Applied Statistics

(Romain Billot et Aziliz Guezou-Philippe)

This course provides a comprehensive blend of theoretical foundations and practical applications in statistics, essential for understanding and effectively applying machine learning algorithms. It equips students with the necessary statistical background to tackle complex data analysis challenges and lays the groundwork for advanced machine learning techniques.

Key concepts covered:

- Descriptive statistics (Measures of Tendency, Dispersion, Distribution, Correlation and covariance)
- Sampling theory, Estimation, confidence intervals and hypothesis testing (parametric and non parametric tests, p-value)
- Data challenges (quantity, quality, representativeness, statistical biases)
- Advanced Regression analysis (multiple linear regression, logistic regression, regularization techniques)
- Dimensionality Reduction (Principal Component Analysis, Multiple Correspondence Analysis, t-SNE and UMAP)
- Bayesian Statistics (inference, Markov Chain Monte Carlo)
- Resampling methods (Bootstrap, Monte Carlo simulation)
- Experimental Design (Design of Experiments, A/B Testing)

By the end of this course, students will be able to:

- [BC-04] Design and conduct experiments using statistical principles for robust data collection and analysis
- [BC-07] Apply appropriate statistical methods to analyze and interpret complex datasets
- [BC-07] Implement advanced regression and dimensionality reduction techniques for data modeling and visualization

Prerequisites :

- Fundamental concepts of probability theory, including random variables and their distributions
- Basic sampling theory and estimation techniques
- Principles of hypothesis testing and p-values
- Fundamentals of linear regression analysis