Big Data (ECON3389) - Syllabus

Spring 2016 - Version Jan 28th*

Lectures Mondays and Wednesdays 8:30-9:45 at Campion Hall 300

Instructors Prof. Stefan Hoderlein
Vitor Hadad (PhD Candidate)

E-mail Please include "ECON3389" in the subject field
Stefan Hoderlein stefan.hoderlein@bc.edu
Vitor Hadad baisihad@bc.edu

Office Hours

Stefan Hoderlein	Mon, Wed 11:15AM-12:00PM
Vitor Hadad	Tue 1:30PM-3:00PM

Description Large-scale data sets ("big data") have become ubiquitous across many applied areas. The goal of this course is to provide an introduction to methods that allow to deal with this situation. We focus on statistical learning techniques and high-dimensional statistics, and show how they can be applied in economics and business administration. Students will learn how to program statistical methods in R or PYTHON, as well as how and when to use the common libraries in these languages.

*Updates are marked with an asterisk.

Prerequisites For students majoring in Economics, we require Economic Statistics (ECON1151) and Econometric Methods (ECON2228). Students coming from different departments are welcome, but should have similar command of statistical methods. A solid knowledge of differential calculus at the level of MATH1102 (the "preferred" co-requisite for ECON2228) is highly recommended, as well as fundamentals of linear algebra (matrix notation, multiplication, inverses, determinants). Prior knowledge of programming is *not* a prerequisite, but student should be willing to learn it.

During the first week, a graded quiz will be handed out to assess the students' mathematical maturity.

Homework There will be weekly homework assignments that will include programming, mathematical problems and applications on real and simulated data sets.

Textbooks

1. *Main textbook:* James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning. New York: Springer. Available for free at: http://www-bcf.usc.edu/∼gareth/ISL/

2. *Other references:*

   (a) Murphy, K. P. (2012). Machine learning: a probabilistic perspective. MIT press.

   (b) Hastie, T., Tibshirani, R., Friedman, J., & Franklin, J. (2005). The elements of statistical learning: data mining, inference and prediction. The Mathematical Intelligencer, 27(2), 83-85. Available for free at: http://statweb.stanford.edu/∼tibs/ElemStatLearn/

   (c) Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.

   (d) Giraud, C. (2014). Introduction to high-dimensional statistics. CRC Press

   (e) Hastie, T., Tibshirani, R., Wainwright, M. (2015). Statistical Learning with Sparsity: The Lasso and Generalizations. CRC Press.

Course Outline This is the first time this course will be taught, so please be advised that the material may change during the semester.

1. Introduction: statistical models, loss functions, optimization
2. Review of multivariate linear regression
3. Beyond linear regression: nonlinear regression, polynomial regression
4. Learning theory: model selection, bias-variance trade-off, overfitting and underfitting, penalization.
5. Regularization: Ridge regression
6. Sparsity: LASSO
7. Ensemble methods: Random forests, Gradient Tree Boosting
8. Classification: Logistic regression
9. Dimensionality reduction: Nearest-neighbors clustering, Principal component analysis (PCA).

Grading

- Homeworks: 30% (best 6 out of 7)*
- Midterm: 30%
- Final project: 40%

	Jan 20th (Wed)	First day of classes
		Homework 0 (math review) handed out
	Jan 25th (Mon)	Homework 0 (math review) due
	Jan 27th (Wed)	Homework 1 handed out
	Feb 8th (Mon)	Homework 1 due
	Feb 10th (Wed)	Homework 2 handed out
	Feb 22nd (Mon)	Homework 2 due
	Feb 24th (Wed)	Homework 3 handed out
Schedule	Mar 2nd (Wed)	Homework 3 due
		Homework 4 handed out
	Mar 14th (Mon)	Homework 4 due
	Mar 16th (Wed)	Homework 5 handed out
	*Mar 30th (Wed)	Homework 5 due
		Midterm
	Apr 6th (Wed)	Homework 6 handed out
	*Apr 20th (Wed)	Homework 6 due
	May 10th (Tue)	**Final Project due**

Please note that there will not be rescheduled or make-up examinations. Homework assignments will not be accepted past their due dates. You must demonstrate your reasoning and show all calculations to receive full grade.

Academic integrity Boston College values the academic integrity of its students and faculty. It is your responsibility to familiarize yourself with the university's policy on academic integrity: www.bc.edu/integrity. If you have any questions, always consult your professor. Violations of academic integrity will be reported to your class dean and judged by the academic integrity committee in your school. If you are found responsible for violating the policy, penalties may include a failing grade as well as possible probation, suspension, or expulsion, depending on the seriousness and circumstances of the violation.

For more information, please make sure to read:
http://www.bc.edu/offices/stserv/academic/integrity.html