Restricted kernel canonical correlation analysis

NINA OTOPAL

Institute of Mathematics, Physics and Mechanics
Ljubljana (Slovenia)
`nina.otopal@gmail.com`

Kernel canonical correlation analysis (KCCA) is a procedure for assessing the relationship between two sets of random variables when the classical method, canonical correlation analysis (CCA), fails because of the nonlinearity of the data. The KCCA method is mostly used in machine learning, especially for information retrieval and text mining. Because the data is often represented with non-negative numbers, we propose to incorporate the non-negativity restriction directly into the KCCA method. Similar restrictions have been studied in relation to the classical CCA and called restricted canonical correlation analysis (RCCA), so that we call the proposed method *restricted kernel canonical correlation analysis* (RKCCA).

With the Karush-Kuhn-Tucker theorem we show that the solution of RKCCA equals an unconstrained solution to a modified CCA problem on two random vectors with known covariance matrix where one or several variables have been excluded. Furthermore we use the idea of sub-vectors and sub-matrices to translate the problem of searching for the kernel canonical correlation under non-negativity restriction into an optimization problem related to eigenvalues of some generalized eigenvalue problem with a real symmetric matrix and a positive definite matrix.