ITMO University

Methods and Models for Multivariate Data Analysis

Course Workload		Assessment form (examination/ graded test/ ungraded test)
ECTS	Hours	
3	108	Exam

Students will improve their knowledge of probability theory and develop skills in probabilistic modelling and statistical evaluation.

Course structure:

1. Probabilistic models for one-dimensional random variables

1.1. Introduction. Basic concepts and properties of the distribution law, distribution function, distribution density.
1.2. Methods for estimating distribution parameters. Probability interval, confidence interval, tolerance interval.

2. Probabilistic Models for Multivariate Random Variables

2.1. Introduction to the section. Regression analysis. Linear and nonlinear regression.
2.2. Correlation analysis. Canonical correlations. Principal component method. Natural orthogonal functions.
2.3. Factor analysis. Multidimensional scaling. Analysis of variance. Multivariate interval estimates of distribution parameters, regression.

3. Multidimensional data classification methods

3.1. Introduction to the concept of multidimensional data classification. Cluster analysis. Determination of the proximity metric of objects and classes in cluster analysis problems. Methods and algorithms for hierarchical clustering. Iterative methods and algorithms for clustering. Assessment of the clustering quality function. Criteria for the significance of differences in clusters.
3.2. Fisher's linear discriminant analysis. Quadratic discriminant analysis. Dimension reduction methods. Exploratory analysis and data visualization techniques.

4. Probabilistic models of stochastic processes

4.1. The concept of a random function and its relationship with temporal processes and fields. The concept of stationarity in the narrow and wide sense. Ergodic processes. Periodically correlated random processes. Gaussian processes. Markov processes.

4.2. Dynamic system model. Regression models for stochastic processes. Correlation analysis of stochastic processes. Autoregressive model. Wold's model. Rice's model. Moving average autoregressive model. Trend modeling.
4.3. Spectral auto and cross analysis. Fourier transform. Wiener-Khinchin theorem.

5. Theory of extreme values for random variables and processes

5.1. The concept of the exact distribution of extreme values. Limit theorems for extremal distributions. Fisher-Tippett theorem. Block maximum method. Pickands-Balkema-de Haan theorem. POT method.
5.2. Joint distributions of extreme values. Equal probability contouring method.

6. Fundamentals of Probabilistic Forecasting

6.1. Probabilistic predictive models. Data assimilation methods. Optimal filter concept. Kalman filtration. Optimal interpolation method. Assessment of the quality of forecasts. Calibration of models.
6.2. Ensemble forecast. The concept of ensemble forecast using the Monte Carlo method. The concept of a multi-model ensemble forecast based on a linear dynamic system model.