

Practical data science and machine learning (3 credits) 1-2-5070

Lecturer: Dr. Scott K. Hansen

This is a survey course, introducing practical techniques for processing data and discovering patterns with computers. Material is presented at a level suitable for non-specialist graduate students in the earth sciences. The course is self-contained, covering mathematical and programming basics that are needed, but no more than those that are needed—the emphasis is on practical skills rather than abstract theory. Three hours per week.

Course contents

Introducing data science and machine learning; review of linear algebra and probability basics; univariate and multivariate regression; model calibration; regularization techniques; Bayesian methods; support vector machines; neural networks; unsupervised clustering; principal component analysis; kernel density estimation; model selection.

Useful texts

Data Science from Scratch by Joel Grus

The Hundred-Page Machine Learning Book by Andriy Burkov

Structure of final grade

Evaluation will be based on take-home problem sets surveying the course material.