Introduction to Data Science

Contents

About this course

0.0.1 Course Content

This course will introduce participants to a fascinating field of statistics. We will see how we can rely on statistical models to gain a deep understanding from data. This often involves finding optimal predictions and classifications. Machine Learning (also known as Statistical Learning) is quickly developing and is being applied in various fields such as business analytics, political science, sociology, and elsewhere.

Machine learning can be divided into supervised learning and unsupervised learning. We cover supervised machine learning. Supervised learning involves models where we have a depended variable - often referred to as labelled data. In unsupervised learning the outcome variable is not known - often referred to as unlabelled data.

0.0.2 Course Objectives

This course aims to provide an introduction to the data science approach to the quantitative analysis of data using the methods of statistical learning, an approach blending classical statistical methods with recent advances in computational and machine learning. The course will cover the main analytical methods from this field focussing on hands-on applications using example datasets. This will allow participants to gain experience with and confidence in using the methods we cover.

0.0.3 Course Prerequisites

Participants are expected to have a solid understanding of linear regression models and preferably know binary models. Prior exposure to the statistical software R is required. The course will not provide an introduction to R.

0.0.4 Agenda

1. Regression (linear models)
2. Classification
3. Cross-validation
4. Subset selection
5. Regularisation
6. Polynomials
7. Tree based models
8. Simulation and Monte Carlo Simulation

Placeholder

1 Linear Regression

1.1 Learning objectives

Placeholder

2 Classification

2.1 Seminar

Placeholder

3 Cross-Validation

3.1 Seminar

Placeholder

4 Subset Selection

4.1 Seminar

Placeholder

5 Regularisation

5.1 Seminar

Placholder

6 Polynomials

6.1 Seminar

Placeholder

7 Tree Based Models

7.1 Seminar

Placeholder

8 Simulation and Monte Carlo Simulation

8.1 Seminar

Placeholder