Spectral Density Estimation of Kernel Matrices with Applications

Mikhail Lepilov

Abstract

Kernel matrices formed from large sets of observations arise frequently in data science, for example during classification tasks. It is desirable to know the eigenvalue decay properties of these matrices without explicitly forming them, such as when determining if a low-rank approximation is feasible. In this talk, I will introduce a new spectral density framework based on quantile bounds. This framework gives meaningful bounds for all the eigenvalues of a kernel matrix while avoiding the cost of constructing the full matrix. The kernel matrices under consideration come from a kernel with quick decay away from the diagonal applied to uniformly-distributed sets of points in Euclidean space of any dimension. I will prove certain results enabling this framework whenever the kernel function satisfies a certain decay condition, and I will give empirical evidence for its accuracy. In the process, I will also prove a very general interlacing-type theorem for finite sets of numbers. Additionally, I will give an application of this framework to the study of the intrinsic dimension of data. In doing so, I introduce a new "local" notion of intrinsic dimension, which has the power to test a certain interpretation of the so-called "manifold hypothesis" for a given dataset.