FINAL EXAM, STUDY GUIDELINE

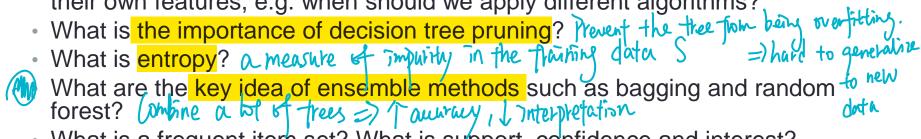
CIS/STA 3920 Dec 14, 2018 3:30~5:30 pm

Exam Structure

- 20% True/False, MC questions
- 80% Short essay questions
- You would not be asked to write R code during the exam
- However, you should be able to read, understand and interpret R outputs, similar to what we have done in labs, practices or assignments
- Short essay questions include applied questions:
 - Given data, software outputs, tables, or plots, use your knowledge to solve problems, or interpret results
- Some calculation formulas you should know
 - Euclidian Distance
 - Bayes Rule
 - Entropy and Information Gain

Sample Questions could be:

 Understand different tree, clustering and association rule algorithms and their own features, e.g. when should we apply different algorithms?



- What is a frequent item set? What is support, confidence and interest?
- Me How to evaluate the classification model performance? Gini Index > Enthopy
- Calculate the probability and make a recommendation using Naive Bayes algorithm P(A|B) = P(B|A) P(A) / P(B)
- Calculate entropy, information gain and construct a decision tree using ID3 algorithm $E_n(s) = \sum_{i=1}^{n} \Pr[\sigma_i, r_i] = F_n(s) \sum_{i=1}^{n} F_n(s) = \sum_{i=1}^{n} F_n(s)$
- How can bagging be used to make a prediction?
 - Use Apriori or FP-tree algorithm to find frequent item sets.
 - Perform clustering using k-means and hierarchical clustering algorithm.
 - How to use information retrieval to find the most similar documents to a given query?