# Regression formula

August 17, 2016

Ross Bannister

**The problem**

Suppose that we have a system of $n+1$ vectors: $\mathbf{x}$, $\mathbf{v}^{(j)}$ $(1 \leq j \leq n)$ and that the $\mathbf{v}^{(j)}$ are predictors of $\mathbf{x}$, such that the following:

$$\sum_{j=1}^{n-1} \mathbf{L}^{(j)} \mathbf{v}^{(j)} + \mathbf{v}^{(n)}$$

is as close as possible to $\mathbf{x}$, such that the difference

$$\mathbf{d} = \sum_{j=1}^{n-1} \mathbf{L}^{(j)} \mathbf{v}^{(j)} + \mathbf{v}^{(n)} - \mathbf{x}$$

is as small as possible given the population. We ask the question: what set of regression matrices, $\mathbf{L}^{(j)}$, achieves this? We may solve the problem using the method of least squares.

**Cost function**

Define a cost function, $J$, that is a function of $\mathbf{L}^{(1)}, \ldots, \mathbf{L}^{(n-1)}$:

$$\begin{aligned}
J[\mathbf{L}^{(1)}, \ldots, \mathbf{L}^{(n-1)}] &= \mathbf{d}^{\mathrm{T}} \mathbf{d}, \\
&= \left( \sum_{j=1}^{n-1} \mathbf{L}^{(j)} \mathbf{v}^{(j)} + \mathbf{v}^{(n)} - \mathbf{x} \right)^{\mathrm{T}} \left( \sum_{j'=1}^{n-1} \mathbf{L}^{(j')} \mathbf{v}^{(j')} + \mathbf{v}^{(n)} - \mathbf{x} \right), \\
&= \sum_{j,j'=1}^{n-1} \mathbf{v}^{(j)\mathrm{T}} \mathbf{L}^{(j)\mathrm{T}} \mathbf{L}^{(j')} \mathbf{v}^{(j')} + \sum_{j=1}^{n-1} \mathbf{v}^{(j)\mathrm{T}} \mathbf{L}^{(j)\mathrm{T}} \left( \mathbf{v}^{(n)} - \mathbf{x} \right) + \sum_{j'=1}^{n-1} \left( \mathbf{v}^{(n)} - \mathbf{x} \right)^{\mathrm{T}} \mathbf{L}^{(j')} \mathbf{v}^{(j')} \\
&\quad + \left( \mathbf{v}^{(n)} - \mathbf{x} \right)^{\mathrm{T}} \left( \mathbf{v}^{(n)} - \mathbf{x} \right).
\end{aligned}$$

Expanding the notation into its components:

$$\begin{aligned}
J[\mathbf{L}^{(1)}, \ldots, \mathbf{L}^{(n-1)}] &= \frac{1}{2} \left\{ \sum_{j,j'=1}^{n-1} \sum_{a,b,c} \mathbf{v}_a^{(j)} \mathbf{L}_{ba}^{(j)} \mathbf{L}_{bc}^{(j')} \mathbf{v}_c^{(j')} + \sum_{j=1}^{n-1} \sum_{a,b} \mathbf{v}_a^{(j)} \mathbf{L}_{ba}^{(j)} \left( \mathbf{v}_b^{(n)} - \mathbf{x}_b \right) \right. \\
&\quad \left. + \sum_{j'=1}^{n-1} \sum_{a,b} \left( \mathbf{v}_b^{(n)} - \mathbf{x}_b \right) \mathbf{L}_{ba}^{(j')} \mathbf{v}_a^{(j')} + \sum_{b} \left( \mathbf{v}_b^{(n)} - \mathbf{x}_b \right)^2 \right\}.
\end{aligned}$$

**The minimum of the cost function with respect to the regression matrices**

Differentiating $J$ with respect to an arbitrary component of an arbitrary regression operator, $\mathbf{L}^{(p)}_{\alpha\beta}$, and assuming $p < n$ gives:

$$\begin{aligned}
\frac{\partial J}{\partial \mathbf{L}^{(p)}_{\alpha\beta}} &= \frac{1}{2}\left\{\sum_{j,j'=1}^{n-1}\sum_{a,b,c}\mathbf{v}^{(j)}_a\delta_{\alpha b}\delta_{\beta a}\delta_{pj}\mathbf{L}^{(j')}_{bc}\mathbf{v}^{(j')}_c\right.\\
&\quad+\sum_{j,j'=1}^{n-1}\sum_{a,b,c}\mathbf{v}^{(j)}_a\mathbf{L}^{(j)}_{ba}\delta_{\alpha b}\delta_{\beta c}\delta_{pj'}\mathbf{v}^{(j')}_c\\
&\quad+\sum_{j=1}^{n-1}\sum_{a,b}\mathbf{v}^{(j)}_a\delta_{\alpha b}\delta_{\beta a}\delta_{pj}\left(\mathbf{v}^{(n)}_b-\mathbf{x}_b\right)\\
&\quad\left.+\sum_{j'=1}^{n-1}\sum_{a,b}\left(\mathbf{v}^{(n)}_b-\mathbf{x}_b\right)\delta_{\alpha b}\delta_{\beta a}\delta_{pj'}\mathbf{v}^{(j')}_a\right\},\\
&= \frac{1}{2}\left\{\sum_{,j'=1}^{n-1}\sum_{c}\mathbf{v}^{(p)}_\beta\mathbf{L}^{(j')}_{\alpha c}\mathbf{v}^{(j')}_c+\sum_{j=1}^{n-1}\sum_{a}\mathbf{v}^{(j)}_a\mathbf{L}^{(j)}_{\alpha a}\mathbf{v}^{(p)}_\beta\right.\\
&\quad\left.+\mathbf{v}^{(p)}_\beta\left(\mathbf{v}^{(n)}_\alpha-\mathbf{x}_\alpha\right)+\left(\mathbf{v}^{(n)}_\alpha-\mathbf{x}_\alpha\right)\mathbf{v}^{(p)}_\beta\right\}.
\end{aligned}$$

In the first term on the penultimate line we can relabel the dummy variables $j' \to j$, and $c \to a$:

$$\begin{aligned}
\frac{\partial J}{\partial \mathbf{L}^{(p)}_{\alpha\beta}} &= \sum_{j=1}^{n-1}\sum_{a}\mathbf{v}^{(j)}_a\mathbf{L}^{(j)}_{\alpha a}\mathbf{v}^{(p)}_\beta+\left(\mathbf{v}^{(n)}_\alpha-\mathbf{x}_\alpha\right)\mathbf{v}^{(p)}_\beta,\\
&= \sum_{j=1}^{n-1}\sum_{a}\mathbf{L}^{(j)}_{\alpha a}\mathbf{v}^{(j)}_a\mathbf{v}^{(p)}_\beta+\left(\mathbf{v}^{(n)}_\alpha-\mathbf{x}_\alpha\right)\mathbf{v}^{(p)}_\beta.
\end{aligned}$$

This is the $(\alpha, \beta)$ element of the following matrix expression:

$$\sum_{j=1}^{n-1}\mathbf{L}^{(j)}\left(\mathbf{v}^{(j)}\mathbf{v}^{(p)\mathrm{T}}\right)+\left(\mathbf{v}^{(n)}-\mathbf{x}\right)\mathbf{v}^{(p)\mathrm{T}}.$$

Setting this to zero for the optimum gives:

$$\sum_{j=1}^{n-1}\mathbf{L}^{(j)}\left(\mathbf{v}^{(j)}\mathbf{v}^{(p)\mathrm{T}}\right)+\left(\mathbf{v}^{(n)}-\mathbf{x}\right)\mathbf{v}^{(p)\mathrm{T}}=0.$$

There are $n-1$ such equations ($1 \le p < n$).

**Solving for the regression matrices**

The outer products are covariance matrices and can be estimated from a population of $\mathbf{x}$ and $\mathbf{v}^{(j)}$ vectors:

$$\begin{aligned}
\mathbf{v}^{(j)}\mathbf{v}^{(p)\mathrm{T}} &\equiv \mathbf{C}^{(jp)},\\
\mathbf{x}\mathbf{v}^{(p)\mathrm{T}} &\equiv \mathbf{C}^{(xp)}.
\end{aligned}$$

Assembling all $n-1$ systems together gives:

$$\begin{pmatrix} \mathbf{L}^{(1)} & \cdots & \mathbf{L}^{(n-1)} \end{pmatrix} \begin{pmatrix} \mathbf{C}^{(1,1)} & \cdots & \mathbf{C}^{(1,n-1)} \\ \vdots & \ddots & \vdots \\ \mathbf{C}^{(n-1,1)} & \cdots & \mathbf{C}^{(n-1,n-1)} \end{pmatrix} = \begin{pmatrix} \mathbf{C}^{(x,1)} & \cdots & \mathbf{C}^{(x,n-1)} \end{pmatrix} - \begin{pmatrix} \mathbf{C}^{(n,1)} & \cdots & \mathbf{C}^{(n,n-1)} \end{pmatrix} .$$

Assuming that different $\mathbf{v}^{(j)}$-vectors are uncorrelated means that $\mathbf{v}^{(j)}\mathbf{v}^{(p)^{\mathrm{T}}} \equiv \mathbf{C}^{(jp)}\delta_{pj}$, which makes the above into:

$$\begin{pmatrix} \mathbf{L}^{(1)} & \cdots & \mathbf{L}^{(n-1)} \end{pmatrix} \begin{pmatrix} \mathbf{C}^{(1,1)} & & \\ & \ddots & \\ & & \mathbf{C}^{(n-1,n-1)} \end{pmatrix} = \begin{pmatrix} \mathbf{C}^{(x,1)} & \cdots & \mathbf{C}^{(x,n-1)} \end{pmatrix} .$$

If all vectors are of equal size then the regression matrices emerge:

$$\mathbf{L}^{(i)} = \mathbf{C}^{(x,i)}\mathbf{C}^{(i,i)^{-1}}$$