

Term 1: Data Science Fundamentals

Orientation - Introduction to team data science; workshop on recognising and addressing issues of diversity and implicit bias.

Ethics of Health Data Science

What is big data ethics? Using big data: public trust and confidence; sharing big data: equitable sharing ethical access; an ethical introduction to data protection (GDPR) and Healthcare and AI.

Introduction to Computational Statistics

A quick review of probability, statistical models, and maximum likelihood estimation is followed by key statistical tools, including hypothesis testing, power and linear regression.

Modern Statistical Methods

Essential statistical tools used in biomedical research, including: logistic regression, log-linear models, mixed effects models, including models for longitudinal and repeated measures data, penalised regression (ridge & lasso), survival analysis, time series, missing data and multiple testing.

Bayesian Modelling

Fundamentals of Bayesian modelling (Likelihood, Prior, Posterior; conjugacy), Bayesian estimation via posterior sampling vs. optimisation, Monte Carlo, rejection sampling, importance sampling, Monte Carlo Markov Chain, Metropolis-Hastings, Gibbs Sampling, Laplace approximation, variational approximations.

Introduction to Machine Learning

Unsupervised learning & clustering; supervised learning, empirical risk minimization & overfitting, generalization, cross-validation, bias-variance trade-off; Bayes classifiers, LDA, QDA, logistic regression; Generative vs. discriminative learning, naïve Bayes, K-NN; evaluating performance, ROC curves; Decision trees.

Machine Learning Topics

Expectation-Maximisation, Hidden Markov Models, applications to genetics; Gibbs sampling; manifold learning.

Deep Learning

Introduction to Neural networks & Convolutional Neural Networks, Model training and Optimisation, Keras/Tensorflow; CNNs for image processing (e.g. Unet), CNNs for signal processing & CNNs for image registration.

Databases

Introduction to data engineering; data structures; data models and constraints (relational and object); data representation (XML, JSON, SQL); transformations.

Data Challenge

Two weeks using at-scale health data to answer a practical question of interest.