[Project Code : BSNB]
Balance Scale using Naive Bayes based Learning Model
Project Duration : 21st Jan 2024 to 10th Feb 2024
Submission Information : (via) CSE-Moodle

Objective:

This data set was generated to model psychological experimental results. Each example is classified as having the balance scale tip to the right, tip to the left, or be balanced. The attributes are the left weight, the left distance, the right weight, and the right distance. The correct way to find the class is the greater of (left-distance x left-weight) and (right-distance x right-weight). If they are equal, it is balanced.

Your task is to build a Naive Bayes classifier learning model for the balance scale task In particular, you shall be doing the following tasks:

1. Based on the dataset (described later), you will write a program to learn a Naive Bayes Classifier. You have to use any method for creation of such a classifier.
2. Compare the results with the results generated by the Naive Bayes classifier learning algorithm from a pre-created package such as sklearn. (Code snippet provided)

Note: *The program can be written in C / C++ / Java / Python programming language from scratch. No machine learning / data science / statistics package / library should be used for model creation.*

Relevant information:

Source: Generated to model psychological experiments reported by Siegler, R. S. (1976). Three Aspects of Cognitive Development. Cognitive Psychology, 8, 481-520.

Dataset Filename: `balance-scale.csv`
Data Description:
Number of Instances: 625 (49 balanced, 288 left, 288 right)
Number of Attributes: 4 (numeric) + class name = 5
Attribute Information:
1. Class Name: 3 (L, B, R)
2. Left-Weight: 5 (1, 2, 3, 4, 5)
3. Left-Distance: 5 (1, 2, 3, 4, 5)
4. Right-Weight: 5 (1, 2, 3, 4, 5)
5. Right-Distance: 5 (1, 2, 3, 4, 5)

Tasks to be done:

1. Starter code provided does this using python: The dataset is not divided into train and validation sets. The first task is to randomly partition the complete dataset into 5 parts:

assign the first part as validation set and the rest for training the classifier. Repeat the process 5 times, assigning the validation sets in a round robin manner. *(5 fold cross-validation)*

2. Naive Bayes Classifier Model:
    a. Implement naive bayes algorithm in your code and mention the same in the report. Do NOT use scikit-learn for this part.
    b. Test the implementation of the Classifier from scikit-learn package (code snippet provided).
3. Classification Report
    a. Create a classification report in tabular form.
    b. You need to calculate precision, recall, f1-score and accuracy of the model.
    c. *Report the average score for the 5 folds.*

Submission Details: (to be submitted under the specified entry in CSE-Moodle)

1. ZIPPED Code Distribution in CSE-Moodle
2. A brief (2-3 page) report/manual of your work
   (with your hyperparameter tuning results also presented in that report)

Submission Guidelines:

1. You may use one of the following languages: C/C++/Java/Python.
2. Your Programs should run on a Linux Environment.
3. You are **not** allowed to use any library apart from these (Also explore all these libraries if doing in Python, or equivalent of these):

```
import numpy # linear algebra
import csv # data processing, CSV file I/O
import pandas # data processing, CSV file I/O
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report
from sklearn.model_selection import KFold
from sklearn.tree import DecisionTreeClassifier # sklearn Decision Tree
import operator
from math import log
from collections import Counter
```

   Your program should be standalone and should **not** use any *special purpose* library for Machine Learning for the decision tree creation algorithm. Numpy and Pandas may be used. And, you can use libraries for other purposes, such as generation and formatting of data.
4. You should submit the program file and README file and **not** the output/input file.
5. You should name your file as <GroupNo_ProjectCode.extension>.
   (e.g., *Group99_BSDT.zip* for code-distribution and *Group99_BSDT.pdf* for report)
6. The submitted program file *should* have the following header comments:
   # Group Number
   # Roll Numbers : Names of members (listed line wise)
   # Project Number

Project Title

7. Submit through CSE-MOODLE only.

You should not use any code available on the Web. Submissions found to be plagiarized or having used ML libraries (except for parts where specifically allowed) will be awarded zero marks.

For any questions about the assignment, contact the following TA:
Sumanta Dey (Email: sumanta.dey@iitkgp.ac.in)