Project A Mathematical Underpinnings

Feature selection

Methods

Compare the following feature selection methods:

- two[1] (*three*[2]) methods using information theory measures (your own implementation incorporating discretization of continuous predictors)
  - a criterion (*two criteria*[2]) which weights relevancy and redundancy such as JMI or SECMI (other possible measures are listed e.g. in Brown et al),
  - a mini-max criterion as proposed e.g. by Fleuret or Shishikn et al.
- two arbitrary methods, which may include wrapper methods like Lasso regularization, decision tree-based feature importance or wrapper methods like recursive feature elimination.

Using more methods than specified is highly encouraged, although not obligatory. Devise stopping rule of your own and check its performance.

Evaluation of the methods

At least two artificial examples (at least one from each bullet point) and three (*five*[2]) real-world data examples must be provided:

1. an artificial example, for which the set of significant features is known (comparing feature selection methods and checking the performance of a stopping rule)
2. an artificial example in which one of the MI-based methods doesn't work
3. examples using e.g. dataset from UCI. The effectiveness of the feature selection methods can be evaluated by fitting a generic classification model and measuring its performance (see Brown et al).

[1] One-person groups
[2] Two-person groups