

Introduction to Machine Learning

Joe Thurbon¹

¹Intersect, Sydney, Australia, joe@intersect.org.au

GENERAL INFORMATION

This is a half-day workshop comprising a series of hands-on exercises strung together with a slide-based presentation of background material. The primary convenor of the workshop is Dr Joe Thurbon, who has a research background in machine learning in both academic and commercial research settings. There are no limits on the number of attendees.

Attendees will leave with two key skills – how to cast their research problem as a machine learning problem, and how to determine an appropriate evaluation technique that is relevant to their research goals. They will also be exposed to and have a chance to experiment with a collection of machine learning algorithms.

This workshop was run at the 2011 eResearch Australasia conference, and received positive feedback from a modest but diverse range of participants. The workshop is now a part of Intersect's Learning and Development program, and this year's workshop content is refined based on feedback from participants.

DESCRIPTION

Machine-learning is the automated, data driven discovery of programs or hypotheses that can perform tasks that are otherwise difficult to program. It includes sub-fields for classification, planning and parameter estimation, and has close ties to robotics, optimisation, medical diagnosis and expert systems. The field is one that is still rapidly evolving, but is well past the point where some of the key techniques are now available in a conveniently packaged form, both in terms of APIs as well as interactive workbenches.

This workshop will introduce some of the fundamental concepts of classification machine learning, data representation and how to design experiments that evaluate machine learning algorithms. Attendees will use one of the standard ML workbenches (WEKA) <http://www.cs.waikato.ac.nz/ml/weka/> for experiments.

OUTLINE

The workshop will cover the following:

1. **Classification and Classifiers.** *90 minutes.* This talk and exercises introduces the fundamentals of classification machine learning. By the end of this topic, attendees will understand
 - what a machine learning problem looks like,
 - some of the available arsenal of algorithms (nearest neighbour, SVMs, decision trees, information gain) and
 - the rudiments of using WEKA to run an algorithm on a problem
2. **Evaluation.** *90 minutes.* This talk and exercises introduces the main concepts in evaluation, including
 - Fundamental measurements of performance (accuracy, sensitivity, specificity, roc curves)
 - Choosing the right measure
 - Working out a test regime (n-fold, leave-one-out, folding)
3. **Advanced Topics.** *30 minutes.* This talk presents some of the next steps that interested attendees could follow up on. They include
 - Multiclass extensions to 2-class learning algorithms (e.g. 1-vs-1 and 1-vs-rest)
 - The challenges of data representation and management
 - Parallel and distributed machine learning

WHO SHOULD ATTEND

This workshop is aimed at people who would like to take a first step into machine learning, either to apply it to their research or to add it to their arsenal of eResearch skills. It assumes no programming background, and only a basic level of mathematical ability (and I won't ask you to do any maths). If logarithms scare you, maybe consider coming with a friend.

WHAT TO BRING

Attendees will need to bring their own laptop. I will provide links to simple-to-install software on the day. It will expedite proceedings if you have Java installed before we commence.

ABOUT THE PRESENTER

Dr Joe Thurbon

Joe Thurbon is the Member Services Manager at Intersect, as well as the eResearch Analyst at Southern Cross University. He has a research background in logic and diagrammatic reasoning, and has practiced software engineering for almost 20 years. For the eight years prior to joining Intersect, Joe worked at CISRA, Canon's Australian R&D company, researching and developing machine learning approaches to image processing problems.

Joe has a BSc (Hons) from the University of Sydney in computer science and psychology, and a PhD in computer science from the University of New South Wales.