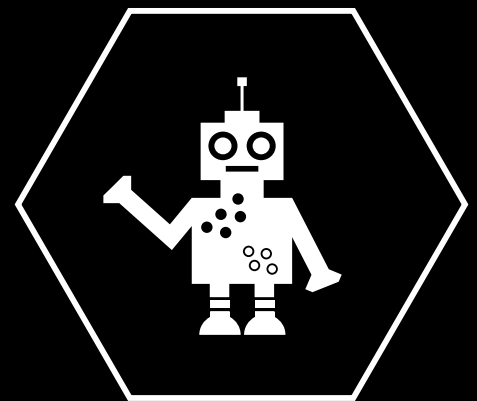


Data Mining and Machine Learning I: Supervised and Un-supervised Learning

Course information sheet 2022-23

Full course, 10 weeks



This course introduces students to machine learning methods and modern data mining techniques, with an emphasis on practical issues and applications.

Prerequisite Knowledge

Learners should have basic experience with the R programming language (e.g. data management and plotting).

Intended Learning Outcomes

By the end of this course learners will be able to:

- apply and interpret methods of dimension reduction such as principal component analysis and the biplot;
- apply and interpret classical methods for cluster analysis;
- apply and interpret a wide range of methods for classification;
- explain and interpret ROC curves and performance measures such as AUC.
- fit support vector machines to data;
- Assess predictive ability objectively.

Syllabus

Week 1 (sample material)

- Dimension reduction in data
- Principal Component Analysis (PCA)
- Performing PCA in R and interpreting its output

Week 2

- Interpreting bivisualisation plot
- Principal Component regression

Week 3

- Classification
- Overfitting
- K-nearest neighbours

Week 4

- Tree based modelling, bagging and random forests
- Applying tree based modelling, bagging and random forests in R

Week 5

- Support vector machines (SVMs)
- Implementing linear SVMs in R
- Kernelised SVMs

Mid-term week break

Week 6

- Peer assessment

Week 7

- Introduction to Model-Based Classification
- Linear Discriminant Analysis and Fisher's Discriminant Analysis

Week 8

- Quadratic and Mixture Model Discriminant Analysis
- Generative vs. Discriminative Classification Models

Week 9

- Cluster analysis
- Reading dendograms
- Choosing the number of clusters

Week 10

- Partitioning cluster analysis
- K-means clustering
- Performing k-means clustering in R and interpreting its output

"The content is very interesting. The different ways of examination provided an excellent challenge."

Online Learning

- Weekly live sessions with tutor(s)
- Weekly learning material (reading material, videos, exercises with model answers)
- Bookable one-to-one sessions with tutor(s)

Textbooks

Hastie, T & Tibshirani, R & Friedman, J (2009) Elements of statistical learning
Smola, A & Vishwanathan, S.V.N (2008) Introduction to machine learning

Assessment

(for credit only)

This will typically be made up of 4 pieces of assessment, including an online quiz, an individual project, an oral assessment and a peer assessment.

Software

To take our courses please use an up-to-date version of a standard browser (such as Google Chrome, Firefox, Safari, Internet Explorer or Microsoft Edge) and a PDF reader (such as Acrobat Reader). Learning material will be distributed through Moodle. We encourage all learners to install R and RStudio and we provide detailed installation instructions, but learners can also use free cloud-based services (RStudio Cloud). Learners need to install Zoom for participating in video conferencing sessions. We recommend the use of a head set for video conferencing sessions.

DATA ANALYTICS 
GLASGOW

School of Mathematics and Statistics

University of Glasgow

<http://gla.ac.uk/>

mdatagov

<http://gla.ai>

Email:

maths-stats-analyticscpd@glasgow.ac.uk