Statistical Foundations of Machine Learning

Course Description

Bielefeld University

Alejandro H. Wences

April 2, 2025

This course provides a rigorous exploration of the statistical principles that form the foundation of modern machine learning. Students will engage with both theoretical concepts and practical applications, developing a deep understanding of statistical inference.

To solidify their knowledge, students will engage in a variety of exercises that blend analytical (mathematical) rigor with simulation-based problem-solving. By the end of the course, participants will be equipped to apply statistical methodologies to machine learning challenges and interpret results with confidence.

Unit	**Topics**	**References**
Unit I: Basic Probability (Review).	Univariate distributions.	Sections 1.5, 1.6 Casella and Berger 1990
	Expectations and variance.	Section 2.2 Casella and Berger 1990
	Multivariate: marginal and conditional distributions.	Sections 4.1,4.2,4.4 Casella and Berger 1990
	Independence and correlation.	Section 4.5 Casella and Berger 1990
	Total hours:	5
Unit II: Notions of Convergence.	Almost sure convergence.	Section 5.5 Casella and Berger 1990
	Convergence in distribution.	
	Strong Law of Large Numbers.	
	Central Limit Theorem.	
	Total hours:	5
Unit III: Big picture: Classical Statistics.	Statistical Models: population, sample, inference.	Kass 2011
	Experiments and the Formal Maximum Likelihood Principle.	Section 6.3 Casella and Berger 1990
	Maximum Likelihood Estimator (MLE). Invariance principle of MLE.	Section 7.2.2 Casella and Berger 1990
	Total hours:	2
Unit IV: Big picture: Bayesian Statistics	Prior and posterior distributions.	Chapter III Robert 2007
	Inference from posterior. Maximum a posterior (MAP) estimator.	Section 7.2.3 Casella and Berger 1990
	Total hours:	2
Unit IV: Point Estimation.	MLE and MAP.	Sections 7.2.2, 7.2.3 Casella and Berger 1990
	Evaluation (Classical): Loss and risk functions.	Section 7.3 Casella and Berger 1990
	Mean squared error and Cramér-Rao bound.	Theorem 7.3.9 Casella and Berger 1990
	Link with Machine Learning Paradigm.	Chapter V Bontempi 2025
	Asymptotic MLE.	Theorems 10.1.6 and 10.1.12 Casella and Berger 1990
	Evaluation (Bayesian): Expected posterior loss.	Section 7.3.4 Casella and Berger 1990
	Total hours:	6
Unit V: Hypothesis Testing.	Likelihood Ration Test (LRT).	Section 8.2.1 Casella and Berger 1990
	Evaluation: Error types, Power function. P-values.	Section 8.3 Casella and Berger 1990
	Asymptotic distributions of LRT.	Section 10.3.1 Casella and Berger 1990
	Bayesian tests.	Section 8.2.2 Casella and Berger 1990
	Loss function optimality.	Section 8.3.5 Casella and Berger 1990
	Total hours:	10
Overall total hours:		**30**

References

Bontempi, Gianluca (2025). Statistical Foundations of Machine Learning: The Handbook. Brussels, Belgium: Machine Learning Group, Computer Science Department, ULB, Université Libre de Bruxelles.

Casella, George and Roger L. Berger (1990). Statistical inference. The Wadsworth & Brooks/Cole Statistics/Probability Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, pp. xviii+650. ISBN: 0-534-11958-1.

Kass, Robert E. (2011). "Statistical inference: the big picture". In: Statist. Sci. 26.1, pp. 1–9. ISSN: 0883-4237. DOI: `10.1214/10-STS337`. URL: `https://doi.org/10.1214/10-STS337`.

Robert, Christian P. (2007). The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation. 2nd ed. Springer Texts in Statistics. New York, NY: Springer-Verlag New York, pp. XXVI, 602. ISBN: 978-0-387-71599-5. DOI: `10.1007/0-387-71599-1`. URL: `https://doi.org/10.1007/0-387-71599-1`.