Exam #2 Review (This is subject to modification. Any changes will be announced in class AT LEAST one week before the exam!!

General information about Exam #2:

- it will be similar to the Midterm in length and types of questions
- it's a mixture of interpretation / conceptual questions and computations
- you will have the entire class period for the exam if you need it

Things you should bring:

- your textbooks
- 2 2-sided pages of your own notes
- Your "journal" of methods/algorithm comparisons (1-2 pages)
- a calculator (there will only be a few calculations on the Final). This may NOT be your phone.

Things you should review:

- demo and homework problems and answers
- (except for items that are not on the Final – see below)
- things we covered in the in-class slides

Areas you should know about/where to find or things you should know how to do:

- various types of plots / visualization techniques – similarities & differences
- the best applications for various distance and similarity measures
- types of correlation measures; examples of algebraic vs. holistic measures
- how to compute mean and variance, and various normalization measures
- how to compute a chi-square statistic
- how to perform equal width vs. equal frequency (equi-depth) partitioning
- how to compute support and confidence of itemsets
- Gini & other decision tree measures/algorithms; rule generation approaches
- "eager" vs. "lazy" learners, and examples of each
- Evaluating accuracy of classifiers: Holdout, cross-validation and bootstrap
    - Improving accuracy of classifiers: bagging, boosting, ensemble
    - clustering vs. association analysis and algorithms, including which algorithms can and cannot detect arbitrarily-shaped clusters (and why)
    - how different clustering algorithms may be applied in different stituations
    - characteristics of lift and related correlation and evaluation measures
    - classification approaches; how to evaluate a classifier's accuracy and quality
- "noise"
    - Comparison of different clustering techniques and their applications
- Outliers, specifically wrt clustering

- Additional material from the DW text, if applicable. Will be reviewed in class.