CS 412HS   Homework#5   Due : April 25, 11:59PM (Wednesday)

1 General Instructions

- This assignment is due at 11:59 PM on the due date. We will be using Compass (http://compass2g.illinois.edu) for collecting the non-programming part of this assignment. Contact TAs if you face technical difficulties in submitting the assignment. We shall NOT accept any late submission!

- The non-programming part of homework MUST be submitted in pdf format. Handwritten answers are not acceptable. Name your pdf file as YourNetid-HW5.pdf

- The programming part of this assignment will be hosted on hackerrank (https://www.hackerrank.com/) as a programming contest. To participate in this contest, please open a hackerrank account with your illinois.edu email id. If your username in hackerrank is different from your net id, let us know by filling out your net id and username in the spreadheet (link provided in Piazza). The contest framework will allow you to verify the correctness of your submission based on a set of sample test cases. We may use additional test cases to grade your submission. Please check the assignment page on course website in a couple of days for accessing the contest.

- It is OK to discuss with your classmates and your TAs regarding the methods, but it is NOT OK to work together or share code. Plagiarism is an academic violation to copy, to include text from other sources, including online sources, without proper citation. To get a better idea of what constitutes plagiarism, consult the CS Honor code (http://cs.illinois.edu/academics/honor-code) on academic integrity violations, including examples, and recommended penalties.There is a zero tolerance policy on academic integrity violations; Any student found to be violating this code will be subject to disciplinary action.

- For each non-programming question, you will NOT get full credit if you only give out a final result. Necessary calculation steps and reasoning are required to be shown.

- Please use Piazza if you have questions about the homework. Also feel free to send TAs emails and come to office hours.

2 Question (2 Points)

An enterprise is trying to build a decision tree to predict the age of it's customer given the customer information. How would you modify the basic (binary) decision tree algorithm to take into consideration the age of each user?

Hint: Assume that age is a numeric value between 0 and 120.

3 Question (2 Points)

Consider the following statement: If each classifier is better than random guess, ensemble of multiple such classifiers will lead to a nontrivial increase of classification accuracy. Do you agree or disagree with this statement? Give reasoning on it.

4 Question (2 Points)

Suppose you are requested to classify micro-array data with 100 tissues and 10000 genes. Which of the following algorithms you would like to choose, and which ones you do not think will work? State your reasons.

1. Decision-tree induction

2. Naive Bayes algorithm

3. SVM

5 Question (4 Points)

The programming part is hosted on hacker-rank.

Url: `https://www.hackerrank.com/cs-412-hw5`