Advanced Statistics

Course Description

Nearly all processes in nature, and most technical and scientific scenarios, are not deterministic but stochastic. Therefore, these processes must be described in terms of probabilities and probability density distributions. After defining and introducing the fundamental concepts of statistics, this course will cover important probability distributions and their prevalence in application scenarios, discuss descriptive techniques to summarize and visualize data effectively, and discuss the Bayesian approach to statistics. Estimating parameters is a key component in optimizing data models, and the course will give a thorough overview of the most important techniques. Hypothesis testing is a crucial aspect in establishing the observation of new effects and determination of the significance of statistical effects. Special focus will be given to the correct interpretation of p-values and the correct procedure for multiple hypothesis tests.

Contents

1. Introduction to Statistics
   1. Random Variables
   2. Kolmogorov Axioms
   3. Probability Distributions
   4. Decomposing Probability Distributions
   5. Expectation Values and Moments
   6. Central Limit Theorem
   7. Sufficient Statistics
   8. Problems of Dimensionality
   9. Component Analysis and Discriminants
2. Important Probability Distributions and their Applications
   1. Binomial Distribution
   2. Gaussian or Normal Distribution
   3. Poisson and Gamma-Poisson Distribution
   4. Weibull Distribution
3. Bayesian Statistics
   1. Bayes’ Rule
   2. Estimating the Prior, Benford’s Law, and Jeffry’s Rule
   3. Conjugate Prior
   4. Bayesian and Frequentist Approaches
4. Descriptive Statistics
   1. Mean, Median, Mode, and Quantiles
   2. Variance, Skewness, and Kurtosis
5. Data Visualization
   1. General Principles of Dataviz/Visual Communication
   2. 1D, 2D Histograms
   3. Box Plot and Violin Plot
   4. Scatter Plot, Scatter Plot Matrix, and Profile Plot
   5. Bar Chart
6. Parameter Estimation
   1. Maximum Likelihood
   2. Ordinary Least Squares
   3. Expectation Maximization (EM)
   4. Lasso and Ridge Regularization
   5. Propagation of Uncertainties
7. Hypothesis Test
   1. Errors of 1st and 2nd Kind
   2. Multiple Hypothesis Tests
   3. p-Values