Online correlation computation

Moo K. Chung
University of Wisconsin-Madison
mkchung@wisc.edu

May 21, 2020

Data a added to $\mathbf{x} = (x_1, x_2, \cdots, x_{m-1})$ such that $\mathbf{x}_a = (x_1, \cdots, x_{m-1}, a)$. Let $\nu(\mathbf{x}) = \sum_{j=1}^{m-1} x_j$. Then

$$\nu(\mathbf{x}_a) = \nu(\mathbf{x}) + a.$$

If data changed from $\mathbf{x}_a$ to $\mathbf{x}_b$. Then we have

$$\nu(\mathbf{x}_b) = \nu(\mathbf{x}_a) - a + b.$$

Similarly, let

$$\omega(\mathbf{x}, \mathbf{x}) = \sum_{j=1}^{m-1} (x_j - \nu(\mathbf{x})/(m-1))^2 = \sum_{j=1}^{m-1} x_j^2 - \nu(\mathbf{x})^2/(m-1).$$

Then

$$\omega(\mathbf{x}_a, \mathbf{x}_a) = \sum_{j=1}^{m-1} x_j^2 + a^2 - \nu(\mathbf{x}_a)^2/m$$

Similarly, we obtain

$$\omega(\mathbf{x}_b, \mathbf{x}_b) = \sum_{j=1}^{m-1} x_j^2 + b^2 - \nu(\mathbf{x}_b)^2/m.$$

Thus,

$$\omega(\mathbf{x}_b, \mathbf{x}_b) = \omega(\mathbf{x}_a, \mathbf{x}_a) - a^2 + b^2 + \big(\nu(\mathbf{x}_a)^2 - \nu(\mathbf{x}_b)^2\big)/m.$$

Let's extend the result to the covariance. Suppose data c added to $\mathbf{y} = (y_1, y_2, \cdots, y_{m-1})$ such that $\mathbf{y}_c = (y_1, \cdots, y_{m-1}, c)$. Suppose data changed from $\mathbf{y}_c$ to $\mathbf{y}_d$. Then let's see how the covariance between $\mathbf{x}_a$ and $\mathbf{y}_c$ changed to the covariance between $\mathbf{x}_b$ and $\mathbf{y}_d$.

Define

$$\omega(\mathbf{x},\mathbf{y}) = \sum_{j=1}^{m-1}(x_j - \frac{\nu(\mathbf{x})}{m-1})(y_j - \frac{\nu(\mathbf{y})}{m-1}) = \sum_{j=1}^{m-1} x_j y_j - \frac{\nu(\mathbf{x})\nu(\mathbf{y})}{m-1}.$$

Following the similar method as before, let's compute

$$\omega(\mathbf{x}_a,\mathbf{y}_c) = \sum_{j=1}^{m-1} x_j y_j + ac - \frac{\nu(\mathbf{x}_a)\nu(\mathbf{y}_c)}{m}.$$

Similarly,

$$\omega(\mathbf{x}_b,\mathbf{y}_d) = \sum_{j=1}^{m-1} x_j y_j + bd - \frac{\nu(\mathbf{x}_b)\nu(\mathbf{y}_d)}{m}.$$

Thus, for covariance, we use the following update scheme:

$$\begin{aligned}
\nu(\mathbf{x}_b) &= \nu(\mathbf{x}_a) - b + a \\
\nu(\mathbf{y}_d) &= \nu(\mathbf{y}_c) - c + d \\
\omega(\mathbf{x}_b,\mathbf{y}_d) &= \omega(\mathbf{x}_a,\mathbf{y}_c) + bd - ac - \frac{\nu(\mathbf{x}_b)\nu(\mathbf{y}_d) - \nu(\mathbf{x}_a)\nu(\mathbf{y}_c)}{m}.
\end{aligned}$$

For transposition between the twins, the above general formula should be used. For the transposition within twin, a simpler formula can be derived since $c = b$ and $d = a$:

$$\begin{aligned}
\omega(\mathbf{x}_b,\mathbf{y}_a) &= \omega(\mathbf{x}_a,\mathbf{y}_b) + \frac{\nu(\mathbf{x}_a)\nu(\mathbf{y}_b) - \nu(\mathbf{x}_b)\nu(\mathbf{y}_a)}{m} \\
&= \omega(\mathbf{x}_a,\mathbf{y}_b) + \frac{(b-a)(\nu(\mathbf{x}) - \nu(\mathbf{y}))}{m} \\
&= \omega(\mathbf{x}_a,\mathbf{y}_b) + \frac{(b-a)(\nu(\mathbf{x}_a) - a - \nu(\mathbf{y}_b) + b)}{m} \\
&= \omega(\mathbf{x}_a,\mathbf{y}_b) + (a-b)^2/m + (b-a)(\nu(\mathbf{x}_a) - \nu(\mathbf{y}_b))/m
\end{aligned}$$