Classification Workflow with Rapidminer
Outline

1. Data Import
2. Preprocessing
3. Classification
4. Evaluation
Data Import

- Import your data into Rapidminer Repository
  - Everything in one place
  - Valuable meta-data for further processing

- Use the import wizard, if available
Preprocessing

- Look at your data
  - What is the target attribute?
    - Is the target attribute already a label?
  - What is the distribution of labeled examples by class?
    - Is my classifier capable of handling imbalanced data?
  - What other attributes are available?
    - Is my classifier able to handle these types of attribute?
  - What are the ranges of the attributes?
    - Is my classifier good in handling various ranges?
  - What attributes correlate?
    - Is my classifier able to handle strongly correlating attributes?

See Exercise 1 for more information.
Set Roles & Normalization

- Set roles for attributes
  - Image showing a workflow with a 'Retrieve data' node, a 'Set Role' node, and an 'Edit List' for additional roles.

- Normalize attribute values
  - Image showing a workflow with a 'Retrieve data', 'Set Role', and 'Normalize' nodes, with options for attribute filter type, range transformation, min, and max values.
Discretize

- Numerical attributes can be divided into bins using discretization
- By Size (equally sized data ranges per bin)
  
- By Frequency (equally sized number of examples per bin)
Balancing

- Sampling (with balancing)
- Multiplication of data
  - Filter under-represented class examples
  - Append them to original example set
Classification

- Input: data set with labels
- Output: classification model

Known Classifiers:
- K-NN
- Naive Bayes
- Decision Tree (Hunts & ID3)
- Rule Induction & Tree to Rules
- Support Vector Machine (libSVM)
- Neural Networks
Evaluation

- Evaluate on dedicated test data set
- Evaluate on one data set using
  - Split validation
  - X-Validation
Performance

• Standard Measures
  • Accuracy
  • Precision
  • Recall

• Task Specific
  • Misclassification Costs
Questions?