Chapter 15. Canonical Correlation Analysis

* Joint study of wo data sets: ask what type of low-dimensional projection helps to find joint structures
* Standard tool for discovery and quantification of associations between two sets of variables
* Basic technique: define an index (projected multivariate variable) that maximally correlates with the index of the other variable for each sample separately.
* Aim: maximize the association (correlation) between the low-dimensional projections of the two data sets

15.1 Most Interesting Linear Combination

* CCA developed by Hotelling (1935)
* for random variables and find such that is maximized.
* ; covariance is invariant to scale, so we can solve
* under the constraints
* define , then the SVD is (or UΛVT)
* where K has rank k = rank() = rank(), Λ is the diagonal matrix with non-zero eigenvalues (sqrt) of KKT and KTK, Γ are the standardized eigenvectors of KKT, and Δ are the standardized eigenvectors of KTK.
* The *canonical correlation vectors* are:
* The *canonical correlation variables* are:
* The *canonical correlation coefficients* are:
  + for *i* = 1, …, *k*