

Modern Statistical Methods (M24)

Dr S. A. Bacallado

The remarkable development of computing power and other technology now allows scientists and businesses to routinely collect datasets of immense size and complexity. Most classical statistical methods were designed for situations with many observations and a few, carefully chosen variables. However, we now often gather data where we have huge numbers of variables, in an attempt to capture as much information as we can about anything which might conceivably have an influence on the phenomenon of interest. This dramatic increase in the number variables makes modern datasets strikingly different, as well-established traditional methods perform either very poorly, or often do not work at all.

Developing methods that are able to extract meaningful information from these large and challenging datasets has recently been an area of intense research in statistics, machine learning and computer science. In this course, we will study some of the methods that have been developed to study such datasets.

Prerequisites

This course requires an undergraduate-level knowledge of Statistics, roughly equivalent to the courses Principles of Statistics and Statistical Modelling from Part II of the Mathematical Tripos.

Literature

1. T. Hastie, R. Tibshirani and J. Friedman. *The Elements of Statistical Learning*. 2nd edition. Springer, 2009.
2. P. Bühlmann and S. van de Geer, *Statistics for High-Dimensional Data*. Springer, 2011.
3. M. J. Wainwright, *High-Dimensional Statistics*, Cambridge Series in Statistical And Probabilistic Mathematics, 2019.

Additional support

Four examples sheets will be provided and four associated examples classes will be given. There will be a one-hour revision class in the Easter Term.