Advanced Data Mining with Weka

Pre-course survey

Class 1 — Time series forecasting
1.1 Introduction: installing Weka packages
1.2 Time series: linear regression with lags
1.3 Using the timeseriesForecasting package
1.4 Looking at forecasts
1.5 Lag creation, and overlay data
1.6 Application: Analyzing infrared data from soil samples

Class 2 — Data stream mining in Weka and MOA
2.1 Incremental classifiers in Weka
2.2 Weka’s MOA package
2.3 The MOA interface
2.4 MOA classifiers and streams
2.5 Classifying tweets
2.6 Application to bioinformatics: Signal peptide prediction

Mid-course assessment

Class 3 — Interfacing to R and other data mining packages
3.1 LibSVM and LibLINEAR
3.2 Setting up R with Weka
3.3 Using R to plot data
3.4 Using R to run a classifier
3.5 Using R to preprocess data
3.6 Application: Functional MRI Neuroimaging data

Class 4 — Distributed processing with Apache SPARK
4.1 What is distributed Weka?
4.2 Installing distributed Weka for Spark
4.3 Using Naïve Bayes and JRip
4.4 Map tasks and Reduce tasks
4.5 Miscellaneous distributed Weka capabilities
4.6 Application: Image classification

Class 5 — Scripting Weka in Python
5.1 Invoking Python from Weka
5.2 Building models
5.3 Visualization
5.4 Invoking Weka from Python
5.5 A challenge, and some Groovy
5.6 Course summary

Post-course assessment