Department of Computer Science

()

COM3004 Data Driven Computing

Summary	This module is intended to serve as an introduction to machine learning and pattern processing, but with a clear emphasis on applications. The module is themed around the notion of data as a resource; how it is acquired, prepared for analysis and finally how we can learn from it. The module will employ a practical Python-based approach to try and help students develop an intuitive grasp of the sophisticated mathematical ideas that underpin this challenging but fascinating subject.
Session	Autumn 2017/18
Credits	20
Assessment	Assignment (50%) and examination (50%).
Lecturer(s)	Dr Jon Barker
Resources	• MOLE (https://vle.shef.ac.uk) () • Unconfirmed practical marks when available (https://www.dcs.shef.ac.uk/intranet/teaching/campus/practicalmarks/com3004.pdf) • Exam Papers, past 2 years (where applicable) (https://www.dcs.shef.ac.uk/intranet/teaching/campus/pastpapers.html)
Aims	This unit aims to: • provide an accessible introduction to key concepts in machine learning and pattern processing, • demonstrate the application of machine learning in a number of recent research areas, • develop an appreciation of the difficulties involved when trying to extract meaning from naturally occurring data with particular reference to data preprocessing, feature extraction, classifier design and efficient learning, • To prepare students for specialised data-driven subjects at level 3/4 such as natural language processing, speech processing and computational biology.
Objectives	By the end of the unit, a student will be able to • demonstrate how to extract features from data for use by machine learning (ML) techniques, • demonstrate the ability to analyze and model data using ML techniques, • demonstrate the ability to apply ML in various areas of Computer Science, e.g. in natural language processing, audio/speech processing, biological applications and vision processing, • demonstrate the ability to use Python for scientific computing.
	Introduction • overview: classification and feature handling • Python programming Multivariate data • review: linear algebra/probability • normal distribution Classification • Bayes decision theory • risk and ROC (receiver operating characteristic) • parameter estimation - maximum likelihood estimation • curse of dimensionality and naive Bayes classifier

Content	Linear classifiers • perceptron • XOR problem Instance based approaches • nearest neighbour and k-nearest neighbour • template matching and edit distance Feature selection • discriminability • feature selection algorithms Feature generation • dimensionality reduction • principle components analysis Unsupervised learning and approaches to clustering. Density estimation and mixture modelling. Case study: Analysis of how techniques have been applied in a real system.
Teaching Method	Lectures, problem classes and laboratory classes.
Feedback	Immediately from problem classes. After each assignment stage through debriefing lecture and individual marking.
Recommended Reading	• Python Programming, https://en.wikibooks.org/wiki/Python_Programming • Stephen Marsland, "Machine Learning: An Algorithmic Perspective", CRC Press, 2009