AR model

Algorithm

AR model (dim=2) is

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \epsilon_t \tag{1}$$

and

$$\epsilon_t \sim N(0, \sigma^2), \tag{2}$$

$$E[y_t] = \mu (\because \text{stationality}) \tag{3}$$

are assumed.

The coefficents of AR model can be estimated using the Yule Walker method, which is based on the maximum likelihood estimation.

By the equations (1)∼(3)

$$\begin{align} E[y_t] &= a_1 E[y_{t-1}] + a_2 E[y_{t-2}] + E[\epsilon_t] \tag{4} \\ \mu &= a_1 \mu + a_2 \mu \tag{5} \\ \mu &= 0. \tag{6} \end{align}$$

By the equation (6), put autocovariance γ;

$$\begin{align} \gamma_k &= Cov(y_t, y_{t-k}) \tag{7} \\ &= E[(y_t - \mu)(y_{t-k} - \mu)] \tag{8} \\ &= E[y_t y_{t-k}]. \tag{9} \end{align}$$

From the assumption (2),

$$Cov(\epsilon_t, \epsilon_t) = \sigma^2 \tag{10}$$

$$Cov(\epsilon_i, \epsilon_j) = 0 \qquad (i \neq j). \tag{11}$$

By the equations (1)(9)(11),

$$\begin{align} E[y_{t-1} y_t] &= E[y_{t-1}(a_1 y_{t-1} + a_2 y_{t-2} + \epsilon_t)] \tag{12} \\ &= a_1 E[y_{t-1} y_{t-1}] + a_2 E[y_{t-1} y_{t-2}] + E[y_{t-1} \epsilon_t] \tag{13} \\ &= a_1 \gamma_0 + a_2 \gamma_1 \tag{14} \\ \therefore \gamma_1 &= a_1 \gamma_0 + a_2 \gamma_1. \tag{15} \end{align}$$

Similary,

$$\begin{aligned}
E[y_{t-2}y_t] &= E[y_{t-2}(a_1 y_{t-1} + a_2 y_{t-2} + \epsilon_t)] && (16)\\
&= a_1 E[y_{t-2}y_{t-1}] + a_2 E[y_{t-2}y_{t-2}] + E[y_{t-2}\epsilon_t] && (17)\\
&= a_1\gamma_1 + a_2\gamma_0 && (18)\\
\therefore \gamma_2 &= a_1\gamma_1 + a_2\gamma_0. && (19)
\end{aligned}$$

The equations(15)(19) are written as simultaneous equations;

$$\begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} = \begin{pmatrix} \gamma_0 & \gamma_1 \\ \gamma_1 & \gamma_0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}. \tag{20}$$

If observed sequence data $y_1, y_2, \ldots, y_T \in R$ is given,

$$\hat{\gamma}_k = \frac{1}{T}\sum_{t=k+1}^{T}(y_t - \hat{\mu})(y_{t-k} - \hat{\mu}) \tag{21}$$

$$\begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{pmatrix} = \begin{pmatrix} \hat{\gamma}_0 & \hat{\gamma}_1 \\ \hat{\gamma}_1 & \hat{\gamma}_0 \end{pmatrix} \begin{pmatrix} \hat{a}_1 \\ \hat{a}_2 \end{pmatrix}. \tag{22}$$

By solving the simultaneous equations(22), the estimated coefficients $\hat{a}_1$ and $\hat{a}_2$ are obtained. Moreover, by the equations (1)(9)(10)(15)(19),

$$\begin{aligned}
E[y_t y_t] &= E[(a_1 y_{t-1} + a_2 y_{t-2} + \epsilon_t)(a_1 y_{t-1} + a_2 y_{t-2} + \epsilon_t)] && (23)\\
&= a_1(a_1\gamma_0 + a_2\gamma_1) + a_2(a_1\gamma_1 + a_2\gamma_0) + \sigma^2 && (24)\\
&= a_1\gamma_1 + a_2\gamma_2 + \sigma^2 && (25)\\
\therefore \sigma^2 &= \gamma_0 - (a_1\gamma_1 + a_2\gamma_2). && (26)
\end{aligned}$$

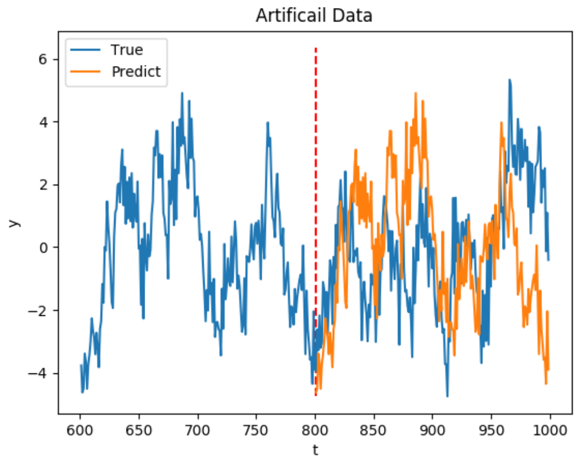
the estimated variance $\hat{\sigma}^2$ is obtained by the equation (26);

$$\hat{\sigma}^2 = \hat{\gamma}_0 - (a_1\hat{\gamma}_1 + a_2\hat{\gamma}_2). \tag{27}$$

Examples

Artificial data

The artificial data is generated by $y_t = 0.5y_{t-1} + 0.4y_{t-2} + \epsilon_t (\sim N(0,1))$.



Economic data

The economic data is Canada in vars package(https://www.rdocumentation.org/packages/vars/versions/1.5-3). This data includes labor productivity, employment, unemployment rate and real wage from 1980 to 2000 in Canada. Since this data is nonstationary, AR model cannot predict successfully.