CSE4334/5334 Data Mining
Module I summary

Sharma Chakravarthy

Department of Computer Science and Engineering
University of Texas at Arlington

Test 1 Details

- **On February 25, 2021 (Thursday) during class for both sections**
  - **I usually make the paper for 80 mins; you will have 100 mins (entire class period)**
  - **Online using lockdown browser + respondus**
  - **Make sure you are familiar with both**
  - **You will do the whole test/exam on separate sheets of paper, scan, and upload to Canvas (will be given extra 15 mins for this)**
  - **Clarifications NOT POSSIBLE (write any assumptions clearly as part of the answer sheet)**
  - **I will provide a sheet with formulas; No need to memorize!**
  - **If you are NEW to lockdown browser + respondus, please let me know. I may make a practice test**

2

Test 1 Details

- **February 25, 2021 (Thursday) during class for both sections**
  - **Make sure you put your id or name, page number on all sheets**
  - **Make sure to put question and sub-question numbers clearly**
  - **Make sure you put any assumptions you have made**
  - **Make sure you have a good scanner (preferably pdf and not images)**
  - **Upload ONE zipped file**
  - **DO NOT upload scratch sheets (ONLY final answer)**
  - **Test 1 is for 15% of overall grade**
  - **It will be 60+5 bonus points (you can only score 60 points)**
  - **I will show some good and bad scans and answer sheets!**

3

Material for Test 1 includes

- Data types
- Distance measures
- Cosine similarity
- Decision tree
  - GINI Index
  - Entropy
  - Miscalculation error
  - GAIN ratio
- Naïve Bayes
  - Formula
  - Working out a small problem
- Common sense

4

Test 1 information

- Will be a collection of short answer question, working out examples (small), intuition behind usage etc.
  - True/false with justification
  - Read questions carefully; Don't write long answers; consider points
    - A question on use of distance measure
    - One or two question on decision tree
    - A question on naïve Bayes
    - A question on comparisons
- Point distribution for each sub-topic may vary
- Where numerical problems are difficult (e.g., SVM), short answer questions on your conceptual understanding will be asked.
- Also, data types and similarity concepts are included.

Thank You !!!

For more information visit:
http://itlab.uta.edu