MP12

Added by Javari, Amin, last edited by Javari, Amin on Jul 14, 2016

Machine Problem 12 (**optional** - *15* points)

Purpose:

- Get deeper understanding and working experience of Naive Bayes classifier through implementation.

Instructions:

- This is an individual assignment, which means it is OK to discuss with your classmates and TAs regarding the methods, but it is not OK to work together or share code.
- Similar libraries or programs of frequent pattern mining algorithms can be found on-line, but you are prohibited from using these resources directly. This means you can not include external libraries, or modify existing programs, since the purpose of this MP is to help you go through frequent pattern mining step by step.
- You can use Java/C++/Python/Matlab as your programming language. No other languages are allowed.
- Copying source code from other students will give you 0 grade. We will run plagiarism detection software.

Requirements:

- For your answers in the Answer Document, you should include 1) A brief description of the learning model 2) the outputs 2) a brief explanation about your results 3) answers to questions
- Put all your codes in a separate folder with the name NetId_MP12_codes. Do not use sub-folders inside this folder. All of your codes should have been successfully compiled before submission. Do not include files other than the codes you write. Put a single readme.txt file in the code folder to briefly describe the functionalities of your codes and how to run them.
- Your **PDF** submission file should be at the same level as your code folder. Compress these two together into a zip file, and name it MP11.netid.zip. Submit this zip file through Compass2g.

NOTE: This is an optional MP. You should submit your MP12.netid.zip file to MP12 submission link on Compass2G. The deadline for this MP is (8/4/16).

Input:

- In this MP, we use the car evaluation dataset from the UCI Machine Learning Repository. The car.names file enumerates the class categories as well as the non-class attributes. You should use car.data file to create your test and training sets. Each car has 4 classes : unacc, acc, good, vgood and is judged using 6 attributes: buying, maint, doors, persons, lug_boot, safety.

Required Outputs:

(a) [8, L3] (Step 1) Implement Naïve Bayes method following the algorithm described in the slides. Obviously, your program should have two components: training (to build a classifier) and test (to assign labels to your test data using the learned classifier). Randomly split the data into training (70%) and test (%30) sets. Apply your program on the obtained sets. Your program should output 8 numbers: True positive, false negative, false positive and true negative for the both training and test data.

(b) [7, L3] Use Laplace smoothing to estimate parameters (add k counts to every possible observations value). How do you choose K?

Investigate the effect of decreasing the size of the training data for the both versions of your program (with and without Laplace smoothing). For this experiment, your program should output accuracy on the test and training sets. Conduct your experiments based on 5%, 10%, 40%, 70% and 100% of the training data . Write down your observations.

No labels