

CSCI 6350/7350 test2 review
(you may bring to the test 2 separate pages (8x11) of notes)

1. Ensemble
 - a. Why ensemble can provide better classification than the underlying base classifier(s)?
 - b. Understand the bagging ensemble approach
 - c. Understand the boosting ensemble approach
 - i. Be able to update the distribution of data based on the performance from “weaklearners”
 - ii. What are the characteristics of set of data end up having higher weights in later iterations?
 - d. Difference between simple majority vote and weighted majority vote
2. Neural Network
 - a. Understand the structure of a feed forward neural network in classification task (what are the input nodes, the hidden layer nodes, and the output layer nodes correspond to)
 - b. Understand the backpropagation training algorithm
 - i. What does the forward phase do? What does the backward phase do?
 - ii. What does BP minimize through weight updation?
 - iii. What does it mean when a FFNN is overfitted? How to prevent overfitting?
 - iv. How to select NN structure (number of hidden layer nodes)?
 - v. Be able to update the weights of the network using BP algorithm
3. Feature Selection and Construction
 - a. What is the purpose of performing feature selection?
 - b. In classification tasks, what is the main underlying approach to various feature selection schemes?
 - c. Understand how mutual information, t-test, chi-squared test, Pearson correlation, signal-noise ratio can be used for feature selection in classification tasks.
 - d. Be able to apply the Feature Selection Mutual Information algorithm and understand how context information in data is used for feature selection
 - e. Be able to apply the Relief algorithm and understand how context information in data is used for feature selection
 - f. Understand the idea of Wrapper methods for feature selection. What is the role of validation data in this method?
 - g. In clustering tasks, what is the main underlying approach to feature selection methods?
 - h. Be able to apply Principle Component Analysis to construct features for clustering problems.
 - i. Be able to compute the eigenvectors and eigenvalues from a given matrix of size 2 or 3.
4. Association Rule
 - a. Be able to compute the support and confidence of an association rule based on data
 - b. Be able to apply the Apriori algorithm in deriving the frequent item sets and the strong association rules
 - c. Understand how the idea of “Apriori property” is used:
 - i. in the frequent item set generation, and
 - ii. in the strong association rule generation
5. Web link analysis
 - a. What are hubs and authoritative nodes?
 - b. How to compute the rank of a page?
 - c. How to compute the overall rank of a page?
 - d. Understand the HITS algorithm, and the Page Rank algorithm. What’s the difference?