

0.1. Notas en la obtención de las ecuaciones de Kriging

0.1.1. definiciones básicas

Consideramos un campo continuo e invariante en el tiempo,

$$\phi : x \in D \rightarrow \phi(x) \in \mathbb{R} : D \subset \mathbb{R}^m$$

Modelamos ϕ como un proceso gaussiano,

$$\Omega(x) = r(x)^T \beta + Z(x), \quad (1)$$

Donde r es un vector de funciones de regresión conocidas, $r(x) = (r_1(x), r_2(x), \dots, r_k(x))^T$ y $\beta = (\beta_1, \beta_2, \dots, \beta_k)$ es un vector de parámetros, también conocidos.

$Z(x)$ es un proceso gaussiano de media zero, cuya función de covarianza modelamos como,

$$\begin{aligned} \text{cov}(Z(x), Z(x')) &= \sigma_z^2 \xi(x, x') \\ \xi(x, x') &= \exp \left[- \sum_i^m \frac{1}{\theta_i} |x_i - x'_i|^{q_i} \right] \end{aligned} \quad (2)$$

Donde x_i es la componente i del vector x . Supongamos que tenemos un conjunto de n medidas ruidosas, de nuestro campo $y = [y(x^{(1)}), y(x^{(2)}), \dots, y(x^{(n)})]$. Nuestras medidas son una realización del proceso gaussiano,

$$Y(x) = R(x)\beta + \mathcal{Z} \quad (3)$$

Donde,

$$R = \begin{pmatrix} r_1(x^{(1)}) & \dots & r_k(x^{(1)}) \\ \vdots & \ddots & \vdots \\ r_1(x^{(n)}) & \dots & r_k(x^{(n)}) \end{pmatrix} \quad (4)$$

con $\mathcal{Z} = (Z(x^{(1)}), \dots, Z(x^{(n)}))^T$

Supongamos que queremos construir un predictor lineal $Y = a_x^T y$ a partir de las medidas. Es decir yo quiero inferir cual sería el valor de mi campo en un punto x , supuesto que conozco los valores en los puntos medidos. De alguna manera, lo que estoy definiendo es una media ponderada de los valores que he medido.

Podemos definir entonces el error cuadrático medio a partir del valor esperado de la diferencia entre las medidas y el modelo gaussiano,

$$\begin{aligned} E[(a_x^T y - Y(x))^2] &= E[a_x^T y y^T a_x + Y(x)^2 - 2a_x^T y Y(x)] = \\ &= E[a_x^T (R\beta + \mathcal{Z})((R\beta + \mathcal{Z})^T a_x) + \\ &= E[(r^T(x)\beta + Z(x))^2] + \\ &= E[-2a_x^T (R\beta + \mathcal{Z})(r^T(x)\beta + Z(x))] \end{aligned} \quad (5)$$

Desarrollamos por separado cada uno de los tres sumandos de la ecuación (5),

1.

$$\begin{aligned} E[a_x^T(R\beta + \mathcal{Z})(R\beta + \mathcal{Z})^T a_x] &= \\ a_x^T R\beta\beta^T R^T a_x + 2a_x^T R\beta E[\mathcal{Z}^T] a_x + a_x^T E[\mathcal{Z}\mathcal{Z}^T] a_x &= \\ = a_x^T R\beta\beta^T R^T a_x + a_x^T \sigma_z^2 K a_x \end{aligned} \quad (6)$$

Donde hemos hecho uso de que el valor esperado del vector \mathcal{Z} tiene que ser nulo. Y además $\mathcal{Z}\mathcal{Z}^T$ Es una matriz de covarianza que podemos definir a partir de la ecuación 2 como $\sigma_z K$ con $K_{i,j} = \xi(x^{(i)}, x^{(j)})$.

2.

$$\begin{aligned} E[(r^T(x)\beta + Z(x))^2] &= \\ r(x)^T \beta\beta^T r(x) + 2r(x)^T \beta E[Z(x)] + E[Z(x)^2] &= \\ = r^T \beta\beta^T r + \sigma_z^2 \end{aligned} \quad (7)$$

De nuevo, ecuación (2), $E[Z(x)] = 0$ y $\text{cov}(Z(x), Z(x)) = \sigma_z^2$

3.

$$\begin{aligned} E[-2a_x^T(R\beta + \mathcal{Z})(r^T(x)\beta + Z(x))] &= \\ -2a_x^T R\beta r(x)^T \beta - 2a_x^T R\beta E[Z(x)] - 2a_x^T E[\mathcal{Z}]r(x)^T \beta + -2a_x^T E[\mathcal{Z}Z(x)] &= \\ = -2a_x^T R\beta r(x)^T \beta - 2a_x^T \sigma_z k_x \end{aligned} \quad (8)$$

Donde, $k_x = \sigma_z^2(\xi(x, x_1), \xi(x, x_2), \dots, \xi(x, x_n))^T$ y las esperanzas de Z y \mathcal{Z} son cero.

Llegamos por tanto a que,

$$\begin{aligned} E[(a_x^T y - Y(x))^2] &= \\ a_x^T R\beta\beta^T R^T a_x + a_x^T \sigma_z^2 K a_x + r(x)^T \beta\beta^T r(x) + \sigma_z^2 - 2a_x^T R\beta r(x)^T \beta - 2a_x^T \sigma_z k_x &= \\ = (a_x^T R\beta - r(x)^T \beta)^2 + a_x^T \sigma_z^2 K a_x + \sigma_z^2 - 2a_x^T \sigma_z^2 k_x \end{aligned} \quad (9)$$

A nosotros nos gustaría minimizar esta expresión, ya que eso es tanto como minimizar el error esperado. Si nos fijamos en el primer término cuadrático, si nuestro ajuste es bueno, deberíamos hacer este término cero: $R^T a_x = r(x)$, Si lo hacemos así, estamos construyendo un estimador sin sesgo. $Y = a_x^T y = a_x^T R\beta$ donde Y es el valor de la variable estocástica que queremos predecir $Y = r(x)^T \beta$.

Por tanto deberíamos minimizar el resto de expresión,

$$\begin{aligned} \min_{a_x} [a_x^T \sigma_z^2 K a_x + \sigma_z^2 - 2a_x^T \sigma_z^2 k_x] \\ \text{sujeto a:} \\ R^T a_x = 0 \end{aligned} \quad (10)$$

Si derivamos con respecto a a^T , igualamos a cero y añandimos un multiplicador de Lagranje $-\lambda R$ a la expresión derivada, obtenemos:

$$\begin{aligned} K a_x - k_x - R\lambda &= 0 \\ R^T a_x &= r(x) \end{aligned} \quad (11)$$

Podemos describir las ecuaciones en forma matricial,

$$\begin{pmatrix} K & R \\ R^T & 0 \end{pmatrix} \begin{pmatrix} a_x \\ -\lambda \end{pmatrix} = \begin{pmatrix} k_x \\ r(x) \end{pmatrix} \quad (12)$$

$$\begin{pmatrix} a_x \\ -\lambda \end{pmatrix} = \begin{pmatrix} K & R \\ R^T & 0 \end{pmatrix}^{-1} \begin{pmatrix} k_x \\ r(x) \end{pmatrix} \quad (13)$$

Podemos calcular la inversa de esta matriz, empleando los bonitos complementos de Shür y la eliminación de Gauss-Jordan por bloques¹,

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \quad M^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B(D - A^{-1}BC)^{-1}CA^{-1} & -A^{-1}B(D - A^{-1}BC)^{-1} \\ -(D - A^{-1}BC)^{-1}CA^{-1} & (D - A^{-1}BC)^{-1} \end{bmatrix}$$

Si lo aplicamos a nuestro sistema,

$$\begin{pmatrix} a_x \\ -\lambda \end{pmatrix} = \begin{pmatrix} K^{-1} - K^{-1}R(K^{-1}RR^T)^{-1}R^TK^{-1} & K^{-1}R(K^{-1}RR^T)^{-1} \\ (K^{-1}RR^T)^{-1}R^TK^{-1} & -(K^{-1}RR^T)^{-1} \end{pmatrix}^{-1} \begin{pmatrix} k_x \\ r(x) \end{pmatrix} \quad (14)$$

Con lo que obtenemos para a_x ,

$$a_x = (K^{-1} - K^{-1}R(K^{-1}RR^T)^{-1}R^TK^{-1})k_x + K^{-1}R(K^{-1}RR^T)^{-1}r(x) \quad (15)$$

Luego nuestra estimación del valor de $Y(x)$ sería,

$$\begin{aligned} Y(x) &= a_x^T y = \\ &[k_x^T(K^{-1} - K^{-1}R(K^{-1}RR^T)^{-1}R^TK^{-1}) + r(x)^T(R^TK^{-1})^{-1}R^TK^{-1}]y = \\ &k_x^TK^{-1}(y - R\beta) + r(x)^T\beta \end{aligned} \quad (16)$$

Que es precisamente la ecuación (18) del artículo con,

$$\beta = (R^TK^{-1})^{-1}R^TK^{-1} \quad (17)$$

En realidad, si somos rigurosos, el valor de β que estamos empleando es una estimación, y deberíamos llamarle $\hat{\beta}$, siguiendo una notación que es más o menos estándar, para dejar claro que es un valor estimado.

Lo que hemos obtenido es el valor esperado para el modelo lineal que suponemos siguen nuestros predictores. Nos faltaría calcular la varianza. En realidad, coincide con el valor cuadrático medio que hemos optimizado,

$$\begin{aligned} \sigma_\phi^2(x) &= E[(a_x^Ty - Y(x))^2] = \\ &E[(k_x^TK^{-1}(y - R\beta) + r(x)^T\beta - r(x)^T\beta - Z(x))^2] = \\ &E[(k_x^TK^{-1}(y - R\beta) + Z(x))^2] = \\ &E[(k_x^TK^{-1}(R\beta + \mathcal{Z} - R\beta) - Z(x))^2] = \\ &E[(k_x^TK^{-1}\mathcal{Z} - Z(x))^2] \end{aligned} \quad (18)$$

Si desarrollamos el cuadrado de la última expresión de la ecuación (18).

¹https://en.wikipedia.org/wiki/Schur_complement

$$\begin{aligned}
\sigma_\phi^2(x) &= E[(k_x^T K^{-1} \mathcal{Z} + Z(x))^2] = \\
E [k_x^T K^{-1} \mathcal{Z} \mathcal{Z}^T K^{-1} k_x - 2k_x^T K^{-1} \mathcal{Z} Z(x) + Z(x)^2] &= \\
k_x^T K^{-1} E[\mathcal{Z} \mathcal{Z}^T] k_x - 2k_x^T K^{-1} \sigma_z k_x + \sigma_z^2 &= \\
k_x^T K^{-1} \sigma_z K K^{-1} k_x - 2k_x^T K^{-1} \sigma_z k_x + \sigma_z^2 &= \\
&= (k_x^T K^{-1} k_x - 1) \sigma_z^2
\end{aligned} \tag{19}$$