Principal Component Analysis

May 12, 2014

Let x be a zero-mean random variable. Suppose we want the direction w such that the projection of x along this direction has maximum variance:

$$\max_w \operatorname{var}(w'x) \quad \text{st.} \quad \|w\| = 1. \tag{1}$$

We have

$$\begin{aligned} \operatorname{var}(w'x) &= \mathrm{E}\, w'xx'w \qquad (2) \\ &= w'\Sigma w. \qquad (3) \end{aligned}$$

The Lagrangian is

$$L = w'\Sigma w + \lambda(w'w - 1). \tag{4}$$

The stationary condition is

$$\begin{aligned} \frac{\partial L}{\partial w} = 2\Sigma w - 2\lambda w &= 0, \qquad (5) \\ \Sigma w &= \lambda w. \qquad (6) \end{aligned}$$

Thus w is an eigenvector of Σ. Since

$$w'\Sigma w = w'(\lambda w) = \lambda, \tag{7}$$

the direction with maximum variance is the largest eigenvector. This procedure can be iterated to get the second largest variance projection (orthogonal to the first one), and so on. For a set of data points, we use the ML estimate of the covariance matrix.

1 High-Dimensional Data

Let X be the $d \times n$ matrix of high-dimensional points ($d > n$). Instead of explicitly estimating the covariance matrix, we use the following trick:

$$\begin{aligned} \left(\frac{1}{n}XX'\right)W &= W\Lambda, \qquad (8) \\ \left(\frac{1}{n}X'X\right)X'W &= X'W\Lambda, \qquad (9) \\ \left(\frac{1}{n}X'X\right)V &= V\Lambda. \qquad (10) \end{aligned}$$

$$\frac{1}{n}X(X'X)F = XFV, \tag{11}$$

$$\Big(\frac{1}{n}XX'\Big)(XF) = (XF)V, \tag{12}$$

so XF are the eigenvectors of $\frac{1}{n}XX'$. Since

$$\frac{1}{n}X'XF = FV, \tag{13}$$

$$(F'X')(XF) = nF'FV \tag{14}$$

$$(XF)'(XF) = nV, \tag{15}$$

we have $E = \frac{1}{\sqrt{n}}XFV^{-1/2}$.

2 Maximum Information Preservation

Let

$$y = w'x, \tag{16}$$

where $\|w\| = 1$.

We want to maximize $I(X;Y)$ or $H(Y)$ under the constraint on w:

$$J = 1/2(\log(2\pi e)^D + \log|Cy|) + aw'w. \tag{17}$$

$$Cy = \mathrm{E}[w'xx'w] = w'\,\mathrm{E}[xx']w = w'Cw. \tag{18}$$

Thus we maximize

$$L = \log|w'Cw| + aw'wdL/dw = 2Cw/|w'Cw| + 2awCw = a|w'Cw|w \tag{19}$$

Thus, w is an eigenvector of the covariance matrix C of x, and the maximum occurs with the eigenvector with the largest eigenvalue.

3 Maximum Likelihood

The data matrix is modeled as linear combinations of a small set of basis vectors plus noise.

$$X = UV + Y \tag{20}$$

X is DxN, where each column is a datum. U is DxK, where each column is a factor. V is KxN, where each column is a vector of coefficients. Y is DxN noise.

Assuming normal noise with equal variance for all points, ML estimation of UV gives a least squares cost function.

$$J = (X - UV)^2. \tag{21}$$

The stationary conditions are

$$dJ/dU = (X-UV)V' = 0dJ/dV = U'(X-UV) = 0XV' = UVV'U'X = U'UV \tag{22}$$

To make the solution unique against rotations and scalings of U and V, we constrain U'U = I and V'V diagonal:

$$V = U'XXX'U = SU1/NXX'U = S/NU \tag{23}$$

The basis functions are the eigenvectors of the covariance matrix, assuming X is zero mean .

4 Generalized PCA

To generalize PCA to other data the noise distribution can be changed and a link function added:

$$X\ p(f(g(A)h(B))). \tag{24}$$