Thoughts about covariance

Andy Smith

September 2012

We have N measurements stored in an array $\mathbf{X}$:

$$\mathbf{X} = [\ \mathbf{x}_1\ \mathbf{x}_2\ \ldots\ \mathbf{x}_N\ ] \tag{1}$$

where $\mathbf{x}_i$ is the i-th measurement vector, and has M elements. If we say that the mean vector is

$$\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i, \tag{2}$$

and will also have M elements, then we can define

$$\mathbf{Y} = [\ \bar{\mathbf{x}}\ \bar{\mathbf{x}}\ \ldots\ \bar{\mathbf{x}}\ ] \ = \ \bar{\mathbf{x}}\,\mathbf{1}_N^{\mathrm{T}} \tag{3}$$

where $\mathbf{1}_N$ is a column vector of length N, with every element set to 1.

Using $\mathbf{X}$ and $\mathbf{Y}$ we can write the covariance, $\mathbf{S}$ as:

$$\begin{aligned}
\mathbf{S} &= \frac{1}{N-1}(\mathbf{X}-\mathbf{Y})(\mathbf{X}-\mathbf{Y})^{\mathrm{T}} && (4)\\
&= \frac{1}{N-1}\left(\mathbf{X}-\bar{\mathbf{x}}\,\mathbf{1}_N^{\mathrm{T}}\right)\left(\mathbf{X}-\bar{\mathbf{x}}\,\mathbf{1}_N^{\mathrm{T}}\right)^{\mathrm{T}} && (5)\\
&= \frac{1}{N-1}\left(\mathbf{X}\mathbf{X}^{\mathrm{T}} - \left(\mathbf{X}\mathbf{1}_N\,\bar{\mathbf{x}}^{\mathrm{T}}\right)^{\mathrm{T}} - \mathbf{X}\mathbf{1}_N\,\bar{\mathbf{x}}^{\mathrm{T}} + \bar{\mathbf{x}}\,\mathbf{1}_N^{\mathrm{T}}\,\mathbf{1}_N\,\bar{\mathbf{x}}^{\mathrm{T}}\right). && (6)
\end{aligned}$$

But $\mathbf{X}\,\mathbf{1}_N = N\bar{\mathbf{x}}$ (since it is the sum of elements in the rows of $\mathbf{X}$), and $\mathbf{1}_N^{\mathrm{T}}\,\mathbf{1}_N = N$, so that:

$$\begin{aligned}
\mathbf{S} &= \frac{1}{N-1}\left(\mathbf{X}\mathbf{X}^{\mathrm{T}} - 2N\bar{\mathbf{x}}\bar{\mathbf{x}}^{\mathrm{T}} + N\bar{\mathbf{x}}\bar{\mathbf{x}}^{\mathrm{T}}\right) && (7)\\
\mathbf{S} &= \frac{1}{N-1}\left(\mathbf{X}\mathbf{X}^{\mathrm{T}} - N\bar{\mathbf{x}}\bar{\mathbf{x}}^{\mathrm{T}}\right). && (8)
\end{aligned}$$

Now imagine that we want to combine two areas of a dataset together:

$$\mathbf{X_1} = [\ \mathbf{x}_1\ \mathbf{x}_2\ \ldots \mathbf{x}_{N_1}\ ] \ \text{ and } \ \mathbf{X_2} = [\ \mathbf{x}_{N_1+1}\ \mathbf{x}_{N_1+2}\ \ldots \mathbf{x}_{N_2}\ ] \tag{9}$$

so that $\mathbf{X} = [\ \mathbf{X_1}\ \mathbf{X_2}\ ]$ gives us the original dataset. In this case, the mean vector can be obtained easily from:

$$\bar{\mathbf{x}} = \frac{N_1\bar{\mathbf{x}}_1 + N_2\bar{\mathbf{x}}_2}{N_1+N_2}. \tag{10}$$

The vast majority of computing time for $\mathbf{S}$ is in the matrix multiplication $\mathbf{X}\mathbf{X}^{\mathrm{T}}$, so a way to combine $\mathbf{X_1}\mathbf{X_1}^{\mathrm{T}}$ and $\mathbf{X_2}\mathbf{X_2}^{\mathrm{T}}$ in a timely fashion would be very pleasing. Happily, since $\left(\mathbf{X}\mathbf{X}^{\mathrm{T}}\right)_{ij} = \sum_k x_{ik}\,x_{jk}$, then $\mathbf{X}\mathbf{X}^{\mathrm{T}} = \mathbf{X_1}\mathbf{X_1}^{\mathrm{T}} + \mathbf{X_2}\mathbf{X_2}^{\mathrm{T}}$.