

WHAT MAKES DATA SCIENCE DIFFERENT? A DISCUSSION INVOLVING STATISTICS2.0 AND COMPUTATIONAL SCIENCES

Christophe Ley

Ghent University, Department of Applied Mathematics, Computer Science and Statistics, christophe.ley@ugent.be

Abstract. Data Science is today one of the main buzzwords, be it in industrial or academic settings. Machine learning, experimental design, data-driven modelling are all, undoubtedly, rising disciplines if one goes by the soaring number of research papers and patents appearing each year. The prospect of becoming a “Data Scientist” appeals to many. A discussion panel organised as part of the European Data Science Conference 2016 in Luxembourg asked the question: “What makes Data Science different?” In this talk, I give a personal and multi-facetted view on this question, from a statistics, machine learning and engineering perspective. In particular, I compare Data Science to Statistics and discuss the connection between Data Science and Computational Science.

Keywords. High-dimensional statistics, Interdisciplinary research, Machine learning, Data Analytics, Modelling

Résumé. Data Science est aujourd’hui l’un des principaux mots à la mode, que ce soit dans un cadre industriel ou universitaire. L’apprentissage automatique, la conception expérimentale, la modélisation basée sur les données sont sans aucun doute des disciplines en ligne de mire si l’on se fie au nombre croissant d’articles de recherche et de brevets paraissant chaque année. La perspective de devenir un “Data Scientist” est attrayante pour beaucoup de personnes. Un panel de discussion organisé dans le cadre de la European Data Science Conference 2016 au Luxembourg a posé la question suivante: “Qu’est-ce qui distingue Data Science des autres domaines?” Dans cet exposé, je donne un point de vue personnel et multiforme sur cette question, en adoptant des perspectives de type statistique, apprentissage automatique et ingénierie. En particulier, je compare Data Science à la Statistique et discute le lien entre Data Science et Computational Science.

Mots-clés. Statistiques en grande dimension, Recherche Interdisciplinaire, Apprentissage automatique, Data Analytics, Modélisation