Faculty of Economics and Business Administration

COURSE (MODULE) DESCRIPTION

Course title	Code
Big Data Analytics	

Staff	Department
Coordinator: Dr Dmitrij Celov **Other(s):**	Department of Econometric Analysis Faculty of Mathematics and Informatics

Study cycle	Course type
First (Bachelor's)	Optional

Form of implementation	Period of implementation	Language of instruction
Face-to-face	Semester 6	English

Requirements for student	
Prerequisites: Statistical Theory, Econometric Theory and Practice	**Additional requirements (if any):**

Number of ECTS credits	Student's workload	Contact hours	Individual work
5	130	36	94

Purpose of the course and competences developed

The purpose of the course is to acquaint students with the principles of analysis of big data, and to enable them to apply high-dimensional models, using tools from statistical, machine learning and econometrics to solve practical problems.

Learning outcomes	Teaching methods	Assessment methods
1.2 understand the problems related to prediction and inference when dealing with big data and/or high-dimensional models	Lectures and individual work	Written exam (70%) Labs (20%) Practicals (10%)
3.2 know and apply methods of supervised learning	Lectures, labs with R, practical training, and individual work.	
3.2 know and apply methods of unsupervised learning		
3.2 understand the specificity of approximations in high dimensions		
3.2 evaluate the empirical adequacy of models		

Course themes	Contact / Individual work: time and assignments

	Lectures	Tutorials	Seminars	Practical classes	Laboratory work	Practice	Contact hours	Individual work	**Assignments due date**
1. Introduction to high-dimensional methods, big data and big p inference	4				1	1	**6**	**14**	[V], [KMO], [ISLR] Ch. 1-2, [ESL] Ch. 1-2,
2. Supervised learning: cross-validation and penalized estimation, regression trees and random forests, bagging and boosting	4				1	1	**6**	**14**	[ISLR] Ch. 3, 5,7,8.2 [ESL] Ch. 3, 5, 7, 8.7, 10
3. Classification and support vector machines	4				1	1	**6**	**14**	[ISLR] Ch. 4, 9 [ESL] Ch. 4, 12
Midterm exam									
4. Unsupervised learning: density estimation, principal components and factor models, clustering, topic models	4				1	1	**6**	**24**	[ISLR] Ch. 10 [ESL] Ch. 14
5. Asymptotic approximations in high dimensions	4				1	1	**6**	**14**	[CGHST], [CHSa],
6. Inference in high-dimensional models	4				1	1	**6**	**14**	[CHSb]
Final exam									
Total	**24**				**6**	**6**	**36**	**94**	

Assessment strategy	**Weight (%)**	**Time of assessment**	**Criteria**
Practical training	10	Regular	A correct solution of 2 equally valued tasks is required to get the maximum.
Labs	20	End of term	4 equally valuated tasks correctly implemented/solved are required to get the maximum.
Midterm examination	35	Mid-term	10 short questions and a solution of 2 exercises.
Final examination	35	End of term	4 points out of 10 from the final exam is required to pass the course. Given this condition holds, the final mark is obtained as a weighted average from the two components.

Author	Published in	Title	Issue No. or Volume	Publishing house or Internet site
Required reading				
[ISLR] James, G., D. Witten, T. Hastie, and R. Tibshirani	2014	An Introduction to Statistical Learning with Applications in R		Springer: http://www-bcf.usc.edu/~gareth/ISL/index.html
[ESL] Hastie, T., R. Tibshirani, and J. Friedman	2009	The Elements of Statistical Learning: Data Mining, Inference, and Prediction		Springer
[CGHST] Chernozhukov, V., M. Gentzkow, C. Hansen, J. Shapiro, M. Taddy	2013	Econometrics of High-Dimensional Sparse Models		*NBER Lectures and Video Materials*: http://www.nber.org/econometrics_minicourse_2013/
[CHSa] Chernozhukov, V., C. Hansen, and M. Spindler	2015	Post-Selection and Post-Regularization Inference in Linear Models with Many Controls and Instruments	105	American Economic Review
[CHSb] Chernozhukov, V., C. Hansen, and M. Spindler	2015	Valid Post-Selection and Post-Regularization Inference: An Elementary, General Approach	forthcom.	Annual Review of Economics
[HK] Hansen, C. and D. Kozbur	2014	Instrumental Variables Estimation with Many Weak Instruments Using Regularized JIVE	182	Journal Econometrics
[KMO] Kleinberg, J., J. Ludwig, S. Mullainathan, and Z. Obermeyer	2015	Prediction Policy Problems	105	American Economic Review: Papers and Proceedings
[V] Varian, Hal R.	2014	Big data: New tricks for econometrics	28	Journal of Economic Perspectives