Instructions for homework 6

Since this homework will be automatically graded, please be sure to follow the naming and instructions.

Predictor.py

This file contains the predictor class you will be creating. The predictor should do training on the given training set and provide a function to do prediction on an given file.
The given example predictor will naively count and calculate the frequency of spam emails and then do prediction according the frequency.

__init__

This function takes the spam folder and ham folder as input, and will do training on these two folders.
This function will not be called explicitly by the grader, so we can make it a private function here. You can calculate and store anything you want in this function.

__train__

This function performs training process on the given spam and ham folder.
In the example, it simply count and calculate the frequency.
This function will not be called explicitly by the grader, so we can make it a private function here. You can calculate and store anything you want in this function.

predict

This function performs predict on given file and return result.
****This function will be called by the grader, so please following the scheme, especially don't change the parameter and return type in this function.****

The input will be a file name. It will be a single text file, not a folder. The content in this file will be the content of an email with the same format as any given training file.
The return type ****must**** be a boolean value. It stands for whether the given file is a spam.

- True - given file is spam
- False - given file is not spam

trainAndSave.py

This is only an example of how to use the training set to train your predictor and how to save the object into a pickle file.
It checks the input folders first, and then perform train on training set.

This file will not be submitted, and will not be used by the grader. You can modify anything inside as you want. The only constrain is to make sure your saved pickle file can be read and used properly.

readAndPredict.py

This is an example of how we are going to read and use the pickle file. It check the input files first, then load pickle and do prediction.
This is exactly the same way we are going to use the pickle file and the predictor you created. The key points are:

* We will use pickle.load() to load your pickle file
* The predictor is not created by using constructor, it is directly loaded by pickle
* The object loaded from the file must have a predict function, which takes a filename as input and returns a boolean value

You may have noticed that we don't need to import Predictor in this file explicitly. But the Predictor.py file must exist to make the loaded predictor work. Simply put the Predictor.py in the same folder with readAndPredict.py, then everything will work.

This file will not be submitted, and will not be used by the grader. The only purpose of this file is to give you a way to check whether the pickle file you create works fine.

What to submit

The files you will submit will be

- predictor.pickle: the pickle file you created to save the predictor
- Predictor.py: the class you created for predictor

All files names, class names and function names are case-sensitive.