

Department of Computer Science, University of Houston

COSC 4335 - Exercise set 2

Spring 2012, Due April 17 at 11.30am

Staple all sheets together. No unstapled homeworks will be accepted.

Remember the academic honesty policy for the course. All work turned in **must** be your own. Copying solutions from anyone else, *from a solution manual*, or working on problems together is disallowed. Searching/downloading from internet or other text books for solutions is disallowed unless explicitly allowed for a question.

1. Take the last digit of your UH ID mod 8. Suppose the value is i . Then, select the asymmetric measures i from Table 6.12, where Goodman-Kruskal is numbered 0, Mutual Information 1 and so on. Study the three properties for the selected measures and give proofs of the properties.
2. Consider a medical database with three attributes as discussed in class. Discuss in detail the algorithms that you will use to find patterns in this database. Focus on the three attributes of waist size, educational attainment and socioeconomic status of the patient.
3. Consider a text database with 10 authors and 30 articles of varying length authored by each author. The goal is to identify markers that determine the author of an article. What methods will you use for this goal? Explain in detail the method and any data transformations that you will use. What will be the baseline against which you will compare your methods?
4. Section 6.10. Exercise 2 parts (a) through (c).
5. Section 6.10. Exercise 4 part (a).
6. Section 6.10. Exercise 12 (a), (b) parts (iv) to (vi) only.
7. Section 8.7. Exercise 17 parts (a) through (d).