CS 5342/4342

The final is comprehensive – covering everything we discussed in class, no exceptions.

Please read the announcements on Canvas (and email copies) about closed book / cheat sheets / date / time, etc.

Please review the assignments, projects, and the midterm.

Solutions to the end-of-chapter exercises from the textbook can be found in the authors' website (but be careful of typos or even wrong solutions!).

Practice Questions

1. Tan (1st ed.) Chap 6: End-of-chapter exercises 2, 3, 6, 7
2. Tan (1st ed.) Chap 8: End-of-chapter exercises 5, 6
3. Tan (1st ed.) Chap 4: End-of-chapter exercises 2, 3, 5, 6, 7, 8
4. Tan (1st ed.) Chap 5: End-of-chapter exercises 6, 7, 8, 10, 11, 12, 13
5. What is the need for "standardization" or "normalization" in data analysis? (Chap 2)
6. Bias-variance decomposition (see class notes for an analytical treatment; Section 5.6.3 for an informal discussion)
7. Computational complexity of the major algorithms discussed
8. Between Ward's method and the centroid method for hierarchical clustering, which one can be considered the hierarchical counterpart of the (partitional) K-means?
9. How can we extend the Naïve Bayesian method to classify text documents where each document is represented by the frequencies of a certain predetermined set of keywords?
10. What is the effect of the value of k in k-NN classification?
11. Consider the usual supervised classification set-up. If two different training data sets are worked upon (separately) by the same learning algorithm (e.g., a decision tree algorithm), are the resulting solutions (classifiers) likely to be the same or different?
12. What is/are the basic condition(s) that must be met for the ensemble method to work?
13. In the context of random forests what is "subspace sampling"? How is it expected to contribute to the ensemble's success?
14. In supervised learning, is a more complex (= more powerful = more expressive) model necessarily better than a less complex one?
15. Briefly discuss the pros and cons of pre-pruning vis-à-vis post-pruning.