Canonical Correlations

Robert L. Wolpert
Department of Statistical Science
Duke University, Durham, NC, USA

Let X be an n-dimensional random vector with finite covariance, and Y an m-dimensional one. Are X and Y independent? If not, how can one quantify the *degree* of dependence between them? Here's one form of answer, borrowed in large part from the Wikipedia.

1 Mutual Information

One classical approach from information theory is to measure the *mutual information*. If the joint distribution of X and Y has a density function $f(x, y)$ with marginals $f_x(x)$ and $f_y(y)$, this is defined to be

$$I(X:Y) = \iint \lg \left[\frac{f(x,y)}{f_x(x)\, f_y(y)} \right] f(x,y)\, dx\, dy$$

(here $\lg x \equiv \log_2 x = \frac{\log x}{\log 2}$; traditionally logarithms in Information Theory are taken to base two, so $I(X:Y)$ is measured in 'bits'). A similar expression applies to discrete random variables, with a sum replacing the integral— or, more generally, to any joint distribution absolutely-continuous with respect to the product of the two marginal distributions (otherwise $I(X:Y) = \infty$). It's easy to show that $I(X:Y) \geq 0$ for all X, Y and that $I(X:Y) = 0$ if and only if X and Y are independent. The quantity $I(X:Y)$ is invariant under affine changes of variables for X and Y.

For univariate normal random variables $X \sim \mathsf{No}(\mu_x, \sigma_x^2)$ and $Y \sim \mathsf{No}(\mu_y, \sigma_y^2)$ with correlation ρ, hence covariance and precision matrices

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix} \qquad \Lambda = \frac{1}{\sigma_x^2\sigma_y^2(1-\rho^2)} \begin{bmatrix} \sigma_y^2 & -\rho\sigma_x\sigma_y \\ -\rho\sigma_x\sigma_y & \sigma_x^2 \end{bmatrix}$$

we have (setting $\mu_x = \mu_y = 0$ without loss of generality)

$$\begin{aligned}
I(X:Y) &= \iint \lg \left[\frac{\left[2\pi\sigma_x^2\sigma_y^2(1-\rho^2)\right]^{-1/2} \exp\left\{-\frac{\sigma_y^2 x^2 - 2\rho\sigma_x\sigma_y xy + \sigma_x^2 y^2}{2\ \sigma_x^2\sigma_y^2(1-\rho^2)}\right\}}{\left[2\pi\sigma_x^2\right]^{-1/2} \exp(-x^2/2\sigma_x^2)\ \left[2\pi\sigma_y^2\right]^{-1/2} \exp(-y^2/2\sigma_y^2)} \right] f(x,y)\,dx\,dy \\
&= \iint \lg \left[\frac{1}{\sqrt{1-\rho^2}} \exp\left\{ \frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2} - \frac{\sigma_y^2 x^2 - 2\rho\sigma_x\sigma_y xy + \sigma_x^2 y^2}{2\ \sigma_x^2\sigma_y^2(1-\rho^2)} \right\} \right] f(x,y)\,dx\,dy \\
&= -\tfrac{1}{2}\lg(1-\rho^2) + \frac{1}{\log 2}\left\{ \frac{\sigma_x^2}{2\sigma_x^2} + \frac{\sigma_y^2}{2\sigma_y^2} - \frac{\sigma_y^2\sigma_x^2 - 2\rho\sigma_x\sigma_y\,\rho\sigma_x\sigma_y + \sigma_x^2\sigma_y^2}{2\ \sigma_x^2\sigma_y^2(1-\rho^2)} \right\} \\
&= -\log(1-\rho^2)/(2\log 2). \qquad (1)
\end{aligned}$$

As expected, $I(X:Y) = 0$ if $X \perp\!\!\!\perp Y$ (i.e. if $\rho = 0$), while $I(X:Y) \to \infty$ as $\rho \to \pm 1$. We'll return to this below.

There are close links between $I(X:Y)$, the Kullback-Leibler divergence from $p(x,y)$ to $p_x(x)\,p_y(y)$, and the mutual (Shannon, not Fisher) information $H(X,Y)$ between X and Y.

2 Canonical Correlations

For $a \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$, what is the highest possible correlation ρ between $U = a'X$ and $V = b'Y$? First let's standardize by setting $c = \Sigma_{XX}^{\frac{1}{2}} a$ and $d = \Sigma_{YY}^{\frac{1}{2}} b$; then

$$\begin{aligned}
\rho &= \frac{a'\Sigma_{XY} b}{\sqrt{a'\Sigma_{XX} a}\,\sqrt{b'\Sigma_{YY} b}} \\
&= \frac{c'\Sigma_{XX}^{-\frac{1}{2}}\Sigma_{XY}\Sigma_{YY}^{-\frac{1}{2}} d}{\sqrt{c'c}\,\sqrt{d'd}}
\end{aligned}$$

so, by the Cauchy-Schwartz inequality,

$$\rho^2 \le \frac{(c'AA'c)\ (d'd)}{c'c \quad d'd} = \frac{c'AA'c}{c'c}$$

where

$$A = \Sigma_{XX}^{-\frac{1}{2}}\Sigma_{XY}\Sigma_{YY}^{-\frac{1}{2}}.$$

The largest this can possibly be is the maximum eigenvalue λ_1 of the positive-definite matrix AA'; it's attained when c is the (say, unit) eigenvector c_1 for AA' and where $d = d_1$ is proportional to $A'c_1$ and (say) of unit length.

We call ρ_1 the "first canonical correlation" and the vectors

$$U_1 = c_1' \Sigma_{XX}^{-\frac{1}{2}} X = a_1' X \qquad V_1 = d_1' \Sigma_{YY}^{-\frac{1}{2}} Y = b_1' Y$$

the first Canonical Components.

We can define similarly the second, third, etc. canonical correlations and components, up to the rank of AA', $(n \wedge m)$.

These may be used to quantify the degree of dependence between X and Y. For example, if all $\rho_i = 0$ for $i > r$, the dependence has rank (at most) r; a classical test of that null hypothesis may be based on the asymptotic (as the number of replicates $p \to \infty$) χ^2_ν distribution of the test statistic

$$Q_r = -\left(p - 1 - \tfrac{1}{2}(m + n + 1)\right) \sum_{j=r+1}^{n \wedge m} \log\left(1 - \hat{\rho}_j^2\right), \tag{2}$$

with degrees of freedom $\nu = (m - r)(n - r)$. As an exploratory tool, a plot of Q_r *vs.* r may suggest the dimensionality of the linear interdependence of X and Y.

2.1 Back to Information

The sum in Equation (2) is just $(-2 \log 2)$ times the mutual information $I(X_r : Y_r)$ between the projections of X and Y onto the orthogonal complements of the spaces spanned by the first r canonical components (see Equation (1)).