Computational Statistical Discrepancies for High-dimensional Datasets

Supervisor: Song Liu (song.liu@bristol.ac.uk)

Statistical Discrepancies play important roles in modern-day machine learning: For example, when classifying images of cats and dogs, we want to find image features that maximize a statistical discrepancy between cat and dog images. When generating artificial images, we want synthesized images that minimize a statistical discrepancy between the synthesized and original images.

Although the computation of statistical discrepancies has been extensively explored, traditional measures such as the Kullback-Leibler (KL) divergence fall short of effectively handling high-dimensional and large-scale datasets, which are commonplace in image processing and other data-intensive fields. Recent advancements have introduced computational methods aimed at estimating existing discrepancies, including the variational approximations of f-divergences and the introduction of computationally efficient alternatives like the Fisher-Hyvarinen divergence.

In this project, we will investigate computational methods for accurately estimating statistical discrepancies in high-dimensional datasets. By leveraging the inherent low-dimensional structures present within these datasets, we aim to enhance the sample efficiency of discrepancy estimation. Our approaches will be justified through real-world applications and theoretical investigations. You can find more about Song Liu's research at https://allmodelsarewrong.net/.