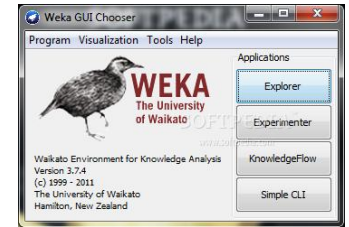


Weka



Weka



- A Java-based machine learning tool
- <http://www.cs.waikato.ac.nz/ml/weka/>
- Implements numerous classifiers and other ML algorithms
- Uses a common data representation format, making comparisons easy
- Comprehensive set of data pre-processing tools and evaluation methods
- 3 modes of operation: GUI, Command Line, Java API

.arff data format

@relation heart-disease-simplified

numeric attribute

@attribute age numeric

@attribute sex { female, male}

nominal attribute

@attribute chest_pain_type { typ_angina, asympt, non_anginal, atyp_angina}

@attribute cholesterol numeric

@attribute exercise_induced_angina { no, yes}

@attribute class { present, not_present}

@data

63,male,typ_angina,233,no,not_present

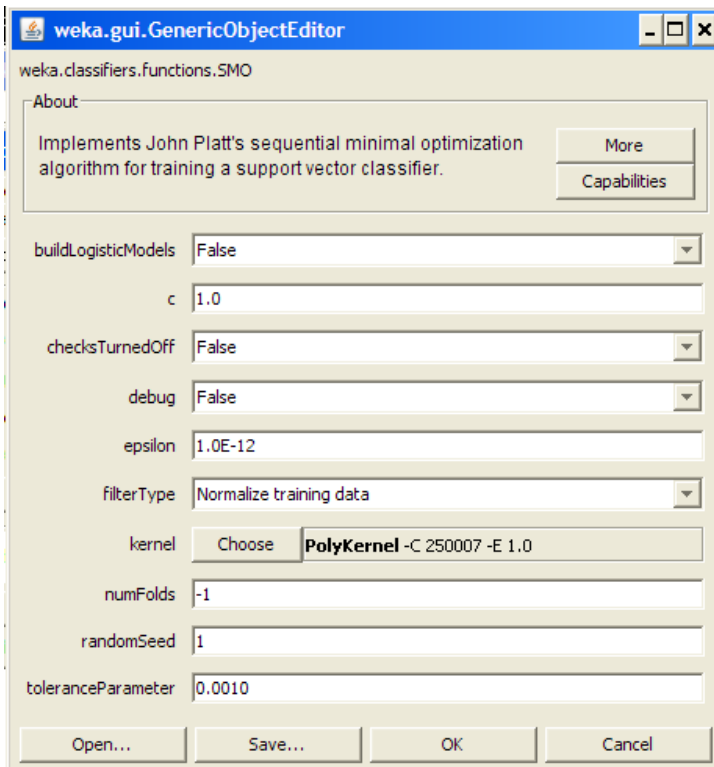
67,male,asympt,286,yes,present

67,male,asympt,229,yes,present

38,female,non_anginal,?,no,not_present

...

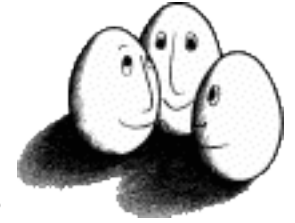
Using SVM in Weka



- SMO is the implementation of SVM used in Weka
- Note that all nominal attributes are converted into sets of binary attributes
- You can choose either the RBF kernel or the polynomial kernel
- In either case, you have the linear versus non-linear options

Weka demo

Weka vs. svm_light vs. ...



- Weka is good for experimenting with different ML algorithms
- Other, more specific tools are much more efficient and scalable
- For SVMs, for example, many use svm_light
- <http://svmlight.joachims.org/>
- Works well for 10K+ features, 100K+ training vectors
- Uses a sparse vector representation
 - Good for many features (e.g., text)