Association rules and
decision trees

Sven Laur
What is machine learning?
Four basic tasks

Classification

Regression

Summarisation

Clustering
Main inference procedure

1. Gather data
2. Choose method
3. Tune parameters
4. Estimate performance

If something fails, go back to Choose method.
Features are most important
Do not learn things you know

- Fitting complex systems takes a lot of data
- If some parts of them are known use this info!
- You may gain both in precision and stability
Rule-based prediction

Features

\[ X_1, X_2, X_k \]

Target

\[ Y \]

Inferred rules:

- \( X_1 = 1 \) and \( X_2 = 0 \) \( \rightarrow \) \( Y = 1 \)
- \( X_3 = 1 \) and \( X_2 = 1 \) \( \rightarrow \) \( Y = 0 \)
Support and confidence

All examples

If part holds

Then part holds
Why rules do not work?

Good rule

Reachable rules
Decision trees
How to find decision trees

Recursive splitting based on four questions

- Which features are used for splitting?
- What is the split criterion?
- When to stop splitting?
- What is the decision on the leaf nodes
Which attribute?

- Relative frequencies
- Entropy
- Information gain
- Optimal split for continuous variables
Pruning

‣ Early stopping with statistical tests
‣ Early stopping with holdout data
‣ Bottom-up pruning strategy
‣ Rules and rule generalisation
‣ Rule prioritisation
Why does it work

- Representative sample (iid sampling)
- Law of large numbers
- Hypothesis space is complete
- Entropy decreases by separating red and blue dots
- Given enough samples, leaf decision is optimal
Tricks and hacks

- Splits with many parts are bad.
- Add a penalty term
- Sometimes attributes are missing
  - Impute attributes, estimate entropy
- Not all attributes are equal
  - Include costs into the model
- Sometimes decision borders are not rectangular
  - Use advanced classifiers