Canonical correlation analysis

- A method of correlating linear relationships between two multidimensional variable.
  - x = [x1, x2, …, xq1], y = [y1, y2, …, yq2]
  - Transform x = w_x x = [w_xx1, w_xx2, …, $w_x$xq1]
  - Transform y = w_y y = [w_yy1, w_yy2, …, $w_y$yq1]
  - Choose w_x and w_y that maximize the correlation between the transformed vectors

Canonical correlation analysis

- Extension of multiple regression – more than one y
  - x = [x1, x2, ..., xq1], y = [y1, y2, ..., yq2]
  - R11 = cor matrix of variables in x
  - R22 = cor matrix of variables in y
  - R12 = cor between x and y (q1 by q2 matrix)
  - E1 = R11inv R12 R22inv R21
- CC combines the DVs to find pairs of new variables (called canonical variables, one for each data table) which have the highest correlation.
- CV's, even when highly correlated, do not necessarily explain a large portion of the variance of the original tables.
  - This make the interpretation of the CV sometimes difficult, but CC is nonetheless an important theoretical tool