Obtaining the AR parameters of AR Model by Yule-Walker Equation

Andersen Ang

First created: between 2015-2013. Last update: 2017-Feb.

The problem

The equation of predicting x_k based on previous data are

$$x_k = f\left(x_{k-1}, x_{k-2}, ...;\, a_1, a_2, ...\right)$$

where $x_{k-1}, x_{k-2}, ...$ are those previous data, and a_i are parameters, there can be any number of previous data or parameters for scuh general model

There is always noise, so additive white Gaussian noise ε should be included

$$x_k = f\left(x_{k-1}, x_{k-2}, ...;\, a_1, a_2, ...\right) + \varepsilon_k$$

Such general model is not easy to solve, thus restrict the problem into linear model

$$x_k = a_1 x_{k-1} + a_2 x_{k-2} + ... + \varepsilon_k$$

$$x_k = \sum_{i=1}^{\infty} a_i x_{k-i} + \varepsilon_k$$

It is still difficult to solve for infinite number of parameters, so the equation is now reduced to finite number of terms, p terms.

$$x_k = \sum_{i=1}^{p} a_i x_{k-i} + \varepsilon_k$$

The AR equation

$$x_k = \sum_{i=1}^{p} a_i x_{k-i} + \varepsilon_k$$

Rearrange

$$\varepsilon_k = x_k - \sum_{i=1}^{p} a_i x_{k-i}$$

Since the goal is to estimate x_k, thus the noise should be minimized

$$\text{minimize the effect of } \varepsilon_k$$

For minimization such random variable, the *power minimization* is used : that is, minimize the expectation of ε_k^2

$$\min\, E\left(\varepsilon_k^2\right)$$

Since we have an expression of ε_k , and thus

$$\min\ E\left(\varepsilon_k^2\right) = \min E\left(\left[x_k - \sum_{i=1}^{p} a_i x_{k-i}\right]^2\right)$$

To find the minima , take derivatives and set to zero

$$\frac{\partial}{\partial a_i} E\left(\left[x_k - \sum_{i=1}^{p} a_i x_{k-i}\right]^2\right) = 0 \qquad \forall i$$

Since derivative and expectation operator is linear, thus they are interchangable

$$E\left(\frac{\partial}{\partial a_i}\left[x_k - \sum_{i=1}^{p} a_i x_{k-i}\right]^2\right) = 0 \qquad \forall i$$

The derivative part can be solved by simple chain rule, for example :

$$\frac{\partial}{\partial a_1}\left[Aa_1 + Ba_2\right]^2 = 2\left(Aa_1 + Ba_2\right)A$$

Therefore the equation now becomes

$$E\left(2\left[x_k - \sum_{i=1}^{p} a_i x_{k-i}\right] x_{k-i}\right) = 0 \qquad \forall i$$

The 2 can be removed

$$E\left(\left[x_k - \sum_{i=1}^{p} a_i x_{k-i}\right] x_{k-i}\right) = 0 \qquad \forall i$$

Now explicitly list out all the equations

$$\begin{cases} E\left(\left[x_k - \sum_{i=1}^{p} a_i x_{k-i}\right] x_{k-1}\right) = 0 \\ E\left(\left[x_k - \sum_{i=1}^{p} a_i x_{k-i}\right] x_{k-2}\right) = 0 \\ \vdots \\ E\left(\left[x_k - \sum_{i=1}^{p} a_i x_{k-i}\right] x_{k-p}\right) = 0 \end{cases}$$

Consider the first equation

$$E\left(\left[x_k - \sum_{i=1}^{p} a_i x_{k-i}\right] x_{k-1}\right) = 0$$

Expand the first equation

$$E\left(\left[x_k - a_1 x_{k-1} - a_2 x_{k-2} - \ldots - a_p x_{k-p}\right] x_{k-1}\right) = 0$$

Multiply the x_{k-1} term inside $[\,]$

$$E\left(x_k x_{k-1} - a_1 x_{k-1}^2 - a_2 x_{k-1} x_{k-2} - ... - a_p x_{k-1} x_{k-p}\right) = 0$$

Expectation operator is linear, so

$$E\left(x_k x_{k-1}\right) - a_1 E\left(x_{k-1}^2\right) - a_2 E\left(x_{k-1} x_{k-2}\right) - ... - a_p E\left(x_{k-1} x_{k-p}\right) = 0$$

Each term of the expectation is an $autocorrelation\, r$ with different time lag

$$r\left(1\right) - a_1 r\left(0\right) - a_2 r\left(1\right) - ... - a_p r\left(p-1\right) = 0$$

Rearrange

$$r\left(1\right) = a_1 r\left(0\right) + a_2 r\left(1\right) + ... + a_p r\left(p-1\right)$$

$$\begin{cases} E\left(\left[x_k - \sum_{i=1}^{p} a_i x_{k-i}\right] x_{k-1}\right) = 0 \\ E\left(\left[x_k - \sum_{i=1}^{p} a_i x_{k-i}\right] x_{k-2}\right) = 0 \\ \vdots \\ E\left(\left[x_k - \sum_{i=1}^{p} a_i x_{k-i}\right] x_{k-p}\right) = 0 \end{cases} \Longrightarrow \begin{cases} r\left(1\right) = a_1 r\left(0\right) + a_2 r\left(1\right) + ... + a_p r\left(p-1\right) \\ r\left(2\right) = a_1 r\left(1\right) + a_2 r\left(2\right) + ... + a_p r\left(p-2\right) \\ \vdots \\ r\left(p\right) = a_1 r\left(p-1\right) + a_2 r\left(p-2\right) + ... + a_p r\left(0\right) \end{cases}$$

Consider notation

$$\mathbf{r} = \left[r(1), r(2), ..., r(p)\right]^T$$

$$\mathbf{a} = \left[a_1, a_2, ..., a_p\right]^T$$

$$\mathbf{R} = \begin{bmatrix} r(0) & r(1) & ... & r(p-1) \\ r(1) & r(2) & ... & r(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(p-1) & r(p-2) & ... & r(0) \end{bmatrix}$$

Thus

$$\mathbf{r} = \mathbf{Ra} \text{ or } \mathbf{a} = \mathbf{R^{-1}r}$$

$-END-$