Computer Science 477

Final Examination

and Final Paper

Particulars

- Thursday, June 8
- 1:00 - 3:00 PM
- Haggard Hall 345
- One page of notes, double sided
  - Handwritten or typeset
- Do bring a calculator

Topics

- Be able to define (articulately):
  - Classification
  - Clustering
  - Association Rule Analysis
- Be able to calculate
  - Entropy
  - Information Gain
  - And understand their uses
- Be able to calculate support, confidence and lift of association rules
  - And understand what they mean

Topic

- Be able to generate frequent itemsets of cardinality k from itemset of cardinality $k - 1$

- Be able to calculate from a given database of transactions how many association rules are possible

- Understand the concept of overfitting and be able to tell how you would know that a classification system overfit

Topics

- Understand the adequacy condition and know how it affects various classification algorithms

- Be able to devise a rational, justifiable method of discretizing a numerical attribute.

Evaluation of Final Paper

- Data Mining – techniques for extracting useful information from datasets otherwise beyond human comprehension

  - *Not* Statistics (though we use statistics)

- Final paper evaluation - *not* on the basis of the actual useful information extracted

- Relevant to evaluation:

  - Application of data mining techniques

  - Quality of writing (organization, professional tone, consistent level of usage, *et cetera*)

  - Ingenuity and completeness of the effort

Some Deadlines

- Final paper due no later than Friday, June 9
- Confirm that equations and diagrams are correctly displayed
- Word or PDF

- I wish you the best in the next stage