



Contribution ID: 33

Type: not specified

Franck Picard - A probabilistic Graph Coupling View of Dimension Reduction

Tuesday, October 4, 2022 3:30 PM (1 hour)

Dimension reduction is a standard task in machine learning, to reduce the complexity and represent the data at hand. Many (and more than many!) methods have been proposed for this purpose, among which the seminal principal component analysis (PCA), that approximates the data linearly with a reduced number of axes. In recent years, the field has witness the emergence of new non linear methods, like the Stochastic Neighbor Embedding method (SNE) and the Uniform Manifold Approximation and Projection method (UMAP), that proposes very efficient low-dimensional representations of the observations. Though widely used, these approaches lack clear probabilistic foundations to enable a full understanding of their properties and limitations. A common feature of these techniques is to be based on a minimization of a cost between input and latent pairwise similarities, but the generative model is still missing. In this work we introduce a unifying statistical framework based on the coupling of hidden graphs using cross entropy. These graphs induce a Markov random field dependency structure among the observations in both input and latent spaces. We show that existing pairwise similarity dimension reduction methods can be retrieved from our framework with particular choices of priors for the graphs. Moreover this reveals that these methods suffer from a statistical deficiency that explains poor performances in conserving coarse-grain dependencies. Our model is leveraged and extended to address this issue while new links are drawn with Laplacian eigenmaps and PCA.