14BD99

A modern crush course in intermediate Statistics and Probability

Overview and Objectives

This course offers a narrative of fundamental ideas in Statistics and Probability. The narrative and the tools developed are indispensable for then being able to follow the intensive courses of the Msc in Data Science, such as Statistical Modelling and Inference and Machine Learning. The focus is on connecting ideas and showcasing the usefulness of results from Mathematics in concrete applications of Statistics.

Course Outline

Overview and elements of axiomatic probability

Models and Data

- densities and random variables
- independence and conditional densities
- Bernoulli and multinomial models for categorical data
- Conditional indpendence and graphical models
- models for continuous data
- transformations of random variables and densities

Expectation

- Working understanding
- Properties
- Inequalities
- KL divergence and entropy
- Incorporating data and conditional expectation
- Generating functions

Multi-Gauss: properties, models, geometry and linear algebra

- Definition
- Linear algebra pt 1: positive-semi-definite matrices and quadratic forms
- Linear transformations
- Linear algebra pt 2: matrix square roots
- multi-Gauss density
- Linear algebra pt 3: determinants

- Linear algebra pt 4: spectral analysis
- Low rank approximations
- Mahalanobis distance and outlier detection
- The two results from multivariate calculus you NEED to know
- Multi-Gauss conditional distributions
- Gaussian graphical models

Classic parametric inference and asymptotic theory

- KL divergence and likelihood
- Fundamental concepts:
  - point estimation, confidence intervals, tests
- Basic limit theory
  - Real analysis 101
  - Borel-Cantelli and monotone convergence
  - Modes of covergence
  - Laws of large numbers and central limit theorems
- Limit theorems for maximum likelihood estimators and a framework for inference

Hypothesis testing

- Simple hypotheses and the Neymann-Pearson lemma
- p-values
- Multiple testing
- Local power and the score statistic
- From tests to confidence intervals: a narrative

Elements of non-parametric statistics

- Empirical CDF and functionals
- The bootsrap
- Bias-variance tradeoff in Statistics and Machine Learning
- Non-parametric tests

Required Activities

+ Class participation

+ Daily homework

14BD99

A modern crush course in intermediate Statistics and Probability

+ Exam

Evaluation

You need a pass grade in an exam that will be given to you at the end each brush up course.

If you fail this exam you need to retake it within 5 days from the first one. We encourage you to hire a tutor to prepare you best for this retake. We can make suggestions for a tutor but we will not cover the cost of tutoring. In order to be able to follow satisfactorily the rest of the program a pass mark in these exams is a must and we need to make sure that you have reached the necessary level before the program starts. The point of the courses and the exams is to make sure we have helped you to reach the necessary level before the real program starts.

Materials

- Structure-wise the course has similarities to what is covered in Chapters 1-10 of Wasserman's All of Statistics and Chapters 1-3 of Wasserman's All of Nonparametric Statistics. Content-wise the course deviates from the book considerably.

- Bishop's Pattern Recognition and Machine Learning Chapters 1,2 and Appendix B are useful read before the brush up to build a high-level narrative (do not try to go into depth)

- Hastie, Tibshirani and Friedman's Elements of Statistical Learning Chapters 1 and 2 are also a useful read before the brush up to build a high-level narrative (do not try to go into depth)