
Sparse GEMINI for Joint Discriminative Clustering and Feature Selection

Louis Ohl*^{1,2}, Pierre-Alexandre Mattei , Frédéric Precioso , and Charles Bouveyron

¹Laboratoire d'Informatique Signaux et Systèmes de Sophia-Antipolis – Laboratoire d'Informatique, Signaux, et Systèmes de Sophia Antipolis, Laboratoire d'Informatique, Signaux, et Systèmes de Sophia Antipolis, Laboratoire d'Informatique, Signaux, et Systèmes de Sophia Antipolis – France

²Equipe Maasai, Inria Sophia Antipolis – Institut National de Recherche en Informatique et en Automatique – France

Abstract

Feature selection in clustering is a hard task which involves simultaneously the discovery of relevant clusters as well as relevant variables with respect to these clusters. While feature selection algorithms are often model-based through optimised model selection or strong assumptions on $p(x)$, we introduce a discriminative clustering model trying to maximise a geometry-aware generalisation of the mutual information called GEMINI with a simple ℓ_1 penalty : *the Sparse GEMINI. This algorithm avoids the burden of combinatorial feature selection on high dimensional data and large amount of samples while only designing a clustering model $p_\theta(y|x)$. We demonstrate the performance on scaled datasets. Our results show that Sparse GEMINI is a competitive algorithm and has the ability to select relevant subsets of features.*

Keywords: clustering, discriminative, unsupervised learning, feature selection, sparsity

*Speaker