Abstract

There is a strong relation between high-dimensional data and functional data. One can convert the densely observed high-dimensional data to functional data by defining a set of functional bases, then set up the coefficients to define the functional data as a linear combination of these bases. Typically, the choice of the bases is not data-driven with a notable exception to the number of dimensions, that often can be derived by cross-validations. As a consequence, several standard bases such as Fourier and related bases, wavelets, splines etc. are typically used to transform observed functional data. Through such a prior and rather arbitrary decision on the basis selection the problem is transformed to a finite-dimensional space of basis coefficients and formally is losing its infinite dimensional character. We propose a strictly data-driven method of basis selection. Since the method is algorithmic and searches the data to find an effective representation of the basis by minimizing overall mean square error across functional samples, the functional basis is strictly tied to the functional character of the data and loses arbitrariness of common approaches. The method itself uses B-splines and O-splines in the machine learning style of functional data mining to find efficiently placed knots. Due to machine learning character of data processing, the method has the potential to further numerically improve and extend beyond the considered scope.

This is a joint work with Krzysztof Podgórski.