

Large Scale Data Analysis in Physics and Astronomy

This course is targeted at graduate students interested in a cogent exposition of the fundamental questions in data analysis, allowing them to compare and contrast between different methods. The material will be illustrated by relevant examples from Physics and Astronomy. Mathematical treatment would be consistent with the training of a Physics and Astronomy graduate student.

Prerequisites: The course is available to all P&A graduate students. The preferred language for coding in this course is Matlab.

Draft Syllabus

1. Probability Distributions
2. Point Estimation and Maximum Likelihood Estimates
3. Significance Tests, Regression and Goodness-of-fit Tests.
4. Bayesian approach and simulation
5. Latent variable models (clustering, mixture models, Hidden Markov Models,)
6. Kernel methods (SVM classification and regression)
7. Neural Networks/Deep Learning.

There will be several homework assignments as well as a data-centric project in the student's chosen sub-discipline.

Suggested main text:

Kevin P. Murphy, Machine Learning: A Probabilistic Perspective