Recapitulación

Estimación de Funciones mediante SVM

Consideramos un conjunto de datos $\{X_i, Y_i\}$, donde los X_i corresponden a los inputs y Y_i a los outputs. Queremos aplicar el método de Support Vector Machine para estimar funciones. Para ello consideramos

$$\begin{gathered} F : \mathbb{R}^n \to \mathbb{R} \\ F(X) = W^T \varphi(X) + B \end{gathered} \tag{1}$$

donde

- $\varphi : \mathbb{R}^n \to \mathbb{R}^{n_h}$ mapeo no lineal con $n \leq n_h$
- $W \in \mathbb{R}^{n_h}$
- $B \in \mathbb{R}$

Buscamos minimizar el riesgo empírico.

$$R_{\text{emp}}(W, B) = \frac{1}{N} \sum_{i=1}^{N} |Y_i - F(X_i)| = \frac{1}{N} \sum_{i=1}^{N} \left| Y_i - W^T \varphi\left(X_i\right) - B \right| \tag{2}$$

Problema de control discretizado en N pasos

Consideramos el siguiente problema de control óptimo

$$\begin{aligned} \min_{u} \quad & J_N(x, u) = \rho\left(x_N\right) + \sum_{i=1}^{N-1} h\left(x_i, u_i\right) \\ \text{sujeto a} \quad & x_{i+1} = f\left(x_i, u_i\right), \quad i = 1, \ldots, N-1, \\ & x_1 \text{ dado.} \end{aligned} \tag{3}$$

donde $\rho(\cdot)$ y $h(\cdot)$ son definidas positivas.

En este proyecto buscamos resolver problemas de este tipo, empleando el método de LS-SVM para estimar los controles.

Formulación

Queremos incorporar la metodología de LS-SVM. Para ello, basándonos en el mecanismo de estimación de funciones usando SVM y el SRM (Structural Risk Minimization), podemos escribir el modelo de mínimos cuadrados como:

$$(P) \qquad \min_{W,B,e} \mathscr{J}_{\mathrm{LS}}(W,B,e) = \frac{1}{2}W^{\mathrm{T}}W + \gamma\frac{1}{2}\sum_{i=1}^{N} e_i^2$$

sujeto a las restricciones de igualdad

$$Y_i = W^{\mathrm{T}}\varphi\left(X_i\right) + B + e_i, \quad i = 1,\ldots,N.$$

donde se ponderan los términos $W^{\mathrm{T}}W$, asociado a la regularización de los parámetros (*no sé bien por qué no se minimiza también sobre el valor de* B), y $\sum_{k=1}^{N} e_k^2$ correspondiente a la suma de los errores de estimación.

Dado que buscamos incluir esta metodología en (P), podemos plantear un nuevo problema de optimización considerando tanto el funcional $J_N(\cdot,\cdot)$, propio del problema (P), como la función de costo del problema de mínimos cuadrados.

$$\begin{aligned}
(\bar{P}) \qquad \min \quad & J_N(x,u) + \frac{1}{2}W^{\mathrm{T}}W + \gamma\frac{1}{2}\sum_{i=1}^{N} e_i^2 \\
\text{sujeto a} \quad & x_{i+1} = f(x_i,u_i), \quad i = 1,\ldots,N-1, \quad (x_1 \text{ dado}) \\
& u_i = W^{\mathrm{T}}\varphi(x_i) + e_i, \quad i = 1,\ldots,N \quad (\text{Notar que no incluye B})
\end{aligned} \tag{4}$$

Esto ya que consideramos $\{x_i,u_i\}_{i=1}^{N} \sim \{X_i,Y_i\}_{i=1}^{N}$ como el conjunto de datos de entrenamiento y usamos la estimación de funciones mediante SVM.

Para resolver este problema debemos plantear su lagrangiano.

$$\begin{aligned}
&\mathscr{L}\left(x_k,u_k,w,e_k,\lambda_k,\alpha_k\right) = \\
&\mathscr{J}_N\left(x_k,u_k\right) + \frac{1}{2}W^{\mathrm{T}}W + \gamma\frac{1}{2}\sum_{k=1}^{N} e_k^2 + \sum_{k=1}^{N}\lambda_k^{\mathrm{T}}\left[x_{k+1} - f\left(x_k,u_k\right)\right] + \sum_{k=1}^{N}\alpha_k\left[u_k - w^{\mathrm{T}}\varphi\left(x_k\right) - e_k\right].
\end{aligned}$$

Y las condiciones de optimalidad están dadas por

$$\begin{cases} \frac{\partial \mathscr{L}}{\partial x_k} = \frac{\partial h}{\partial x_k} + \lambda_{k-1} - \left(\frac{\partial f}{\partial x_k}\right)^{\mathrm{T}} \lambda_k - \alpha_k \frac{\partial}{\partial x_k}\left[W^{\mathrm{T}}\varphi\left(x_k\right)\right] = 0, & k = 2, \ldots, N, \\ \frac{\partial \mathscr{L}}{\partial x_{N+1}} = \frac{\partial \rho}{\partial x_{N+1}} + \lambda_N = 0, & \\ \frac{\partial \mathscr{L}}{\partial u_k} = \frac{\partial h}{\partial u_k} - \lambda_k^{\mathrm{T}} \frac{\partial f}{\partial u_k} + \alpha_k = 0, & \\ \frac{\partial \mathscr{L}}{\partial W} = W - \sum_{k=1}^{N} \alpha_k \varphi\left(x_k\right) = 0, & \\ \frac{\partial \mathscr{L}}{\partial e_k} = \gamma e_k - \alpha_k = 0, & k = 1, \ldots, N, \\ \frac{\partial \mathscr{L}}{\partial \lambda_k} = x_{k+1} - f\left(x_k, u_k\right) = 0, & k = 1, \ldots, N, \\ \frac{\partial \mathscr{L}}{\partial \alpha_k} = u_k - W^{\mathrm{T}}\varphi\left(x_k\right) - e_k = 0. & \end{cases} \tag{6}$$

De la cuarta ecuación obtenemos que $W = \sum_{k=1}^{N} \alpha_k \varphi\left(x_k\right)$ y al reemplazarlo en la séptima ecuación se tiene que

$$u_k = W^{\mathrm{T}}\varphi\left(x_k\right) + e_k = \sum_{l=1}^{N} \alpha_l \varphi^{\mathrm{T}}(x_l)\varphi\left(x_k\right) + e_k$$

Luego, siguiendo la formulación del apartado 5 del paper de referencia, tomamos $\gamma \to \infty$, lo cual implica, de la quinta ecuación, que $e_k \to 0$. Por lo cual, el control puede ser escrito como

$$u_k = W^{\mathrm{T}}\varphi\left(x_k\right) = \sum_{l=1}^{N} \alpha_l \varphi^{\mathrm{T}}(x_l)\varphi\left(x_k\right) = \sum_{l=1}^{N} \alpha_l K(x_l, x_k) \tag{7}$$

Esta representación de u_k nos permite reformular el problema (P) inicial pues hemos mostrado que al usar LS-SVM aplicado al problema de control óptimo discretizado en N pasos, el control óptimo queda caracterizado por los coeficientes α_k y el kernel $K(x_l, x_k)$.

De esta forma, podemos reescribir el problema de control óptimo en N pasos considerando la expresión del control obtenida anteriormente.

$$\begin{aligned} \min_{x,\alpha} \quad & J_N(x,u) + \lambda \sum_{i=1}^{N-1} \alpha_i^2 \\ \text{s.a.} \quad & x_{i+1} = f(x_i, u_i), \quad i = 1, \ldots, N-1 \ (x_1 \text{ dado.}) \\ & u_i = \sum_{l=1}^{N} \alpha_l K(x_l, x_i), \quad i = 1, \ldots, N-1. \end{aligned} \tag{8}$$

donde se incorpora el término $\lambda \sum_{i=1}^{N-1} \alpha_i^2$ asociado a la regularización de los parámetros $(\alpha)_{i=1}^{N}$, el cual penaliza valores grandes, manteniéndolos así cercanos a 0.

A diferencia de otras formulaciones presentadas en el paper, para resolver este problema no se utilizan las ecuaciones adicionales del Lagrangiano. En su lugar, se aborda el problema de optimización considerando únicamente x y α como variables. Esto reduce significativamente el número de variables a determinar, simplificando así el sistema a resolver.