

Mínimos cuadrados: tres interpretaciones (caso multivariable)

Antonio Sala

Notas de clase sobre control de sistemas multivariables

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

*Presentación en vídeo disponible en:

<http://personales.upv.es/asala/YT/V/ls3im.html>

Presentación

Motivación:

Las fórmulas de mínimos cuadrados pueden ser obtenidas como resultado de una optimización, de unas operaciones con matrices pseudoinversas, de inferencia estadística.

Objetivos:

Comprender las relaciones entre las tres interpretaciones del significado del óptimo mínimo-cuadrático.

Contenidos:

Interpretación determinista (optimización). Interpretación geométrica (proyección ortogonal). Interpretación estadística (mínima correlación, mínima varianza de error). Conclusiones.

I Mínimos cuadrados: interpretación determinista

Objetivo: obtener un modelo que permita predecir $y_{p \times 1}$ a partir de $x_{m \times 1}$, en la forma $y_{predic} = \Theta_{p \times m} \cdot x$, considerando x **medible**. Denotemos error como $e = (y - y_{predic})$.

Mínimos cuadrados con N datos (trasladados a **media cero**):

<minimizar

$$J(\Theta) = \frac{1}{2} \sum_{k=1}^N e_k^T e_k = \frac{1}{2} \sum_{k=1}^N (y_k - \Theta x_k)^T \underbrace{(y_k - \Theta x_k)}_{p \times 1} >$$

Solución del problema de optimización

El problema puede entenderse como p problemas de mínimos cuadrados escalares, o sólo uno, considerando un “vector” de parámetros $\hat{\boldsymbol{\theta}}_{1 \times (p \cdot m)}$, poniendo todas las filas de $\boldsymbol{\Theta}$ una detrás de otra.

$$\text{En efecto, } J(\boldsymbol{\Theta}) = \sum_{k=1}^N \sum_{i=1}^p e_{i,k}^2 = \sum_{i=1}^p \overbrace{\left(\sum_{k=1}^N (y_{i,k} - \boldsymbol{\Theta}^{[i]} x_k)^2 \right)}^{J_i(\boldsymbol{\Theta}^{[i]})}$$

donde $\boldsymbol{\Theta}^{[i]}$ es la fila i de $\boldsymbol{\Theta}$.

Como los parámetros a ajustar en cada $J_i(\boldsymbol{\Theta}^{[i]})$ son diferentes, el mínimo del problema completo es la suma de los mínimos de cada problema individual $J_i(\boldsymbol{\Theta}^{[i]})$.

Haciendo las manipulaciones adecuadas, se tiene:

$$\boldsymbol{\Theta}_{opt, p \times m} = \left(\sum_{k=1}^N y_k x_k^T \right)_{p \times m} \cdot \left(\sum_{k=1}^N x_k x_k^T \right)_{m \times m}^{-1}$$

II Mínimos cuadrados: interpretación geométrica

Si se construye matriz de datos $X_{m \times N}$, vector de salidas $Y_{p \times N}$ tenemos:

$$\left(\sum_{k=1}^N x_k x_k^T \right)_{m \times m} = XX^T, \quad \left(\sum_{k=1}^N y_k x_k^T \right)_{p \times m} = YX^T$$

denotemos la matriz de error $E_{p \times N} := (Y - \Theta X)$, con elementos $e_{i,k}$.

$$J_{1 \times 1} = \sum_{i=1}^p \sum_{k=1}^N e_{i,k}^2 = \|E\|_F^2 = \text{traza}(EE^T)$$

$$\Theta_{opt, p \times m} = YX^T(XX^T)^{-1} = YX^\dagger$$

► **Todas las filas de la matriz E son ortogonales a todas las filas de la aproximación $\hat{Y} = \Theta_{opt}X$, esto es, $E \cdot \hat{Y}^T = 0$**

En efecto:

$$\begin{aligned} (Y - \Theta_{opt}X)_{p \times N} \cdot (X^T \Theta_{opt}^T)_{N \times p} &= (Y - \overbrace{YX^T(XX^T)^{-1}X}^{\Theta_{opt}}) \cdot X^T \overbrace{(XX^T)^{-1}XY^T}^{\Theta_{opt}^T} = \\ YX^T(XX^T)^{-1}XY^T - YX^T(XX^T)^{-1}XX^T(XX^T)^{-1}XY^T &= \mathbf{0}_{p \times p} \end{aligned}$$

II Mínimos cuadrados: interpretación geométrica

► La aproximación $\hat{Y} = \Theta_{opt}X = Y \overbrace{X^T(XX^T)^{-1}X}$ es la “**proyección ortogonal**” de $Y_{1 \times N}$ sobre el subespacio m -dimensional generado por las filas de X .

$Y = \overbrace{(Y - \Theta_{opt}X)}^E + \overbrace{\Theta_{opt}X}^{\hat{Y}}$, expresada como suma de dos matrices ortogonales ($E\hat{Y}^T = 0_{p \times p}$), una matriz $\hat{Y}$ cuyas filas son combinación lineal de filas de X , y otra (E) cuyas filas son perpendiculares las de X .

$$EX^T = (Y - \Theta_{opt}X)X^T = YX^T - (YX^T(XX^T)^{-1}X)X^T = 0$$

Es una “proyección” porque la “proyección de la proyección” ya no cambia nada: $\hat{Y} \cdot X^T(XX^T)^{-1}X = YX^T(XX^T)^{-1}X \cdot X^T(XX^T)^{-1}X = \hat{Y}$

III Mínimos cuadrados: interpretación estadística

Matrices de varianzas-covarianzas relacionadas muestrales, considerando x , y variables incrementales, esto es, restando su media muestral:

$$\Sigma_{xx} = E[xx^T] = \left(\frac{1}{N-1} \sum_{k=1}^N x_k x_k^T \right)_{m \times m} = \frac{1}{N-1} XX^T$$

$$\Sigma_{yx} = E[yx^T] = \left(\frac{1}{N-1} \sum_{k=1}^N y_k x_k^T \right)_{p \times m} = \frac{1}{N-1} YX^T$$

$$\Sigma_{ex} = E[ex^T] = \left(\frac{1}{N-1} \sum_{k=1}^N e_k x_k^T \right)_{p \times m}$$

III Mínimos cuadrados: interpretación estadística

$$\Theta_{opt, p \times m} = YX^T(XX^T)^{-1} = YX^T \frac{1}{N-1} \left(\frac{1}{N-1} \right)^{-1} (XX^T)^{-1} = \Sigma_{yx} \Sigma_{xx}^{-1}$$

Modelo óptimo mínimos cuadrados $\frac{dJ}{d\Theta} = 0 \Leftrightarrow$

$\Sigma_{ex} = 0$, **covarianza** (muestral) entre “error” y “datos entrada” es **cero**.

Predicción óptima en sentido “estadístico” (error no correlado con regresores, expresión en función de matrices de varianza-covarianza) coincide con óptimo “determinista” de mínimos cuadrados (si datos tienen media 0)... y, también volviendo a estadística, con **minimización** de la **variación total** del error $VT := \min_{\Theta} \text{traza}(E(ee^T))$.

Conclusiones

- El problema de **mínimos cuadrados** tiene interpretación **determinista** como **optimización** de un índice de coste (derivadas parciales igual a cero).
Ajuste de parámetros
- También, interpretación **geométrica** como: (a) error ortogonal a aproximación, (b) aproximación es la **proyección ortogonal** de medidas sobre *subespacio de regresores*.
Identificación subespacio
- Por último, interpretación **estadística** como: (a) error **no** tiene **correlación** (muestral) con regresores, (b) minimización de **variación total** del error.

Identificación con ruido, filtro de Kalman, mínima varianza