

Functional Data Clustering via Hypothesis Testing k-means

Ronaldo Dias¹

¹ UNICAMP

Functional data clustering procedures seek to identify subsets of curves with similar shapes and estimate representative mean curves of each such subset. In this work, we propose a new approach for functional data clustering based on a combination of a hypothesis test of parallelism and the test for equality of means. These tests use all observations, which come from an underlying functional model, to compute a measure that determines to which smoothed cluster center each subject's data belongs. This measure is incorporated into a modified k-means algorithm to partition subjects into clusters and find the cluster centers. While competing algorithms require a fixed amount of smoothing for all curves, the proposed test-based procedure performs unsupervised clustering to curves with different degrees of smoothing. Extensive numerical experiments were examined and the results on simulated and real datasets suggest that the proposed algorithm outperforms other clustering approaches in most cases.