice data set to create feature set X by taking first second to five columns and target set Y as the first column**
  - X = balance_data.values[:, 1:5]
  - Y = balance_data.values[:,0]
- **Spliting feature set and target set both into trainingset and test set**
  - X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size = 0.3, random_state = 100)

Decision Tree based Classification Contd...

- **Create Decision Tree Classifier using gini index**
  - clf_gini = DecisionTreeClassifier(criterion = "gini", random_state = 100, max_depth=3, min_samples_leaf=5)
- **Train the created Decision Tree classifier model on training data set of feature set and target set**
  - clf_gini.fit(X_train, y_train)
- **Make prediction using gini index based Decision Tree Classifier on a random new data [4, 4, 3, 3]**
  - clf_gini.predict([[4, 4, 3, 3]])

Decision Tree based Classification Contd...

- **Make prediction of target using gini index based Decision Tree Classifier for test data set of feature set data**
  - y_pred = clf_gini.predict(X_test)
- **Print Predictions using Gini_index criteria**
  - print ("Predictions using Gini_index“)
  - print (y_pred)
- **Print accuracy of gini index based Decision Tree Classifier for the test data set of target set**
  - print ("Accuracy of Predictions using Gini_index is ", accuracy_score(y_test,y_pred)*100)

Decision Tree based Classification Contd...

- **Create Decision Tree Classifier using entropy**
  - clf_entropy = DecisionTreeClassifier(criterion = "entropy", random_state = 100, max_depth=3, min_samples_leaf=5)
- **Train Decision Tree classifier model on training data set of feature set and target set**
  - clf_entropy.fit(X_train, y_train)
- **Make prediction of target using information gain based Decision Tree Classifier for test data set of feature set data**
- y_pred_en = clf_entropy.predict(X_test)

Decision Tree based Classification Contd...

- **Print Predictions using Entropy Measure**
  - print ("Predictions using Entropy Measure“)
  - print (y_pred_en)
- **Print accuracy of information gain based Decision Tree Classifier for the target set**
  - print ("Accuracy of Predictions using Entropy Measure is ", accuracy_score(y_test,y_pred_en)*100)