15D013

Topics in Big Data Analytics II

3 ECTS

Overview (Matthew Eric Bassett)

Topological Data Analysis leverages mathematical topology to provide qualitative descriptions of high dimensional datasets, especially when the datasets come with neither a natural coordinate system nor a metric, and has found applications in neuroscience, fluid dynamics, genomics, etcetera. This module will introduce the topological concepts useful for dealing with high dimensional datasets: simplicial complexes, functoriality, and persistent homology, and show how these concepts can extract relevant features from their samples. The goal is to provide students with a firm scientific understanding of the methods in this area of research, while also understanding examples where these methods have seen success. This module will be theoretical and will only briefly cover applications or computational strategies.

* This part is primarily focused on the *speculative science* behind data mining from high dimensional datasets. It is an introduction to active research.

* Ideas will be introduced from pure mathematics that will build a framework that allows us to "geometrize" data to find intrinsic properties.

* Though primarily focused on theory, we will touch on applications in neuroscience, genomics and fluid dynamics.

* Finally, we'll develop the relationship between topological data analysis and other geometric methods, like manifold learning.

Overview (Hrvoje Stojic)

Bayesian optimization is a framework for modeling situations where one wants to learn functions in a supervised manner, but also smartly select inputs with the idea of making the maximization more efficient. Related ideas can be found in other fields of machine learning and statistics, such as active learning or optimal experimental design. This framework has been used in modeling autonomous agents and solving certain types of reinforcement learning problems, but its usage recently blossomed due to applications in optimizing (hyper)parameters of models that are costly to evaluate. Use-cases include optimizing the architecture of deep learning networks or optimizing parameters of scheduling and planning optimizers that have a long runtime.

The objective of the course is to provide you with a basic understanding of the theory behind the Bayesian optimization, and equip you with sufficient amount of know-how to immediately incorporate it into your workflow and substitute less efficient grid-search methods. Bayesian optimization is a general framework and we will illustrate other use-cases.

15D013

3 ECTS

Topics in Big Data Analytics II

Course Outline

Introduction to Reinforcement Learning and Gaussian Processes

We will first briefly introduce the reinforcement learning problem and distinguish it from other types of machine learning problems, providing an overview of basic concepts, such as multi-armed bandit problems or exploration-exploitation trade-off. We will spend more time on basics of the Gaussian processes - great supervised learning method which gives you the full Bayesian nonparametric treatment. We will focus on regression problems and you will learn how to deploy Gaussian processes in practice.

Bayesian Optimization

In this section we will examine contextual multi-armed bandit problem in more details and uncertainty-based decision strategies for balancing the exploration-exploitation trade-off. Together with Gaussian processes this will allow us to formulate the Bayesian optimization algorithm and illustrate some of its many potential uses. We will focus on one particular use case - you will learn how to use it for optimizing hyper-parameters.

Materials

Rasmussen, C. E., & Williams, C. K. I. (2006). Gaussian processes for machine learning. MIT Press.

Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., & de Freitas, N. (2016). Taking the Human Out of the Loop: A Review of Bayesian Optimization. Proceedings of the IEEE, 104(1), 148–175. http://doi.org/10.1109/JPROC.2015.2494218

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian Optimization of Machine Learning Algorithms. In Advances in Neural Information Processing Systems (pp. 2951–2959). http://doi.org/2012arXiv1206.2944S

Overview (Robert Gimeno, Pablo Hidalgo, Gaston Besanson)

Internet of Things (IoT) is the next ground for Analytics. IoT is spread across the length and breadth of the industry. From consumer electronics, automobiles, aviation, energy, oil and gas, manufacturing, banking, and so on, almost every industry is benefiting from IoT. Data can come from different devices or sensors, data can have different characteristics, diverse latencies, different importance or be plagued with missing values or be exposed through vulnerabilities in the pipeline security.

15D013

3 ECTS

Topics in Big Data Analytics II

We, data scientists, are used to developing our solutions on cloud computing architectures, where a server takes care of the entire computation. However, with the IoT ecosystem, this is not always efficient neither cost-effective so we need to rethink our paradigms.

Here is where the “edge computing” (the devices/sensor layer) presents itself as a solution. Edge computing is an architecture where the process of data, applications, and services are pushed away from the centralized cloud to the logical extremes of the network. These devices are equipped with the enough computing power and data storage facilities to fulfill the task. After computing, only the rich and condensed yet reusable data is transmitted back to the cloud. Low cost alternative storage can also be considered for the remaining data.

The objective of the course is to provide you with a hands-on experience on IoT Analytics that will leverage your already obtained knowledge throughout the Master and equip you with basic skills to deploy your first IoT Analytics solutions.

Required Activities

Attendance to theory classes and problem sets.