NORBIS

The national research school in bioinformatics, biostatistics and systems biology

NORBIS

Prediction

Time:	January 16-20, 2017
Place:	University of Oslo
Organisers:	Manuela Zucknick, Arnoldo Frigessi, ++
Credits:	5 ECTS
Registration:	by 1st December, at norbis.no

Course description

This course focuses on prediction of future and/or unmeasured outcomes based on a variety of high-dimensional molecular data. What we predict may for example include whether or not a therapy given to a patient is successful (binary or categorical outcome, also called classification), bone mineral density or the expression in so called eQTL studies (continuous outcomes), survival after cancer surgery or time to recurrence of a disease (time to event outcomes). The course is based on various models which exploit molecular data as input data (genomics, metabolomics, proteomic, epigenetic data, for example) in addition to other individual variables (demographic, clinical, exposure data). All of these data are characterised by their huge dimension (e.g. all genes or SNPs, so a large number p of variables), compared to a smaller number (n) of individuals in a study. Variables can be discrete, categorical, continuous and also related to more complex structures like ontologies and pathways (networks).

Learning outcome

After completing the course the student should be able to:

- know what prediction is, in contrast to estimation, testing, and clustering
- know which steps are involved in a prediction task, and which pitfalls need to be avoided
- be able to identify appropriate methods for a given problem, and to perform prediction tasks using R and Bioconductor packages
- be able to assess methods in the literature, and to put these in a wider context
- be able to assess the performance of prediction results, as they are typically reported in publications

Course program

The course will be given as an intensive one-week long course, with lectures in the mornings and practical hands-on sessions in the afternoons. The students will do a project after the course and deliver a written report within a month. The students will be required to give a brief oral presentation during the course and deliver the written report of the take-home project in order to pass the class.

Prerequisites

Basic knowledge in linear algebra and statistics is expected. The practicals will be run using the statistical computing environment R and Bioconductor. We expect students to be familiar with performing data analysis in R/ Bioconductor.