In-Depth Dependency Estimation Analysis

Estimating the dependency of two or more variables is a fundamental task in data analysis. It allows to understand data and to identify the variables required to answer specific questions. Reducing data to its most relevant components also improves the execution of learning algorithms, in terms of both runtime and quality of the outcome. In recent work of ours, we have introduced Monte Carlo Dependency Estimation (MCDE), a theoretical framework to estimate multivariate dependency in large data sets. MCDE quantifies dependency as the average discrepancy between marginal and conditional distributions via Monte Carlo simulations.

While our current experiments show the benefits of our approach, there is still a lot to discover about its properties. In particular, the deployment of MCDE to the streaming setting is most interesting. **As a student research assistant, you will aim at answering the following questions:**

- Our current experiments assume Gaussian noise. How would the method react to different noise model? Would the result look different with uniform or white noise, for example?
- The method requires the computation of an index as a preprocessing step, which is typically expensive to compute. How to compute this index incrementally, e.g., in a stream fashion?
- In practice, MCDE is instantiated using a statistical test. The current implementation employs the well-known Mann-Whitney U test. How would the system behave with different statistical tests?

As a research assistant, you will contribute to open-source research software. **The following abilities are required:**

- Programming skills. Experience with Scala or Java and Git is preferred.
- Communication skills. You should be able to communicate and present your results clearly.
- Scientific skills. You should be able to read and understand current research papers and to implement ideas of your own.

Throughout this work, the student will acquire a deep knowledge of dependency estimation and sharpen his/her Data Science skills. The student will learn how to conduct controlled experiments to compare results of his/her work to the state of art. **Furthermore, as a student research assistant, you will enjoy the following benefits:**

- Access to the chair’s computing infrastructure for research purposes.
- The possibility to write and publish papers with our support, based on your results.
- Scientific discoveries and the opportunity to contribute to the state of the art.
- Flexible working time, from 20 to 40 hours per month.