Best linear predictor: matrix version

Gustavo Pereira

November 12, 2019

Let $M(n,k)$ denote the linear space of all matrix of dimension $n \times k$.

Suppose we have random vectors $(\mathbf{y}(\omega), \mathbf{z}(\omega))'$. We know additionally that $\mathbf{y} \in M(n,1)$ and $\mathbf{z} \in M(k,1)$ and these vectors have finite mean and variance. Denote their mean by

$$\begin{bmatrix} \mu_y \\ \mu_z \end{bmatrix}$$

and their variance matrix by

$$\begin{bmatrix} \Sigma_{yy} & \Sigma_{yz} \\ \Sigma_{zy} & \Sigma_{zz} \end{bmatrix}$$

We define the **best linear predictor** of $\mathbf{y}$ given $\mathbf{z}$ as the random variable $\mathbf{w}$ such that

$$\mathbf{w}^* = \alpha^* + \beta^*(\mathbf{z} - \mu_z)$$

where $\alpha^* \in M(n,1)$ and $\beta^* \in M(n,k)$ solve the minimzation problem

$$\min_{\alpha,\beta} \mathbf{E}\left[\|\mathbf{y} - \alpha - \beta(\mathbf{z} - \mu_z)\|^2\right]$$

You can solve it either by using calculus – which can be cumbersome if you're not used to matrix derivatives – or by noting that the minimand is a squared norm generated by the inner product

$$\langle \mathbf{y}, \mathbf{w} \rangle := \mathbf{E}[\mathbf{w}'\mathbf{y}]$$

of all vectors of the type $\mathbf{y} - \mathbf{w}$ where $\mathbf{w} = \alpha + \beta(\mathbf{z} - \mu_z)$ for some α, β.

Let $\epsilon := \mathbf{y} - \mathbf{w}^*$ denote the residual of the minimization problem. Then ϵ must be orthogonal (by Hilbert's projection theorem) to every $\mathbf{w} = \alpha + \beta(\mathbf{z} - \mu_z)$.

Taking $\beta = 0$, we see that $\mathbf{w}^*$ must satisfy

$$0 = \langle \mathbf{y} - \mathbf{w}^*, \alpha \rangle = \mathbf{E}\left[\alpha'\mathbf{y}\right] - \mathbf{E}\left[\alpha'\alpha^*\right]$$

for all vectors $\alpha \in M(n,1)$. Taking these to be the elements of the canonical basis, we conclude that

$$\alpha^* = \mu_y$$

Now take $\alpha = 0$. The orthogonality condition now implies that for any $\beta \in M(n,k)$,

$$0 = \langle \mathbf{y} - \beta^*(\mathbf{z} - \mu_z), \beta(\mathbf{z} - \mu_z) \rangle = \mathbf{E}\left[(\mathbf{z} - \mu_z)'\beta' y\right] - \mathbf{E}\left[(\mathbf{z} - \mu_z)'\beta'\beta^*(\mathbf{z} - \mu_z)\right]$$

Use the properties of the trace – namely, that it's linear and that matrix multiplication commutes inside it – and of the expectation operator to conclude that

$$\mathrm{tr}\left(\beta'\mathbf{E}\left[\mathbf{y}(\mathbf{z} - \mu_z)'\right]\right) = \mathrm{tr}\left(\beta'\beta^*\mathbf{E}\left[(\mathbf{z} - \mu_z)(\mathbf{z} - \mu_z)'\right]\right)$$

note that $\mathbf{E}[\mathbf{y}(\mathbf{z} - \mu_z)'] = \Sigma_{yz}$ and $\mathbf{E}[(\mathbf{z} - \mu_z)(\mathbf{z} - \mu_z)'] = \Sigma_{zz}$. The equation above then implies that

$$\mathrm{tr}(\beta'\Sigma_{yz}) = \mathrm{tr}(\beta'\beta^*\Sigma_{zz})$$

should hold for all matrices $\beta \in M(n,k)$. That implies,[1]

$$\Sigma_{yz} = \beta^*\Sigma_{zz}$$

which in turn yields $\beta^* = \Sigma_{yz}\Sigma_{zz}^{-1}$ whenever Σ_{zz} has an inverse. In that case, the BLP is

$$\mathbf{w}^* = \mu_y + \Sigma_{yz}\Sigma_{zz}^{-1}(\mathbf{z} - \mu_z) \tag{1}$$

1 Appendix: the Trace operator

- let $A(i,j)$ denote the entry (i,j) of any matrix
- Let A be a $m \times n$ matrix. The trace is defined as

  $$\mathrm{tr}A = \sum_{i=1}^{\min\{m,n\}} A(i,i)$$

  in other words, it's just the sum of elements in the main diagonal.
- Some properties of the trace:
  1. $\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B)$ whenever A and B have similar dimensions

[1] See the appendix on the trace operator for details.

2. $\text{tr}(kA) = k\,\text{tr}(A)$ for all scalars k
3. $\text{tr}(AB) = \text{tr}(BA)$ whenever dimensions are such that both multiplications make sense

Curiosity: any operation $\tilde{\text{tr}}$ that satisfies the properties above is equal to tr (modulo multiplication by a constant)

- The trace and expectation operators commute:

$$\text{tr}(\mathbf{E}A) = \mathbf{E}(\text{tr}A)$$

- Suppose $A \in M(m, n)$ and you want to select element (i, j) from it. Note that

$$A(i, j) = e_i' A \varepsilon_j = tr(e_i' A \varepsilon_j) = tr(\varepsilon_j e_i' A)$$

where e_i is the i-th element in the canonical basis of R^m and ε_j is the j-th element of the canonical basis of R^n.

Hence for any (i, j), letting $B = \varepsilon_j e_i' \in M(n, m)$ we have

$$A(i, j) = \text{tr}(BA)$$

- This implies that if A and $\tilde{A}$ are fixed $m \times n$ matrices, and

$$\text{tr}(BA) = tr(B\tilde{A})$$

holds for every $B \in M(n, m)$, then

$$A = \tilde{A}$$