Mínimos cuadrados recursivos

El algoritmo de mínimos cuadrados se aplica en forma recursiva a medida que se dispone de nuevos datos. El vector $\boldsymbol{\theta}(k)$ en el instante k, se calcula con los datos (pares $\left(y(k)\,,\boldsymbol{\varphi}(k)\right)$) disponibles hasta el instante k. Para tal fin se parte de la ecuación normal y se reescribe en forma recursiva.

$$\hat{\boldsymbol{\theta}}(N)=\left(\sum_{k=1}^{N}\boldsymbol{\varphi}(k)\,\boldsymbol{\varphi}^{T}(k)\right)^{-1}\left(\sum_{k=1}^{N}\boldsymbol{\varphi}(k)\,y(k)\right)=\boldsymbol{P}(N)\left(\sum_{k=1}^{N}\boldsymbol{\varphi}(k)\,y(k)\right) \tag{1-a}$$

$$\boldsymbol{P}(N)\triangleq\left(\sum_{k=1}^{N}\boldsymbol{\varphi}(k)\,\boldsymbol{\varphi}^{T}(k)\right)^{-1}\in\ \Re^{d\times d} \tag{1-b}$$

De la definición de $\boldsymbol{P}(N)$

$$\boldsymbol{P}(k)^{-1}=\sum_{j=1}^{k}\boldsymbol{\varphi}(j)\,\boldsymbol{\varphi}^{T}(j)=\sum_{j=1}^{k-1}\boldsymbol{\varphi}(j)\,\boldsymbol{\varphi}^{T}(j)+\boldsymbol{\varphi}(k)\,\boldsymbol{\varphi}^{T}(k)=\boldsymbol{P}(k-1)^{-1}+\boldsymbol{\varphi}(k)\,\boldsymbol{\varphi}^{T}(k) \tag{2}$$

la ecuación normal es:

$$\hat{\boldsymbol{\theta}}(k)=\boldsymbol{P}(k)\left(\sum_{j=1}^{k}\boldsymbol{\varphi}(j)\,y(j)\right)=\boldsymbol{P}(k)\left(\sum_{j=1}^{k-1}\boldsymbol{\varphi}(j)\,y(j)+\boldsymbol{\varphi}(k)\,y(k)\right) \tag{3}$$

teniendo en cuenta que

$$\sum_{j=1}^{k-1}\boldsymbol{\varphi}(j)\,y(j)=\boldsymbol{P}(k-1)^{-1}\,\hat{\boldsymbol{\theta}}(k-1)=\boldsymbol{P}(k)^{-1}\,\hat{\boldsymbol{\theta}}(k-1)-\boldsymbol{\varphi}(k)\,\boldsymbol{\varphi}^{T}(k)\,\hat{\boldsymbol{\theta}}(k-1) \tag{4}$$

reemplazando la (4) en la (3), se tiene:

$$\hat{\boldsymbol{\theta}}(k)=\boldsymbol{P}(k)^{-1}\left(\boldsymbol{P}(k)^{-1}\,\hat{\boldsymbol{\theta}}(k-1)-\boldsymbol{\varphi}(k)\,\boldsymbol{\varphi}^{T}(k)\,\hat{\boldsymbol{\theta}}(k-1)+\boldsymbol{\varphi}(k)\,y(k)\right)=$$

$$=\hat{\boldsymbol{\theta}}(k-1)-\boldsymbol{P}(k)\,\boldsymbol{\varphi}(k)\,\boldsymbol{\varphi}^{T}(k)\,\hat{\boldsymbol{\theta}}(k-1)+\boldsymbol{P}(k)\,\boldsymbol{\varphi}(k)\,y(k)=$$

$$=\hat{\boldsymbol{\theta}}(k-1)+\boldsymbol{P}(k)\,\boldsymbol{\varphi}(k)\left(y(k)-\boldsymbol{\varphi}^{T}(k)\,\hat{\boldsymbol{\theta}}(k-1)\right)=\hat{\boldsymbol{\theta}}(k-1)+\boldsymbol{K}(k)\,\varepsilon(k)$$

luego

$$\hat{\boldsymbol{\theta}}(k)=\hat{\boldsymbol{\theta}}(k-1)+\boldsymbol{K}(k)\,\varepsilon(k) \tag{5}$$

con

$$\boldsymbol{K}(k)=\boldsymbol{P}(k)\,\boldsymbol{\varphi}(k) \tag{6}$$

$$\varepsilon(k)=y(k)-\boldsymbol{\varphi}^{T}(k)\,\hat{\boldsymbol{\theta}}(k-1) \tag{7}$$

luego para que el algoritmo funcione necesitamos calcular en forma recursiva $\boldsymbol{K}(k)$ o lo que es equivalente $\boldsymbol{P}(k)$ lo que se tiene hasta el momento es el cálculo recursivo de $\boldsymbol{P}(k)^{-1}$ mediante la (2), para resolver el problema se emplea el siguiente Lema.

Lema de inversión de matrices

Sean $\boldsymbol{A}$, $\boldsymbol{C}$ y $\boldsymbol{C}^{-1}+\boldsymbol{D}\boldsymbol{A}^{-1}\boldsymbol{B}$ matrices cuadradas no-singulares, entonces:

$$(\boldsymbol{A}+\boldsymbol{B}\boldsymbol{C}\boldsymbol{D})^{-1}=\boldsymbol{A}^{-1}-\boldsymbol{A}^{-1}\boldsymbol{B}\left(\boldsymbol{C}^{-1}+\boldsymbol{D}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1}\boldsymbol{D}\boldsymbol{A}^{-1}$$

Prueba: La prueba se obtiene por sustitución directa

$$(\boldsymbol{A}+\boldsymbol{B}\boldsymbol{C}\boldsymbol{D})\left(\boldsymbol{A}^{-1}-\boldsymbol{A}^{-1}\boldsymbol{B}\left(\boldsymbol{C}^{-1}+\boldsymbol{D}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1}\boldsymbol{D}\boldsymbol{A}^{-1}\right)=$$

$$=\boldsymbol{I}+\boldsymbol{B}\boldsymbol{C}\boldsymbol{D}\boldsymbol{A}^{-1}-\boldsymbol{B}\left(\boldsymbol{C}^{-1}+\boldsymbol{D}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1}\boldsymbol{D}\boldsymbol{A}^{-1}-\boldsymbol{B}CD\boldsymbol{A}^{-1}B\left(\boldsymbol{C}^{-1}+\boldsymbol{D}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1}\boldsymbol{D}\boldsymbol{A}^{-1}=$$

$$=\boldsymbol{I}+\boldsymbol{B}\boldsymbol{C}\boldsymbol{D}\boldsymbol{A}^{-1}-\boldsymbol{B}\boldsymbol{C}\left(\boldsymbol{C}^{-1}+\boldsymbol{D}\boldsymbol{A}^{-1}\boldsymbol{B}\right)\left(\boldsymbol{C}^{-1}+\boldsymbol{D}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1}\boldsymbol{D}\boldsymbol{A}^{-1}=$$

$$=\boldsymbol{I}+\boldsymbol{B}\boldsymbol{C}\boldsymbol{D}\boldsymbol{A}^{-1}-\boldsymbol{B}\boldsymbol{C}\boldsymbol{D}\boldsymbol{A}^{-1}=\boldsymbol{I}$$ **Q.E.D.**

Desarrollando $\boldsymbol{P}(k)$ de la (1-b) se tiene

$$\boldsymbol{P}(k)=\left(\boldsymbol{\Phi}^T(k)\boldsymbol{\Phi}(k)\right)^{-1}=\left(\boldsymbol{\Phi}^T(k-1)\boldsymbol{\Phi}(k-1)+\varphi(k)\varphi^T(k)\right)^{-1}=\left(\boldsymbol{P}(k-1)^{-1}+\varphi(k)\varphi^T(k)\right)^{-1}$$

aplicando el lema de inversiónde matrices con: $\boldsymbol{P}(k-1)=\boldsymbol{A}$, $\varphi(k)=\boldsymbol{B}$, $\boldsymbol{I}=\boldsymbol{C}$ y $\varphi^T(k)=\boldsymbol{D}$ se tiene

$$\boldsymbol{P}(k)=\boldsymbol{P}(k-1)-\boldsymbol{P}(k-1)\,\varphi(k)\left(\boldsymbol{I}+\varphi^T(k)\,\boldsymbol{P}(k-1)\,\varphi(k)\right)^{-1}\varphi^T(k)\,\boldsymbol{P}(k-1) \tag{8}$$

La (8) implica que

$$\boldsymbol{K}(k)=\boldsymbol{P}(k)\,\varphi(k)=\boldsymbol{P}(k-1)\,\varphi(k)\left(\boldsymbol{I}+\varphi^T(k)\,\boldsymbol{P}(k-1)\,\varphi(k)\right)^{-1} \tag{9}$$

La recurrencia se obtiene al reemplazar la (9) en la (8)

$$\boldsymbol{P}(k)=\left(\boldsymbol{I}-\boldsymbol{K}(k)\,\varphi^T(k)\right)\boldsymbol{P}(k-1) \tag{10}$$

Notar de la (8) ó (9) que para calcular $\boldsymbol{P}(k)$ de todos modos debe invertirse una matriz (la matriz $\left(\boldsymbol{I}+\varphi^T(k)\,\boldsymbol{P}(k-1)\,\varphi(k)\right)$), pero si el sistema es de una salida $y(k)$ la matriz $\left(\boldsymbol{I}+\varphi^T(k)\,\boldsymbol{P}(k-1)\,\varphi(k)\right)$ es un escalar y solo se debe invertir un número. (recordar que $\boldsymbol{P}(k-1)\in\Re^{d\times d}$ y que $\varphi(k)\in\Re^{d}$).

Todo el desarrollo anterior se resume en el siguiente teorema (que ya se a probado)

Teorema: (Estimación por mínimos cuadrados recursivos)

Asuma que la matriz $\boldsymbol{\Phi}(k)$ es de rango completo para $k \geq k_0$. Si $\hat{\boldsymbol{\theta}}$ es la estima en el sentido de los mínimos cuadrados, entonces se satisfacen las ecuaciones recursivas.

$$\hat{\boldsymbol{\theta}}(k) = \hat{\boldsymbol{\theta}}(k-1) + \boldsymbol{K}(k)\left(y(k) - \boldsymbol{\varphi}^T(k)\,\hat{\boldsymbol{\theta}}(k-1)\right) \tag{11}$$

$$\boldsymbol{K}(k) = \boldsymbol{P}(k)\,\boldsymbol{\varphi}(k) = \boldsymbol{P}(k-1)\,\boldsymbol{\varphi}(k)\left(\boldsymbol{I} + \boldsymbol{\varphi}^T(k)\,\boldsymbol{P}(k-1)\,\boldsymbol{\varphi}(k)\right)^{-1} \tag{12}$$

$$\boldsymbol{P}(k) = \left(\boldsymbol{I} - \boldsymbol{K}(k)\,\boldsymbol{\varphi}^T(k)\right)\boldsymbol{P}(k-1) \tag{13}$$

Observaciones

☞ La Ec.(11) tiene una interpretación intuitiva. La estima del vector de parámetros $\hat{\boldsymbol{\theta}}(k)$ se obtiene adicionando un factor de corrección a la estima del instante anterior $\hat{\boldsymbol{\theta}}(k-1)$. La corrección es proporcional a $y(k) - \boldsymbol{\varphi}^T(k)\,\hat{\boldsymbol{\theta}}(k-1)$, donde el último término puede interpretarse como el valor de $y(k)$ predicho por el modelo de la Ec.(75) $(\hat{y}(k\,|\,\boldsymbol{\theta}) = \boldsymbol{\varphi}^{\boldsymbol{T}}(k)\,\boldsymbol{\theta} = \boldsymbol{\theta}^T\,\boldsymbol{\varphi}(k))$. El término de corrección es entonces proporcional a la diferencia entre los valores medidos y predichos de $y(k)$ basados en los parámetros disponibles para el cálculo. Si la diferencia entre los valores medidos y predichos de $y(k)$ es nula entonces $\hat{\boldsymbol{\theta}}(k) = \hat{\boldsymbol{\theta}}(k-1) = \boldsymbol{\theta}^*$.

☞ Todo método de estimación recursiva da como resultado ecuaciones similar a la Ec.(11), donde lo que cambia de un método a otro es la forma de calcular el vector $\boldsymbol{K}(k)$.

Notar que el teorema exige que la matriz $\boldsymbol{P}(k)$ esté definida y esto ocurre cuando la matriz $\boldsymbol{\Phi}^T(k)\,\boldsymbol{\Phi}(k)$ es no-singular (invertible). Como

$$\boldsymbol{\Phi}^T(k)\,\boldsymbol{\Phi}(k) = \sum_{j=1}^{k} \boldsymbol{\varphi}(j)\,\boldsymbol{\varphi}^T(j)$$

Se sigue que la matriz $\boldsymbol{\Phi}^T(k)\,\boldsymbol{\Phi}(k)$ siempre es singular si k es lo suficientemente pequeño (Se dispone de pocos datos y se deben calcular muchos parámetros o en término de sistemas de ecuaciones se tienen pocas ecuaciones y muchas incógnitas por lo que el sistema no está determinado), se debe verificar que $k \geq d$ (el número de filas de $\boldsymbol{\Phi}(k)$ debe ser mayor o igual que el número de columnas, en general $k \gg d$).

Para evitar el problema descripto y que el algoritmo se pueda ejecutar desde el instante inicial se inicializa la matriz $\boldsymbol{P}$ de la siguiente manera:

$$\boldsymbol{P}(0) = \boldsymbol{P}_0$$

Donde $\boldsymbol{P}_0$ es una matriz definida positiva, entonces

$$\boldsymbol{P}(k) = \left(\boldsymbol{P}_{\boldsymbol{0}}^{-1} + \boldsymbol{\Phi}^T(k)\,\boldsymbol{\Phi}(k)\right)^{-1}$$

$\boldsymbol{P}(k)$ puede ser tan próxima a $\left(\boldsymbol{\Phi}^T(k)\,\boldsymbol{\Phi}(k)\right)^{-1}$ como se quiera eligiendo $\boldsymbol{P}_0$ lo suficientemente grande, en general se elige

$$\boldsymbol{P}_0 = \alpha\,\boldsymbol{I}$$

con $\alpha > 10^6$

El método de mínimos cuadrados recursivo se resume en el siguiente algoritmo:

Algoritmo (Mínimos cuadrados recursivo)

1- **Inicializar** $\boldsymbol{\theta}(0) = 0$**,** $\boldsymbol{P}(0) = \boldsymbol{P}_0$**,** α **(cota en el error del modelo permitida es un número pequeño)**

2- $k = 1$

3- **Calcular** $\boldsymbol{P}(1) = \left(\boldsymbol{P}_{\boldsymbol{0}}^{-1} + \boldsymbol{\Phi}^T(1)\,\boldsymbol{\Phi}(1)\right)^{-1} = \left(\boldsymbol{P}_{\boldsymbol{0}}^{-1} + \boldsymbol{\varphi}^T(1)\,\boldsymbol{\varphi}(1)\right)^{-1}$

$\boldsymbol{K}(1) = \boldsymbol{P}_0\,\boldsymbol{\varphi}(1)\left(\boldsymbol{I} + \boldsymbol{\varphi}^T(1)\,\boldsymbol{P}_0\,\boldsymbol{\varphi}(1)\right)^{-1}$

$\boldsymbol{\theta}(1) = \boldsymbol{K}(1)\,y(1)$

4- **Incrementar** k **(**$k = k+1$**)**

5- **Calcular** $\boldsymbol{K}(k) = \boldsymbol{P}(k)\,\boldsymbol{\varphi}(k) = \boldsymbol{P}(k-1)\,\boldsymbol{\varphi}(k)\left(\boldsymbol{I} + \boldsymbol{\varphi}^T(k)\,\boldsymbol{P}(k-1)\,\boldsymbol{\varphi}(k)\right)^{-1}$

$\boldsymbol{P}(k) = \left(\boldsymbol{I} - \boldsymbol{K}(k)\,\boldsymbol{\varphi}^T(k)\right)\boldsymbol{P}(k-1)$

$\hat{\boldsymbol{\theta}}(k) = \hat{\boldsymbol{\theta}}(k-1) + \boldsymbol{K}(k)\left(y(k) - \boldsymbol{\varphi}^T(k)\,\hat{\boldsymbol{\theta}}(k-1)\right)$

$\hat{y}(k) = \boldsymbol{\varphi}^T(k)\,\boldsymbol{\theta}(k)$

$|\,y(k) - \hat{y}(k)\,|$

6- **Si** $|\,y(k) - \hat{y}(k)\,| > \alpha$ **incrementar** k **y retornar a 5, de lo contrario terminar.**