

Data Mining 2009

Examination 8. 2. 2010

You may answer either in english or in finnish. Answering to the bonus question may yield you some extra points.

1. Give two (sufficiently different) examples of applying data mining to real life situations.
2. Describe what are *decision trees* and how they are constructed from training data. Show how the decision tree can be used to classify a new example. Discuss possible difficulties and their remedies.
3. Explain in sufficient detail the following validation methods:
 - cross-validation;
 - leave-one-out validation; and
 - bootstrapping.

What are they used for? What are the reasons for selecting a particular validation method?

4. What are *bagging* and *boosting*? Explain these techniques and discuss their common properties, differences and (possible) performance gains.

Bonus question. What is *noise*? Where does it come from? How can data mining methods work in noisy conditions (or can they)?