PE10: Programming Exercise

Instructions

Description

The objective of this assignment is to learn about k-nearest neighbors. Given two data sets, complete the functions in the separate knnWrapperClass class within knn_wrapper.py, and edit the parameter on the train method call from the main method to achieve 100% accuracy. The knnWrapperClass object is to implement a k-nearest classifier wrapper and classify test data. Note that knn_wrapper_main.py with the main method and starter file knn_wrapper.py with class with data has already been provided (download attachment). In addition, you should use the pandas and KNeighborsClassifier library from sklearn, and you need to achieve a 100% classification rate on test data.

As part of the assignment, provide which 'neighbor' parameter works best for each dataset and describe tips on choosing the 'neighbor' value. Keep in mind to always comment and document your class and methods.

Documentation reference

Mertz, J. (n.d.). Documenting Python Code: A Complete Guide. https://realpython.com/documenting-python-code/

Additional explanations

1. Files are in pair of "data" and "target," where their lows are synced, and each data low is labeled with classes at the target file.

   - **Train files**:
     - [iris_data.csv, iris_target.csv],
     - [wine_data.csv, wine_target.csv]
   - **Test files**:
     - [iris_test_data.csv, iris_test_target.csv],
     - [wine_test_data.csv, wine_test_target.csv]

2. Result should achieve five 0s, five 1s, five 2s:

```
[0 0 0 0 0 1 1 1 1 1 2 2 2 2 2]
[0 0 0 0 0 1 1 1 1 1 2 2 2 2 2]
```

Expected result

1. knn_wrapper_main.py
   This is the Main python file that is already provided and contains the main and test procedure, which calls methods implemented on knn_wrapper.py.
2. knn_wrapper.py
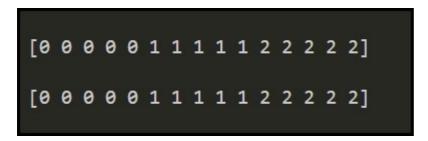   This class contains such methods as init, train, predict, import_train_file, import_test_file.

- init(): this method has already been provided. Please use the 'knn' and 'df_xxx' global field throughout class implementation.
- train(neighbors): the method trains knn classifier using KNeighborsClassifier and fit method from it. Parameters: neighbors value
- predict(): the method classifies df_test data and returns the result
- import_train_file(trainDataFile, trainTargetFile): the method imports train data and target CSV file to df_train and df_train_target global field. Please consider using 'pandas' library and to_list() function for 'target' CSV file
- import_test_file(testDataFile, testTargetFile): the method imports test data and target CSV file to df_test and df_test_target global field. Please consider using 'pandas' library and to_list() function for 'target' CSV file

Example output

KNN

```
[0 0 0 0 0 1 1 1 1 1 2 2 2 2 2]

[0 0 0 0 0 1 1 1 1 1 2 2 2 2 2]
```

Results

As part of the assignment, provide which 'neighbor' parameter works best for each dataset and describe tips on choosing the 'neighbor' value.