



Ciclo de palestras por jovens doutorados - 2ª edição (DEEC-JD2)

Quinta-feira, 16 de maio de 2019, 12h30
Anfiteatro EA2

Challenges in Data Science and Big Data processing: weaving theory with applications in healthcare, public administration, and business

Cláudia Soares

Research Assistant, Instituto de Sistemas e Robótica, Lisboa

ABSTRACT

Recent computational and data accessibility developments have made it possible to transform raw data into useful predictions and detailed descriptions of the world. For those who deal with real-world data, the problems of large amounts of heterogeneous, noisy, and incomplete datasets come hand-in-hand with the benefits of rich, ubiquitous, and up-to-date information. I will present some of the work I have done with my group with such data, interfacing with domain experts from healthcare, public administration, and businesses. These efforts exposed scientific challenges to the data analysis tools available today. Such challenges fueled my research, and I will illustrate how principled theory matched real-world applications. Tools from statistical data analysis, to numerical and optimization techniques, are enablers for the principled research developments that can meet those challenges. Like all scientific areas, data science has its tradeoffs: the expressiveness of probability distributions that can be used to model virtually all questions we pose to data, is balanced out by hardness of inference or intensive data needs. This calls for distributed and approximate inference algorithms, and structure-inducing strategies, while taking into account heterogeneous data beyond the reals, like ordinal or categorical variables, and missing data, adding extra uncertainty to the problem.

Finally, I will sketch the near future landscape of my research, where I envision as central topics privacy and explainability of distributed Big Data models.

SHORT BIO



Cláudia Soares holds a degree on Electrical and Computer Engineering from IST, Universidade de Lisboa, and on Literature, Cultures and Languages, from NOVA University. She completed her PhD in IST on Electrical and Computer Engineering in 2016. She is now a Research Assistant at ISR Lisboa and an invited assistant professor at the Systems, Decision and Control Scientific Area of the Department of Electrical and Computer Engineering of IST. She worked as an entrepreneur copywriter

while pursuing her Engineering degree, and later as an engineer at Chipidea and SISCOG. At ISR Lisboa her research interests include Big Data processing, distributed statistical estimation from data, and learning from space and time correlated data.