

# Identifying informative distance measures in high-dimensional feature spaces

Alessandro Laio  
(SISSA, Italy)

Real-world data typically contain a large number of features that are often heterogeneous in nature, relevance, and also units of measure. When assessing the similarity between data points, one can build various distance measures using subsets of these features. Finding a small set of features that still retains sufficient information about the dataset is important for the successful application of many statistical learning approaches.

We introduce an approach that can assess the relative information retained when using two different distance measures, and determine if they are equivalent, independent, or if one is more informative than the other. This test can be used to identify the most informative distance measure out of a pool of candidates. We will discuss applications of this approach to different branches of physics and science.