Regresión Funcional

CC52A - Inteligencia Artificial

Gonzalo Ríos D.

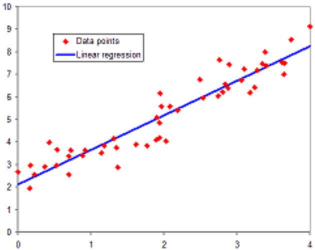
DCC - UChile

Otoño 2010

Definición

Dado un conjunto de puntos $D{=}((\vec{x}_1, y_1), \ldots, (\vec{x}_m, y_m))$, *donde* $\vec{x}_i = (x_{i1}, \ldots, x_{in})$, *un espacio de funciones* F *y una función de error* $E : F \times \mathbb{R}^{m*(n+1)} \to \mathbb{R}$, *deseamos encontrar la función* $\bar{f} \in F$, *llamada regresor, tal que*

$$\bar{f} = \min \arg E(f, D)$$



Proposición

Las funciones de error mas comunes son:

- *Error cuadrático:* $E(f, D) = \sum_{i=1}^{m} (f(\vec{x}_i) - y_i)^2$
- *Error relativo:* $E(f, D) = \sum_{i=1}^{m} \frac{(f(\vec{x}_i) - y_i)^2}{y_i^2}$
- *Error con pesos:* $E(f, D) = \sum_{i=1}^{m} wi * (f(\vec{x}_i) - y_i)^2$
- *Error con tolerancia:*
  $$E(f, D) = \sum_{i=1}^{m} \begin{cases} (f(\vec{x}_i) - y_i)^2 - d^2 & si \quad (f(\vec{x}_i) - y_i)^2 \geq d^2 \\ 0 & si \quad (f(\vec{x}_i) - y_i)^2 \leq d^2 \end{cases}$$

La función de error tendrá de forma implícita la "geometría" del espacio, es decir, la distancia en dicho espacio

Proposición

Los espacios de funciones más comunes son:

- *Lineales:* $F = \{a + bx | a, b \in \mathbb{R}\}$
- *Polinomial:* $F = \{a_0 + a_1 x + ... + a_n x^n | a_0, ..., a_n \in \mathbb{R}\}$
- *Exponenciales:* $F = \{e^{a_0 + a_1 x + ... + a_n x^n} | a_0, ..., a_n \in \mathbb{R}\}$
- *Racionales:* $F = \{\frac{a_0 + a_1 x + ... + a_n x^n}{b_0 + b_1 x + ... + b_m x^m} | a_0, ..., a_n, b_0, ..., b_m \in \mathbb{R}\}$
- *Trigonométricos:* $F = \{a_0 + a_1 \sin(\frac{2\pi x}{T}) + ... + a_n \sin(\frac{2\pi n x}{T}) + b_1 \cos(\frac{2\pi x}{T}) + ... + b_n \cos(\frac{2\pi n x}{T}) | a_0, ..., a_n, b_1, ..., b_n, T \in \mathbb{R}\}$

El espacio de funciones dirá la "forma" apriori de la función a buscar.

El problema de regresión funcional es un problema de optimización, pero en un espacio de funciones.

Proposición

Si el espacio de funciones es paramétrico, es decir, cada función del espacio es generado por una cierta cantidad de parámetros $a_1, ..., a_p$, *entonces existe una función* $\tilde{E} : \mathbb{R}^p \times \mathbb{R}^{m*(n+1)} \longrightarrow \mathbb{R}$ *tal que*

$$E(f, D) = \tilde{E}(a_1, ..., a_p, D)$$

Proposición

Si $\tilde{E}$ *es derivable en los parámetros* $a_1, ..., a_p$, *entonces*

$$\begin{aligned} F(\bar{a}_1, ..., \bar{a}_p) &= \bar{f} \\ \nabla \tilde{E}(\bar{a}_1, ..., \bar{a}_p, D) &= \vec{0} \end{aligned}$$

Proposición

Las siguientes recurrencias son iteraciones de optimización:

- *Gradiente:* $\vec{a}_{n+1} = \vec{a}_n - \alpha_n * \nabla E(\vec{a}_n, D)$
- *Newton:* $\vec{a}_{n+1} = \vec{a}_n - \nabla^2 E(\vec{a}_n, D)^{-1} * \nabla E(\vec{a}_n, D)$
- *Cuasi-Newton:* $\vec{a}_{n+1} = \vec{a}_n - H_n * \nabla E(\vec{a}_n, D)$, $H_n \approx \nabla^2 E(\vec{a}_n, D)^{-1}$

Existe una gran lista de métodos de optimización no lineal, pero nos enfocaremos solo en casos particulares, sin preocuparnos de los criterios de convergencia ni condiciones de optimalidad, ya que son temas que se escapan del temario del curso.

Definición

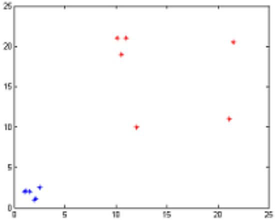
Sea un conjunto de datos $D = ((\vec{x}_1, c_1), ..., (\vec{x}_m, c_m))$, *donde* $c_i \in C = \{-1, 1\}$, *conjunto de clases binarias. Deseamos encontrar una función* $g : \mathbb{R}^n \rightarrow \mathbb{R}$ *tal que*

$$c_i = \begin{cases} 1 & si \quad g(\vec{x}_i) \geq 0 \\ -1 & si \quad g(\vec{x}_i) < 0 \end{cases}$$

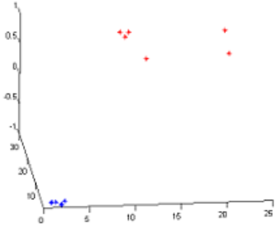
Proposición

Si las clases son linealmente separables, es decir, es posible trazar una linea entre las clases, entonces

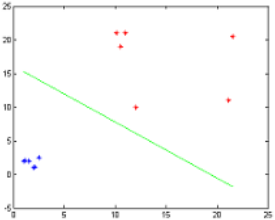
$$\begin{aligned} g(\vec{x}) &= a + \vec{b} \cdot \vec{x} \\ (a, \vec{b}) &= \arg\min \sum_{i=1}^{m} (a + \vec{b} \cdot \vec{x}_i - c_i)^2 \end{aligned}$$



25
20
15
10
5
0
0
5
10
15
20
25



1
0.5
0
-0.5
-1
30
20
10
0
5
10
15
20
25

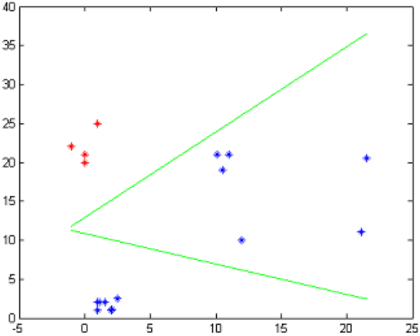


25
20
15
10
5
0
-5
0
5
10
15
20
25

Proposición

En el caso que $C = \{c_1, .., c_n\}$ y todas las clases sean linealmente separables entre si, entonces el siguiente algoritmo separa las n clases:

- *Para k=1...n-1*
    - *aplicar el algoritmo binario para $C^1 = c_k$ y $C^2 = [c_{k+1}, .., c_n]$*
    - *obtener la recta g_k, que separa la clase c_k de las demas clases*
- *End*
- $g(\vec{x}) = \begin{cases} c_1 & si & g_1(x) \geq 0 \\ c_2 & si & g_1(x) < 0 \wedge g_2(x) \geq 0 \\ \dots & \dots & \dots \\ c_n & si & g_1(x) < 0 \wedge g_2(x) < 0 \wedge \dots \wedge g_{n-1} < 0 \end{cases}$



Esto se puede generalizar a cualquier otro espacio de funciones y curvas. ¿Qué sucede si cambiamos la función lineal por $g\left(\vec{x}\right) = \frac{2}{1+\exp^{a+\vec{b}\cdot\vec{x}}} - 1$?

En el caso que $C = \{c_1, .., c_n\}$ no son linealmente separables entre si, entonces el algoritmo anterior falla, por lo que debemos buscar otra opción. Un enfoque es la noción de funciones discriminantes.

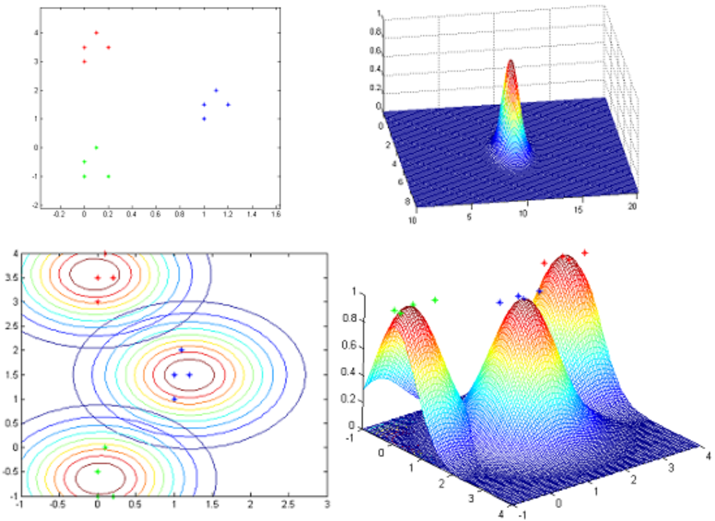
Definición

Dadas las clases $C = \{c_1, .., c_n\}$, las funciones $\{g_1, ..., g_n\}$ son funciones discriminantes si $g(x) = \arg\max_i g_i(x)$ discrimina las clases, es decir, $g(x) = i$ si $c(x) = c_i$.

Proposición

Una función de error para el problema anterior, con el espacio de funciones gaussianas, es:

$$\sum_{i=1}^{n} \sum_{x, c(x)=c_i} (g_i(x) - 1)^2 + \sum_{j \neq i} (g_j(x) + 1)^2$$



Definición

Se llama Series de Tiempo a un conjunto de observaciones sobre valores que toma una variable (cuantitativa) en diferentes momentos del tiempo. Denotaremos como Y_t al valor de la variable Y en el instante t. Ejemplos: Precios de un articulo, tasas de desempleo, tasa de inflación, índice de precios, precio del dólar, precio del cobre, precios de acciones, ingreso nacional bruto.

Definición

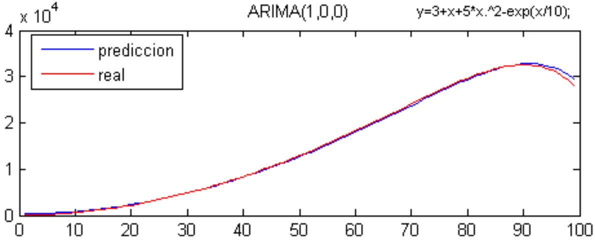
En los modelos autorregresivos, el proceso se representa como una suma ponderada de observaciones pasadas de la variable. El número de rezagos (p) determina el orden del modelo autorregresivo.

$$Y_t = \delta + \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + \varepsilon_t$$

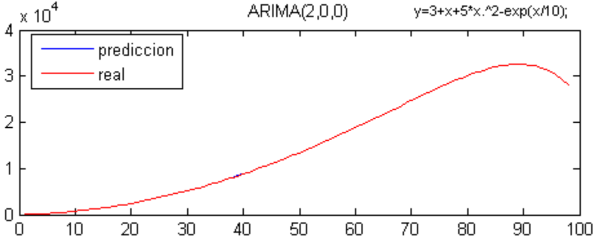
Proposición

Dada la serie $Y_1, ..., Y_n$, podemos crear el conjunto de datos $D = \{(\vec{x}_1, Y_{1+p}), .., (\vec{x}_{n-p}, Y_n)\}$, donde $\vec{x}_i = (1, Y_i, ..., Y_{i+p-1})$, el espacio de funciones lineales en $\mathbb{R}^p$ y la función de error $E(f, D) = \sum_{i=1}^{n-p} (\phi_0 + \phi_1 Y_i + ... + \phi_p Y_{i+p-1} - Y_{i+p})^2$. Luego, el regresor f es el modelo autorregresivo óptimo.

$$\vec{\phi} = (X' * X)^{-1} * X' * \vec{Y}_{1+p}^{n}$$



x 10^4
ARIMA(1,0,0)
y=3+x+5*x.^2-exp(x/10);
prediccion
real
0
1
2
3
4
0
10
20
30
40
50
60
70
80
90
100



x 10^4
ARIMA(2,0,0)
y=3+x+5*x.^2-exp(x/10);
prediccion
real
0
1
2
3
4
0
10
20
30
40
50
60
70
80
90
100