Course description

Course title	Course code
Econometrics of Big Data	

Lecturer(s)	Department
Coordinator: Dmitrij Celov **Other lecturers: -**	Department of Econometric Analysis Faculty of Mathematics and Informatics

Cycle	Level of course	Type of course
Second		Optional

Mode of delivery	Semester or period when the course is delivered	Language of instruction
Face-to-face	2	English

Prerequisites and corequisites	
Prerequisites: Multivariate Statistics Parametric and Nonparametric Econometrics	**Corequisites (if any)**

Number of ECTS credits	Student's workload	Contact hours	Individual work hours
5	125 hours	50	75

Course objectives: competences to be developed		
To understand problems related with the analysis of big data and application of high-dimensional models and to get acquainted with the relevant tools coming from statistical learning, machine learning and econometrics.		
Learning Objectives. **Students should be able to**	**Learning methods**	**Assessment methods**
- understand the problems related to prediction and inference when dealing with big data and/or high-dimensional models	Lectures and individual work	Participation activity in seminars and practical training leans on the results of individual work and is evaluated on a regular basis. The mid-term and final examinations are of the written form.
- know and apply methods of supervised learning	Lectures, labs, practical training, and individual work	
- know and apply methods of unsupervised learning		
- understand the specificity of approximations in high dimensions		
- evaluate the empirical adequacy of models		

Course content	Contact hours						Individual work hours and assignments	
	Lectures	Consultations	Seminars	Recitation hours	Labs	Practical training	Total contact hours	Individual work hours	Assignments
1. Introduction to high-dimensional methods, big data and big p inference	6				2	1	**9**		[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	6				2	1	**9**		[ISLR] Ch. 3, 5,7,8.2 [ESL] Ch. 3, 5, 7, 8.7, 10
3. Classification and support vector machines	4				2	1	**7**		[ISLR] Ch. 4, 9 [ESL] Ch. 4, 12
Midterm exam									
4. Unsupervised learning: density estimation, principal components and factor models, clustering, topic models	4				2	1	**7**		[ISLR] Ch. 10 [ESL] Ch. 14
5. Asymptotic approximations in high dimensions	6				2	1	**9**		[CGHST], [CHSa],
6. Inference in high-dimensional models	6				2	1	**9**		[CHSb]

Final exam									
Total	30	2			12	6	50	75	

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	Publication year	Title	Volume and/or publication number	Publication place and publisher
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