COMP90051 Statistical Machine Learning

Semester 2, 2015

Subject Description

This subject COMP90051 Statistical Machine Learning is designed to give postgraduate students an understanding and appreciation of different learning approaches, how they relate to other techniques, understand their rationale and how they are applied to solve problems. It is presented as a combination of theory and hands-on practice. The list of topics covered is available in the LMS (`http://app.lms.unimelb.edu.au/` and handbook).

Pre-requisites

The official pre-requisite is the completion of either COMP30018 Knowledge Technologies or COMP90049 Knowledge Technologies. However, if you believe you have the necessary background to take this course, have a look at the following details and please talk to Ben if you have queries.

We assume you have studied or have familiarity with the following topics:

- Basic probability — for example marginalisation, Bayes rule;
- Naive Bayes
- Decision trees;
- kNN;
- Basic clustering (e.g., k-means).

In addition, all learning approaches involve some form of maths and/or algorithms, so a pre-requisite is familiarity with formal mathematical notation. We will be running an optional session during the first week: please attend if you need to brush up on background material.

There are no restrictions on which programming language you use to implement and experiment on, but the workshops will use a combination of R, Matlab, and there will be a need to submit code for projects. Your lecturers will be able to read code in the following languages: C, C#, Java, Python, R, Matlab. More details about this will be given when the specifications for project 1 are released. Be aware that the projects often involve a language more performant than R or Matlab, so proficiency beyond these is *highly* advised.

Lecturers

Name	Weeks	Email	Room
Dr Benjamin Rubinstein	1–5	benjamin.rubinstein@unimelb.edu.au	DMD 7.21
Dr Andrey Kan	6–12	andrey.kan@unimelb.edu.au	WEHI

Office Hours will be held DMD 7.02 Wed noon weekly throughout teaching semester.

Ben is the overall subject coordinator and will be available throughout the semester should you have administrative questions.

Lectures and Workshops

Each week, there are two lectures and one (repeated) workshop. We will make the lecture slides available on LMS before the lecture. Lectures are recorded—the link to recordings can be found in LMS—however you are strongly encouraged to attend lectures. **Use recordings only to revise material, NOT instead of attending lecture.** In addition, references from text books or paper readings may be made available from time to time. To succeed in this subject, we encourage you to study the supplied references as well as attending the lectures and workshops, though unless stated otherwise the readings are optional.

Seeking Assistance

The LMS should be your first port of call when seeking help. Issues that affect multiple students are likely to have been picked out and answered via the "Announcements" page. More specific questions can be posted via the LMS Discussion Board for us or other students to answer. We encourage you to read the discussion board and answer the questions of your fellow students, as you are likely to benefit from reading and answering them.

In addition, please feel free to approach us. Email requests and short questions after lectures will get you help quickly. After that we will hold a weekly office hour—see LMS "Staff Information" for time and place through semester (it will change depending on lecturer). Finally, if the above fail email us to request appointments.

Assessment

There are two assignments, a mid-semester test, and a final written exam. The exam is worth 50% of your overall mark. The remaining 50% will be the maximum of

- The sum of your projects counting 25% each; or
- The sum of your projects counting 20% each plus your mid-semester result counting 10%

In this way the mid-semester is an "optional bonus" but can make up for grades lost in projects, and is an extremely useful mechanism for providing you with *feedback* during the semester.

To pass the subject, you must obtain a) at least 50% overall when all marks are combined; b) [hurdle1] obtain 50% of the total marks possible for the two projects combined; and c) [hurdle2] obtain 50% of the total marks possible for the final exam.

Texts and References

Required:

- Hastie, Tibshirini, Friedman (2001), The Elements of Statistical Learning: Data Mining, Inference and Prediction. (free at `http://www-stat.stanford.edu/~tibs/ElemStatLearn`).

Recommended:

- Russell, Norvig (1995), Artificial Intelligence: A Modern Approach (available in library; reading references will be made to the 2nd edition).

Other references for additional reading or deeper understanding:

- Bishop (2007), Pattern Recognition and Machine Learning.
- Koller, Friedman (2009), Probabilistic Graphical Models: Principles and Techniques.
- Shawe-Taylor, Cristianini (2004), Kernel Methods for Pattern Analysis.

Welcome to Statistical Machine Learning, 2015!

Benjamin Rubinstein and Andrey Kan,
Department of Computing and Information Systems,
July 2015.