CISC3023 Machine Learning

Course Project

Traffic Sign Classification

In this project, you are given a dataset as following:

Images of Traffic Signs.

Number of Instances (records in your data set): 15540
Number of Classes (types of Traffic Signs): 15

Relevant Information:

This dataset contains 15540 images for 15 traffic signs, the image for each data is around 32X32 pixels, but not fixed. One image is illustrated below for your reference.

Basically, the images are saved in ppm files in different folders. Each folder represents one class of traffic signs. In each folder, there is a csv file that includes the filenames in the folder, the width and height of each image, the Region of interest (x1,y1,x2,y2 illustrated in the image below) of each image, and the class ID or the output label.

For the access of ppm file and csv file, the short demo code attached in UMMoodle provides some simple solutions.
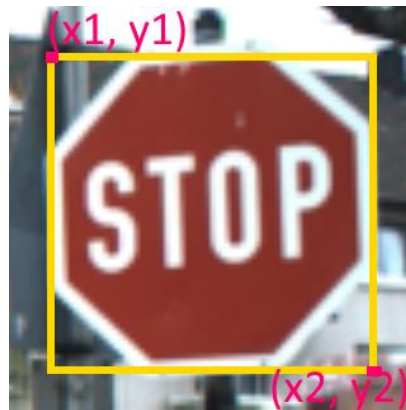
Project requirement:

In this project, the student should design **3 different classifiers (KNN is not permitted)** to classify the data into 15 classes:

- Implement a python program (**trainer**) which can be used to import the historical data from ppm files, preprocess data and train 3 classifiers. Then save the three classifiers.
- Provide some preprocessing on the loaded data to generate some good inputs for the classifier
- Implement another python program (**tester**) which can be used to import the testing data from a folder, and test the trained classifiers over the ppm images in the folder
- Please use cross validation to choose the best classifier.

- *Input:*
  Your programs should be able to accept (read) historical data files (ppm files and csv files) for **trainer** and testing data files (ppm files and csv files) for **tester**
- *Output:*
  Your **trainer** program should output some performance indexes and the trained classifiers. You **tester** program should output classification result for testing data.

What to Submit at UMMoodle before the deadline:

1.   Well-commented python source code for training and testing.

2.   Saved three classifiers, and **trainer** program and the **tester** program that can used them to test data in a folder.

3.   Project report must contain the detail explanation on the design of:

   (1) Instruction to use your code
   (2) Details of the three classifiers used
   (3) The handling of multi-class problem
   (4) How do you preprocess your data to get a better result?
   (5) How do you set the parameters of different classifiers?
   (6) How do you split the data into training and validation sets? (How about cross-validation?)
   (7) What is the performance of the three classifiers?
   (8) What are the labels that are hard to be classified?
   (9) How do you handle the testing data?

4.   You should include a conclusion section for your own additional comments and discussions,

5.   **Your tester should be able to handle ppm images with corresponding csv files, and at the same time, if the csv file is not available, the tester should find a decent way to classify the ppm file as below.**

6.   **Detailed discussion on the comparison of different classifiers and the selection of the best classifier should be included**.

7.   Submit above all your files in a zip file. Provide detailed **instructions** to review and run your codes and program.

IMPORTANT NOTE

1.   Each student may be asked for the demonstration.
2.   Plagiarism will not be tolerated.
3.   Late submission will not be accepted. It is your responsibility to submit via UMMoodle Assignment Function before the deadline. Do not submit it at the last minutes. If the assignment icon is closed after the deadline and if you haven't submitted for any reason, then I will not be able to help you. I will not accept any email submission.