Universidad de Santiago de Chile
Facultad de Ingeniería
Depto. de Ingeniería Informática

Taller de minería de datos avanzada
Capítulo VI
"SVM Regresión y recurrencia"

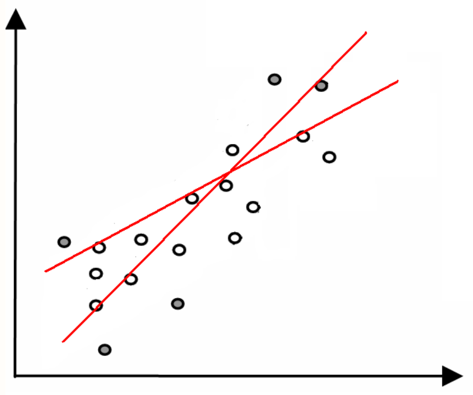
Profesor: Dr. Max Chacón

Objetivos

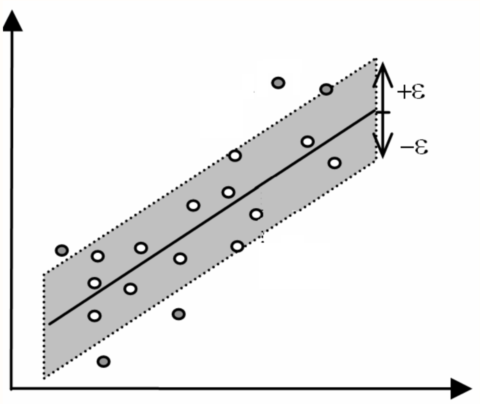
- Introducción
- El problema de la recta de regresión
- El problema de Support Vector Regression (SVR)
- El problema de modelamiento temporal
- Recurrencias externas.

6.1. Support Vector Regression

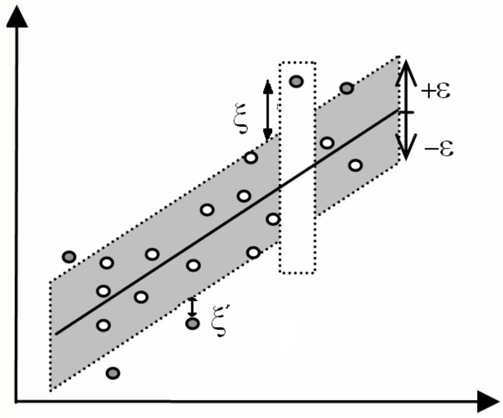
Supongamos que se requiere buscar la mejor recta que ajusta a una nube de puntos $f(\vec{x}) = \vec{\omega}^* \bullet \vec{x} + b$ el problema es encontrar los coeficientes $\vec{\omega}$ que determinan la ecuación de la recta.



Existirá un conjunto de datos cerca de la recta que no influyen en su posición, pero los datos que se encuentran a una distancia > ε controlaran mejor su ángulo.

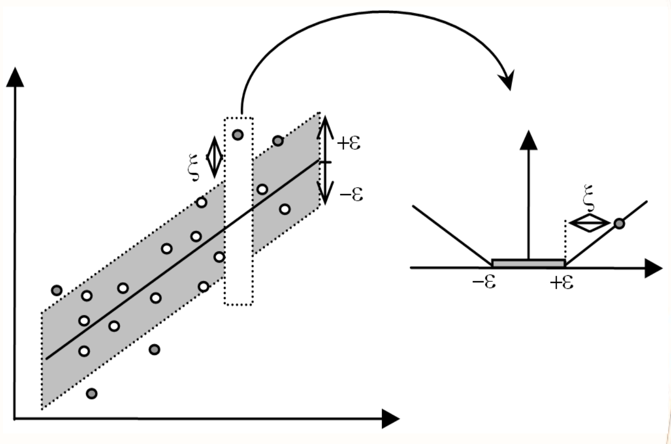


A medida que los datos se alejan cobran más importancia para determinar la posición de la recta.



Seria bueno contar con una ponderación que aumente la importancia de los casos a medida que se alejan.

Función ε-insensitiva lineal



Elimina todos los datos en un tubo de radio menor a ε y **pondera** linealmente los casos afuera del tubo de radio ε

6.2. El problema de optimización de SVR

Se requiere determinar $\vec{\omega}$ que genere una recta que tenga la mínima inclinación posible. Lo más parecida al modelo que se ajusta al promedio de los datos (minimizar norma de ω) .

Se requiere que la desviación entre $f(\vec{x}_i)$ e y_i sea de distancia máxima ε.

Además es necesario ponderar los datos que quedan fuera del tubo, al igual que los vectores acotados de SVM

$$Minimizar \quad \frac{1}{2}\|\varpi\|^2 + C\sum_{j=1}^{n}(\xi_j - \xi_j^*)$$

$$s.a. \quad \begin{aligned} & y_i - (\varpi \bullet \vec{x}_i) - b \le \varepsilon + \xi_i \\ & (\varpi \bullet \vec{x}_i) + b - y_i \le \varepsilon + \xi_i^* \\ & \xi_j, \xi_j^* \ge 0 \end{aligned}$$

- El Lagrangeano aumentado será:

$$\begin{aligned} L(\varpi,\xi_i,\xi_i^*,\lambda_i,\lambda_i^*,\beta_i,\beta_i^*,b) = & \frac{1}{2}\|\varpi\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \\ & - \sum_{i=1}^{n}\lambda_i(\varepsilon + \xi_i - y_i + (\varpi \bullet \vec{x}_i) + b) \\ & - \sum_{i=1}^{n}\lambda_i(\varepsilon + \xi_i^* + y_i - (\varpi \bullet \vec{x}_i) - b) \\ & - \sum_{i=1}^{n}(\beta_i\xi_i + \beta_i^*\xi_i^*) \end{aligned} \quad (i)$$

$\varpi, \xi_i, \xi_i^*, b$ Variables del primal

$\lambda_i, \lambda_i^*, \beta_i, \beta_i^*$ Multiplicadores de Lagrange (duales)

Derivando el Lagrangeano e igualando a cero, se esta en el punto silla que permite obtener las ecuaciones en el espacio dual.

$$\frac{\partial L}{\partial \omega} = \varpi + \sum_{i=1}^{n} \lambda_i \vec{x}_i + \sum_{i=1}^{n} \lambda_i^* \vec{x}_i = 0 \qquad (ii)$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{n} (\lambda_i - \lambda_i^*) = 0 \qquad (iii)$$

$$\frac{\partial L}{\partial \xi_i} = C - \lambda_i - \beta_i^* = 0 \qquad (iv)$$

$$\frac{\partial L}{\partial \xi_i^*} = C - \lambda_i^* - \beta_i^* = 0 \qquad (v)$$

Sustituyendo *ii a v* en *i*, se tiene la función objetivo en el espacio dual.

Problema dual

Máximizar: $\theta(\lambda_i, \lambda_i^*) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (\lambda_i - \lambda_i^*)(\lambda_j - \lambda_j^*)(\vec{x}_i \bullet \vec{x}_j)$

$$- \varepsilon \sum_{i=1}^{n} (\lambda_i - \lambda_i^*) + \sum_{i=1}^{n} y_i (\lambda_i - \lambda_i^*)$$

$$\text{s.a.} \quad \sum_{i=1}^{n} (\lambda_i - \lambda_i^*) = 0; \quad 0 \le \lambda_i \le C; \; 0 \le \lambda_i^* \le C$$

La solución óptima esta dada por la multiplicación de las variables duales y las restricciones del problema original (KKT).

$$\lambda_i (\varepsilon + \xi_i - y_i + (\varpi \bullet \vec{x}_i) + b) = 0 \qquad (vi)$$

$$\lambda_i^* (\varepsilon + \xi_i^* + y_i - (\varpi \bullet \vec{x}_i) - b) = 0$$

$$\beta_i \xi_i = (C - \lambda_i) \xi_i = 0 \qquad (vii)$$

$$\beta_i \xi_i^* = (C - \lambda_i^*) \xi_i^* = 0$$

De las ultimas dos ecuaciones *(vii)* se puede concluir que el ξ_i asociada a una muestra $(\vec{x}_i, y_i)$ sólo puede ser distinta de cero si su λ_i es igual a C.

De las dos primeras ecuaciones *(vi)*, se puede obtener los valores ***b***, puesto que si ***0<λi<C***, la variable ξ_i será cero, entonces.

$$b = y_i - (\vec{\omega} \bullet \vec{x}_i) - \varepsilon \quad \forall \lambda_i \in \left]0, C\right[$$
$$b = y_i - (\vec{\omega} \bullet \vec{x}_i) + \varepsilon \quad \forall \lambda_i^* \in \left]0, C\right[$$

También se observa (de *vi*) que no pueden existir dos variables duales λ_i y λ_i^* que sean simultáneamente distintas de cero.

De *(vi)* se puede ver que para distancias de $f(x)<\varepsilon$ el segundo factor es $\neq 0$, por lo tanto $\lambda_i = 0$.

En el caso contrario de (*vi*) se puede ver que para distancias de $f(x)>\varepsilon$ el segundo factor es $= 0$, por lo tanto pueden ser $\lambda i > 0$.

Lo anterior implica que sólo los $\lambda i > 0$ son los que contribuyen a la solución del problema y no se necesitan todas las $i=1\ldots n$, muestras para describir el problema.

Por lo tanto la ecuación de la recta original puede ser reescrita de la siguiente manara

$$f(\vec{x}) = \vec{\omega}^* \bullet \vec{x} + b = \sum_{i=1}^{l} (\lambda_i - \lambda_i^*)(\vec{x}_i \bullet \vec{x}) + b$$

Donde *l* representa los casos con $\lambda i > 0$, ***Vectores Soporte***

6.3. Tiempo y Recurrencias externas

Identificación de Sistemas (IS)

Metodología que consiste en construir modelos matemáticos del sistema estudiado.

Reglas:

- Obtención de datos de entrada y salida del sistema que se desea identificar.

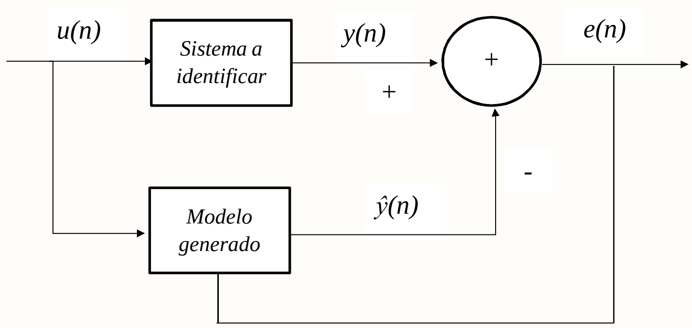
- Pre-procesamiento de los datos de entrada y salida. (Examinar, limpiar y filtrar los datos).

- Elección del conjunto de modelos.

- Obtención de los parámetros del modelo y validación.

-Si modelo tiene buenos resultados (alcanzo un error aceptable), terminar sino realizar cambios en punto anterior e iterar.

Estructura general de la Identificación de sistemas



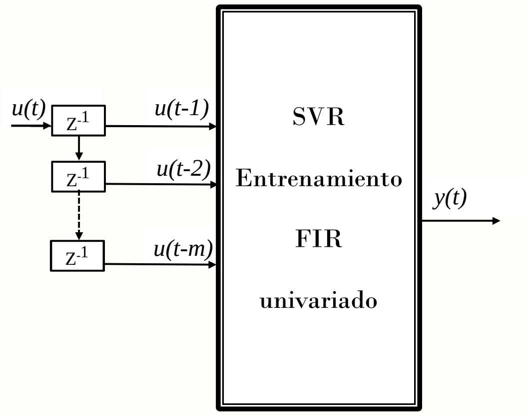
$e(n)= y(n)- \hat{y}(n)$

- $\hat{y}(n)$ la salida del modelo generado y ***e*** el error. La identificación de los parámetros de los modelos para Sistemas en el tiempo se basa en el error de predicción $e(n)= y(n)- \hat{y}(n)$

Sistemas con recurrencias externas en SVM

- Modelos FIR (o NFIR)

- Respuesta finita al impulso.
- Solo poseen memorias las entradas.
- Su estructura de aprendizaje es la siguiente



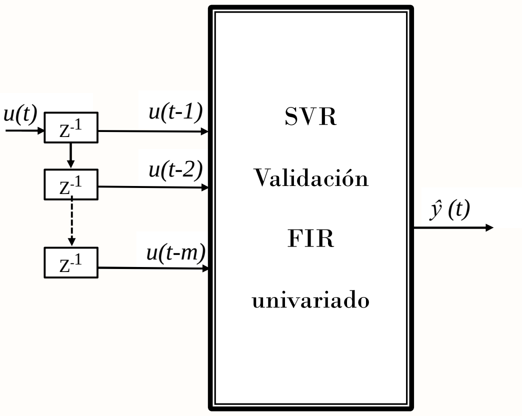
Donde el parámetro de retardos (***m***) de las entradas también es un parámetro que se debe ajustar por el método de grilla.

Entonces para un modelo FIR se requiere ajustar 3 parámetros para el caso lineal y 4 para el caso no-lineal.

C, ε, ***m*** FIR y C, ε, ***m*** y γ NFIR.

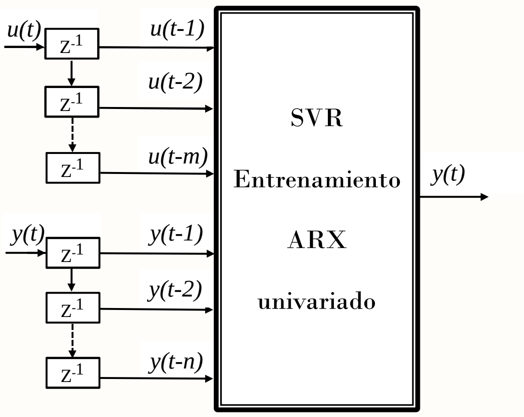
Existen algunas implementaciones, donde el ancho del tubo de la sección ε-insensitiva se normaliza, para ser una proporción de los vectores de soporte.
En ese caso el parámetro se denomina ν y varia de 0 a 1, en general se estima con pasos de 0,1.

El modelo para estimar una nueva salida (producción) es el siguiente:



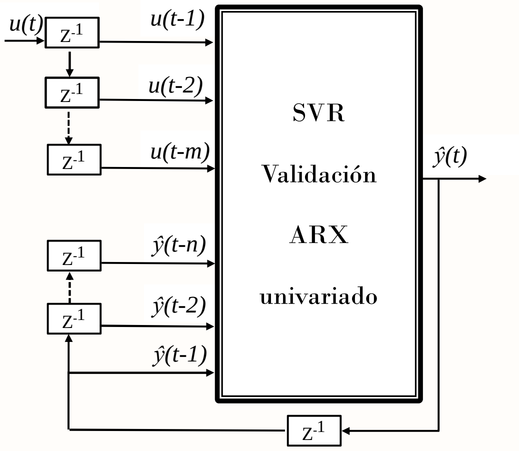
- Modelos ARX (o NARX)

Su estructura de entrenamiento es:



Esta estructura es operada One-Step-Ahead (OSA)

El modelo para estimar una nueva salida (producción) es el siguiente:



Esta estructura es operada Multiple-Prediction-Output (MPO)

Entonces para un modelo ARX se requiere ajustar 4 parámetros para el caso lineal y 5 para el caso no lineal NARX.

C, ε, ***m*** y ***n*** ARX y C, ε, ***m***, ***n*** y γ NARX.

O el caso de donde ε es remplazado por ν