Mínimos Cuadrados Recursivos: relación con fórmulas clásicas pseudoinversa

Antonio Sala Piqueras

Dept. Ing. Sistemas y Automatica (DISA)
Universitat Politècnica de València (UPV)

Vídeo-presentación en:
http://personales.upv.es/asala/YT/V/mcmr.html

Introducción, Motivación,Objetivos

Motivación: Los mínimos cuadrados “clásicos” se basan en pseudoinversa. Los recursivos tienen una fórmula aparentemente diferente.

Objetivos: Comprender la relación con los mínimos cuadrados clásicos no recursivos: equivalentes si incertidumbre inicial tiende a infinito.

Contenidos: Relación mínimos cuadrados recursivos y no recursivos: lema de inversión de matrices.

Revisión: Mínimos cuadrados recursivos sin olvido

Disponemos de un estimado previo (media y varianza):

$$\theta = \mathcal{N}(\hat{\theta}_0, \Sigma_{\theta,0})$$

Objetivo: conocida la distribución **a priori** de θ, calcular la mejor predicción de θ a partir de dicha información a priori y una nueva serie de datos (y, X), supuesto modelo $y = X\theta + v$.

Mejor predicción incorporando nueva información (y, X):

*media: $$\hat{\theta}_1 = \hat{\theta}_0 + \Sigma_{\theta,0} X^T (X \Sigma_{\theta,0} X^T + V)^{-1} (y - X\hat{\theta}_0)$$

*varianza: $$\Sigma_{\theta,1} = \Sigma_{\theta,0} - \Sigma_{\theta,0} X^T (X \Sigma_{\theta,0} X^T + V)^{-1} X \Sigma_{\theta,0}$$

Interpretación alternativa: Filtro de Kalman para $\theta_{k+1} = \theta_k$ (ec. estado) junto con $y_k = X_k \theta_k + v$ (ec. salida).

Relación con mínimos cuadrados "clásicos"

Sin información previa, fórmula no recursiva $\hat{\theta} = X^{\dagger} y = (X^T X)^{-1} X^T y$.

Escalando a $V = I$ (normalización), consideramos la ganancia del corrector $K = \Sigma_{\theta,0} X^T (X \Sigma_{\theta,0} X^T + I)^{-1}$. Aplicamos el lema de inversión de matrices: http://www.ee.ic.ac.uk/hp/staff/dmb/matrix/identity.html:

$$K = \Sigma_{\theta,0} X^T (I + X \Sigma_{\theta,0} X^T)^{-1} = (\Sigma_{\theta,0}^{-1} + X^T X)^{-1} X^T$$

*Demostración: $(\Sigma_{\theta,0}^{-1} + X^T X) \Sigma_{\theta,0} X^T = X^T + X^T X \Sigma_{\theta,0} X^T = X^T (I + X \Sigma_{\theta,0} X^T)$

$$\Sigma_{\theta,0} X^T = (\Sigma_{\theta,0}^{-1} + X^T X)^{-1} X^T (I + X \Sigma_{\theta,0} X^T)$$

$$\Sigma_{\theta,0} X^T (I + X \Sigma_{\theta,0} X^T)^{-1} = (\Sigma_{\theta,0}^{-1} + X^T X)^{-1} X^T$$

- Si la varianza inicial $\Sigma_{\theta,0}$ tiende a infinito (desconocimiento total a priori), entonces $\Sigma_{\theta,0}^{-1} \to 0$, y $K \to (X^T X)^{-1} X^T$ es la pseudoinversa usual.

Entonces da igual la media $\hat{\theta}_0$ elegida, porque $\hat{\theta}_1$ no depende de ella:

$$\hat{\theta}_1 = \hat{\theta}_0 + (X^T X)^{-1} X^T (y - X \hat{\theta}_0) = (X^T X)^{-1} X^T y$$

Conclusiones

- Los algoritmos recursivos de mínimos cuadrados incorporan información de nuevos datos junto a a información “a priori”.
- Reducen a mínimos cuadrados no recursivos si varianza inicial a priori fuera infinita.