CHAPMAN UNIVERSITY
University Honors Program
One University Drive
Orange, CA 92866

COURSE SYLLABUS

HON 385-01 **Summer Term**
(Is Big Data Enough? A Conceptual Exploration of Data Science)

Instructor: Dr. Domenico Napoletani
E-mail: napoleta@chapman.edu
Office Hours: Wednesday 10am-1pm, or by appointment. Location will be announced in class.

Catalog Description: *Prerequisite: acceptance to the University Honors Program, or consent of instructor.* (Offered as needed.) 3 credits. In this course we will explore the computational, mathematical and philosophical concepts underpinning the use of large collections of data to solve problems. We will ask whether it is possible to preserve a role for our reason, when so much of what we understand and what we decide is ultimately shaped by data-driven algorithms.

Course Learning Outcomes:

By the end of this course, students will have:

Obtained a structured knowledge of philosophical, mathematical and computational concepts relating to data collection and analysis.
Developed the ability to rigorously and creatively analyze and assess the core algorithms and ideas of data science.
Sharpened their research skills by exploring how the reliance on large and unstructured collections of data has an impact on what it means to understand natural phenomena.

Honors Program Learning Outcomes:
Upon completing a course in the University Honors Program students will have:

a. Obtained a starting point for integrative exploration of the development of cultures and intellectual achievements through a variety of disciplinary and interdisciplinary perspectives;
b. Sharpened their ability to critically analyze and synthesize a broad range of knowledge through the study of primary texts and through engagement in active learning with fellow students, faculty, and texts (broadly understood);

c. Understood how to apply more integrative and interdisciplinary forms of understanding in the advancement of knowledge and in addressing complex challenges shaping the world;
d. Developed effective communication skills, specifically in the areas of written and oral exposition and analysis.

Content:

The course will be comprised of two parts: an introduction to the core ideas and methods of data science; and a philosophical analysis of its methodology.

We will introduce the mathematical ideas that are at the basis of data science, i.e. the use of algorithms to solve problems on the basis of intensive data collection. The focus will be on understanding the fundamental ideas on which most data science methods are based, and on being able to decide which specific method is relevant for a given problem.

We will explore the ways models can be fit on data, and the key role of optimization in finding such models. The techniques and ideas we will introduce are exemplary of data science and include: linear regression, logistic regression and support vector machines; regularization techniques and the curse of dimensionality; nearest neighbor methods, hierarchical and k-means clustering; neural networks and deep learning.

However, rather than considering data science as a collection of loosely related techniques, we will take the view that there is a coherent set of methodological and philosophical principles on which data science can be based. We will study the structure of these principles and the impact of data science on our modalities of understanding reality. This philosophical approach will bring to the fore not only the potential of data science, but its limits as well, and will suggest a narrow path for a future science that is still driven by us, and not ultimately shaped only by our machines.

Current Required Texts:

Data science for business: what you need to know about data mining and data-analytic thinking by Foster Provost and Tom Fawcett. O' Reilly Media, 2013.

The following papers will be an integral part of our analysis and the starting point of the philosophical discussion.

D. Napoletani, M. Panza, and D.C. Struppa. 2014. Is big data enough? A reflection on the changing role of mathematics in applications. Notices of the American Mathematical Society 61(5): 485-490. Available at: http://www.ams.org/notices/201405/rnoti-p485.pdf

D. Napoletani, M. Panza, and D.C. Struppa. Forcing optimality and Brandt's principle. Accepted for publication in the forthcoming volume *Mathematics as a Tool* edited by J. Lenhard and M. Carrier. Boston Studies in the Philosophy and History of Science. Springer. Preprints will be made available in class.

Other relevant texts will be on reserve at the library, to give the necessary background to some more specialized lectures and projects, for example:

A. J. Izenman. *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning.* Springer, 2013.

T. Hastie, R. Tibshirami, J. Friedman. *The Elements of Statistical Learning.* Springer, New York, 2001. Also available as a free pdf file at: http://statweb.stanford.edu/~tibs/ElemStatLearn/

The Oxford Handbook of Philosophy of Science. Ed. Paul Humphreys, Oxford University Press, 2016.

P. Humphreys. *Extending Ourselves: Computational Science, Empiricism, and Scientific Method.* Oxford University Press, 2007.

J. M. Epstein. *Generative Social Science: Studies in Agent-Based Computational Modeling.* Princeton University Press, 2006.

Instructional strategies:

The course will include daily, in-depth discussions, based on a variety of readings. Mathematical and computational topics will be introduced by lectures. A period of guided written reflection will be set aside at the end of each class. Moreover, students will write an extensive and rigorously argued analysis of a specific topic agreed with the instructor by the second week of class. They will also present their research in class and moderate the resulting discussion.

Methods of Evaluation:

Assessment of student performance will be based on the following items:

In-class discussions and written reflections (30% of the grade). Each student is expected to be directly involved in the discussion during each meeting. Because of this, daily attendance is required and at most one justified absence is permitted. For each additional absence, 5% of the grade will be deducted from the overall grade. At the end of each class, students will be given time to engage in a written reflection on a problem that puts in relief the mathematical and philosophical arguments discussed on that day.

Midterm evaluation (25% of the grade). There will be an in-class evaluation of conceptual and technical understanding of data analysis methods. The emphasis will be on making sure that students have developed the ability to manipulate and generalize the key ideas presented in the course and to understand how to apply them to realistic scenarios.

Final paper (30% of the grade). Students are expected to submit a final paper of 15 pages (12pt font, 1" margin and double spaced) by the last day of class. By the end of the second week of class, students need to confirm a topic for their paper related to the main themes of the course, chosen among a wide selection of given computational, mathematical and philosophical topics. **Papers should offer a synthesis and an original viewpoint on the chosen topic, with an extensive review of relevant literature and a clear grasp of the core concepts of data science (minimum of 10 scholarly citations).**

Presentation and discussion moderation (15% of the grade). All students will present the research topic of their final paper and moderate the discussion for a total of half an hour, during the last week of class. Students will be assessed based on their grasp of the topic at hand; the creativity used in framing the topic and linking it with other discussions; the ease and effectiveness of their moderation.

Chapman University Academic Integrity Policy:
Chapman University is a community of scholars which emphasizes the mutual responsibility of all members to seek knowledge honestly and in good faith. Students are responsible for doing their own work, and academic dishonesty of any kind will not be tolerated anywhere in the university

Students with Disabilities Policy:
In compliance with ADA guidelines, students who have any condition, either permanent or temporary, that might affect their ability to perform in this class are encouraged to inform the instructor at the beginning of the term. The University, through the Center for Academic Success, will work with the appropriate faculty member who is asked to provide the accommodations for a student in determining what accommodations are suitable based on the documentation and the individual student needs. The granting of any accommodation will not be retroactive and cannot jeopardize the academic standards or integrity of the course.

Last revised: 04/03/2017