Canonical Variate Regression

Chongliang Luo[1], Jin Liu[2], Dipak Dey[1], Kun Chen[1*]

[1]Department of Statistics, University of Connecticut , Storrs, CT, US
[2]Centre for Quantitative Medicine, Duke-NUS Graduate Medical School, Singapore
[*]Corresponding author, kun.chen@uconn.edu

In many fields, multi-view datasets, measuring multiple distinct but interrelated sets of characteristics on the same set of subjects, together with data on certain outcomes or phenotypes, are routinely collected. For example, in cancer study, microscopic organ tissue measurements, genetic profiles, and organ function test results may all be available. The objective in such a problem is often twofold: both to explore the association structures of multiple sets of measurements and to develop a parsimonious model for predicting the future outcomes. We study a unified canonical variate regression framework to tackle the two problems simultaneously, allowing them to flexibly borrow strength from each other and hence reinforce each other. The proposed criterion integrates multiple canonical correlation analysis with predictive modeling, balancing between the association strength of the canonical variates and their joint predictive power on the outcomes. Moreover, the proposed criterion seeks multiple sets of canonical variates simultaneously to enable the examination of their joint effects on the outcomes, and is able to handle multivariate and non-Gaussian outcomes through a general loss function formulation. An efficient algorithm based on variable splitting and Lagrangian multipliers is developed. Simulation studies show the superior performance of the proposed approach compared to existing alternative methods. We demonstrate the effectiveness of the proposed approach in an F_2 intercross mice study and an alcohol dependence study.

Keywords: Canonical Correlation Analysis; Integrative Analysis; Reduced-Rank Regression; Supervised Learning.