



Data Mining with Weka

Summary

Ian H. Witten

Summary

- ❖ There's no magic in data mining
 - *Instead, a huge array of alternative techniques*
- ❖ There's no single universal "best method"
 - *It's an experimental science!*
 - *What works best on your problem?*
- ❖ Weka makes it easy
 - *... maybe too easy?*
- ❖ There are many pitfalls
 - *You need to understand what you're doing!*
- ❖ Focus on evaluation ... and significance
 - *Algorithms differ in performance – but is it significant?*

Summary

What have we missed?

- ❖ Filtered classifiers

Filter training data but not test data – during cross-validation

- ❖ Cost-sensitive evaluation and classification

Evaluate and minimize cost, not error rate

- ❖ Attribute selection

Select a subset of attributes to use when learning

- ❖ Clustering

Learn something even when there's no class value

- ❖ Association rules

Find associations between attributes, when no "class" is specified

- ❖ Text classification

Handling textual data as words, characters, n-grams

- ❖ Weka Experimenter

Calculating means and standard deviations automatically ... + more

Summary

❖ Data

- *recorded facts*

❖ Information

- *patterns, or expectations, that underlie them*

❖ Knowledge

- *the accumulation of your set of expectations*

❖ Wisdom

- *the value attached to knowledge*