SCF Iteration for Orthogonal Canonical Correlation Analysis

Ren-Cang Li
University of Texas at Arlington

October 29, 2020

Canonical Correlation analysis (CCA) is a standard statistical technique and widely-used feature extraction paradigm for two sets of multidimensional variables. It finds basis vectors for the two sets of variables such that the correlations between the projections of the variables onto these basis vectors are mutually maximized. Mathematically, CCA is an optimization problem that can be turned into a singular value problem. Orthogonal CCA (OCCA) is a term that was coined broadly as a collection of variants of CCA imposing orthogonality among basis vectors. The most simple variant is the plain CCA followed by performing the Gram-Schmidt orthgonalization process on the two sets of basis vectors. A more straightforward way is to impose orthogonality while optimizing the same objective function as CCA. It has been observed that orthonormal bases by the latter are more effective than the most simple way in data science applications. However, directly optimizing the objective function with orthogonality constraints on the basis vectors is nontrivial. In the data science community, today it is solved most by generic optimization methods. In this talk, we will present an alternating numerical scheme whose core is a customized self-consistent-field (SCF) iteration for a maximization problem on the Stiefel manifold. Along the line, an orthogonal multiset CCA (OMCCA) will be discussed. Extensive experiments are conducted to evaluate the proposed algorithms against existing methods including two real world data science applications: multi-label classification and multi-view feature extraction.

This talk is based a recent joint work with Lei-hong Zhang (Soochow University), Li Wang (UT Arlington), and Zhaojun Bai (UC Davis).