

# Ideas geométricas en el análisis de datos funcionales y regresión no paramétrica.

Alejandro Cholaquidis

CMAT-Facultad de Ciencias, Udelar

Seminario de Probabilidad y Estadística

## 1 Introducción

## 2 Métricas en espacios de funciones

## 3 Aplicación: Clasificación de espectrometrías

## 1 Introducción

## 2 Métricas en espacios de funciones

## 3 Aplicación: Clasificación de espectrometrías

# Regresión y clasificación en general

## Idea:

A partir de una muestra de entrenamiento  $(X_1, Y_1), \dots, (X_n, Y_n)$ , construir y entrenar un modelo  $g$  que permitirá, dada una nueva observación  $X$ , predecir la categoría a la que pertenece o algún valor relevante de salida.

## Ejemplos:

- 1) Predecir si un correo es spam o no spam.
- 2) Predecir si un paciente es propenso a tener una determinada enfermedad.
- 3) Estimar la tasa de ozono en una ciudad teniendo en cuentas variables climáticas.
- 4) Predecir la ausencia o presencia de una especie en determinado medio.
- 5) Identificar cifras manuscritas de códigos postales en sobres.

# El modelo y las variables

Queremos construir un modelo de predicción  $g$  a partir de una muestra de entrenamiento que permitirá, dado un nuevo *input*  $X$ , la predicción del *output*  $Y$ , es decir  $Y = g(X)$ .

- $X$  (input): variable independiente, explicativa, de entrada.  
En FDA:  $X$  toma valores en  $(\mathcal{E}, d)$  separable y completo.  
 $Y$  (output): variable dependiente, de salida reales, multidimensionales, continuas, funcionales, categóricas, binarias, etc.
- Si  $Y$  es categórica: problema de clasificación.
- Datos: muestra de entrenamiento  $D_n = \{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\}$ .

Suponemos que se cumple el modelo

$$Y = g(X) + \varepsilon$$

donde  $g$  una función desconocida,  $\varepsilon$  un error. Se asume  $\mathbb{E}(\varepsilon|X) = 0$ .

## Regresión y clasificación en FDA.

Un posible criterio para encontrar  $g$

Minimizar MSE:  $\mathbb{E}\|Y - g(X)\|^2$ .

Teorema

La función que minimiza el MSE es:  $\eta(x) = \mathbb{E}(Y|X = x)$ .

En clasificación binaria es:  $\eta(X) = P(Y = 1|X)$ .

Vamos a *aproximar*  $\eta$  por:

$$\eta_n(X) = \sum_{i=1}^n W_{ni}(X) Y_i$$

para ciertas funciones de *pesos*  $W_{ni}(X) = W_{ni}(X, X_1, \dots, X_n)$ .

En clasificación obtenemos la regla:

$$g_n(X) = \mathbb{I}_{\{\eta_n(X) \geq 1/2\}}.$$

Si  $(\mathcal{E}, d)$  es separable, localmente compacto. Bajo ciertas hipótesis sobre  $W_{ni}$ .

Teorema (Consistencia)

$$\mathbb{E}\left(P\left(\mathbb{I}_{\{\eta_n(X) \geq 1/2\}} \neq Y | D_n\right)\right) \rightarrow P\left(\mathbb{I}_{\{\eta(X) \geq 1/2\}} \neq Y\right).$$

# K-NN

En el caso particular en que

$$W_{ni}(X) = \frac{1}{k} \mathbb{I}_{\{X_i \in k(X)\}},$$

donde  $X_i \in k(X)$  indica que  $X_i$  es uno de los  $k$  datos más próximos (respecto de  $d$ ) de  $X$ , se obtiene el estimador por vecinos más cercanos.

La idea es tomar  $\mathcal{E} = \{f : [0, 1] \rightarrow \mathbb{R}\}$  y definir una métrica que haga que el espacio sea localmente compacto.

## Observación

- *Los e.v.t son localmente compactos si y solo si son de dimensión finita.*
- *Las métricas usuales (supremo,  $L^p$  etc) no son loc. compactas.*

1 Introducción

2 Métricas en espacios de funciones

3 Aplicación: Clasificación de espectrometrías



# Métrica de Skorohod, 1955

## Definición

Una función  $f : [0, 1] \rightarrow \mathbb{R}$  se dice que es una función càdlàg si para todo  $t \in [0, 1]$

$$1) \exists f(t^-) = \lim_{x \rightarrow t^-} f(x) \quad 2) \exists f(t^+) = \lim_{x \rightarrow t^+} f(x) = f(t)$$

Denotaremos  $D$  al espacio de las funciones càdlàg.

## Definición

Denotemos  $\Lambda$  al conjunto de las biyecciones continuas, crecientes, de  $[0, 1]$  en  $[0, 1]$ . Sean  $f, g \in D$ , la distancia de Skorohod es:

$$d_S(f, g) = \inf_{\lambda \in \Lambda} \max \{ \|\lambda - Id\|_\infty, \|f - g \circ \lambda\|_\infty \}.$$

## Observación

- $d_S$  restringida al espacio de las funciones continuas  $C[0, 1]$  coincide con la métrica uniforme.
- $(D, d_S)$  es separable, por lo tanto  $X_n : ([0, 1], \mathcal{B}, \mu)^n \rightarrow (D, d_S)$  es medible,

$$X_n(t) = \sqrt{n}(F_n(t) - t) \quad F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{[0,t]}(U_i) \quad 0 \leq t \leq 1.$$



# Métrica de Levy y distancia de Hausdorff entre conjuntos

## Definición

$F$  y  $G$  dos distribuciones en  $\mathbb{R}$ , la distancia de Lévy es:

$$d_L(F, G) = \inf \{ \varepsilon > 0 : F(x - \varepsilon) - \varepsilon \leq G(x) \leq F(x + \varepsilon) + \varepsilon \}.$$

Para definir una nueva distancia entre funciones vamos a necesitar:

## Definición

Sean  $A, C$  subconjuntos compactos no vacíos de  $\mathbb{R}^d$

$$d_H(A, C) = \max \left\{ \sup_{a \in A} d(a, C), \sup_{c \in C} d(c, A) \right\},$$

siendo

$$d(x, C) = \inf \{ d(x, c) : c \in C \}$$

$d$  es una métrica localmente compacta en los subconjuntos compactos.

# USC e hipografos

## Definición

Sea  $f : [0, 1] \rightarrow \mathbb{R}$  es semicontinua por arriba en  $x_0$  si  $\limsup_{x \rightarrow x_0} f(x) \leq f(x_0)$ .

## Definición

Sea  $f : [0, 1] \rightarrow \mathbb{R}$  USC, el hipografo de  $f$ , que denotaremos  $H_f$  se define como

$$H_f = \{(x, y) \in \mathbb{R}^2 : x \in [0, 1], -\infty < y \leq f(x)\}.$$

## Proposición

Sea  $f : [0, 1] \rightarrow \mathbb{R}$  no negativa y USC entonces

- 1) existe  $z \in [0, 1]$  tal que  $\sup_{x \in [0, 1]} f(x) = f(z)$ .
- 2)  $H_f^+ = \{(x, y) \in \mathbb{R}^2 : x \in [0, 1], 0 \leq y \leq f(x)\}$  es compacto.

# Distancia de Hausdorff entre hipografos

## Definición

Sean  $f, g : [0, 1] \rightarrow \mathbb{R}$  USC y acotadas inferiormente definimos

$$d(f, g) = d_H(H_{f+k}, H_{g+k}),$$

siendo  $k = \max \{0, -\inf f(x), -\inf g(x)\}$ .

Notación:  $(\mathcal{E}, d)$ .

## Observación

- $d(f_n, f) \rightarrow 0 \not\Rightarrow \forall x f_n(x) \rightarrow f(x)$  ej:  $f_n(x) = nx$ .
- $\forall x f_n(x) \rightarrow f(x) \not\Rightarrow d(f_n, f) \rightarrow 0$  ej:  $f_n(x) = x^n(1 - x^n)$ .
- $f_n$  y  $f$  USC y acotadas inferiormente si  $d(f_n, f) \rightarrow 0$  entonces

$$\max_{x \in [0, 1]} f(x) = \lim_{n \rightarrow +\infty} \max_{x \in [0, 1]} f_n(x).$$

# Un algoritmo para aproximar $d$

Observemos que la distancia de Hausdorff varía si cambiamos un valor funcional, por lo tanto un algoritmo exacto que para cualquier función USC no es posible. Vamos a aproximar  $d$  para algunas funciones, a partir de  $(t_1, f(t_1)), \dots, (t_n, f(t_n))$ . Para eso:

## Lema

$f, g : [0, 1] \rightarrow \mathbb{R}$  no negativas, USC, existen  $u \in \partial H_f$  y  $v \in \partial H_g$  tal que  $d(f, g) = d(u, v)$ .

## Observación

- Si  $f$  y  $g$  son funciones USC, acotadas inferiormente, como  $H_f$  y  $H_g$  son compactos,  $d(f, g) \leq d_H(\partial H_f, \partial H_g)$  no obstante, aún en el caso en que  $f$  y  $g$  son continuas y positivas, la igualdad no es cierta en general.
- No es cierto en general que si  $A$  y  $B$  son subconjuntos compactos de  $\mathbb{R}^2$  existen  $a \in \partial A$  y  $b \in \partial B$  tal que  $d(a, b) = d_H(A, B)$  basta tomar  $B = B(0, 1)$  y  $A = B(0, 1) \setminus B(0, 3/4)$ .

# Un algoritmo para aproximar $d$

## Proposición

Sean  $f, g : [0, 1] \rightarrow \mathbb{R}$  no negativas, continuas sean  $x$  e  $y$  tal que  $d(f, g) = d(x, y)$  entonces existen  $t \in [0, 1]$  y  $s \in [0, 1]$  tal que  $x = (t, f(t))$  e  $y = (s, g(s))$ .

## Observación

*Ambas funciones tienen que ser continuas:*

Consideremos  $g(x) = \mathbb{I}_{[1/2, 1]}(x)$  y

$$f(x) \begin{cases} -4x + 1/2 & \text{si } 0 \leq x \leq 1/8 \\ 0 & \text{si } 1/8 < x < 3/8 \\ 8(x - 3/8) & \text{si } 3/8 \leq x < 1/2 \\ 1 & \text{si } 1/2 \leq x \leq 1 \end{cases}$$

$f$  es continua,  $g$  es USC,  $d(f, g) = \sup_{x \in H_f} d(x, H_g) = 1/2$ .

# Un algoritmo para aproximar $d$

## Teorema

Sean  $f, g : [0, 1] \rightarrow \mathbb{R}$  no negativas USC, entonces

$$d(f, g) := \max \left\{ \sup_{\left\{ \begin{array}{l} x=(x_1, x_2) \in \partial H_g \\ g(x_1) \geq f(x_1) \end{array} \right\}} d(x, \partial H_f), \sup_{\left\{ \begin{array}{l} y=(y_1, y_2) \in \partial H_f \\ f(y_1) \geq g(y_1) \end{array} \right\}} d(y, \partial H_g) \right\}$$

Si tenemos  $(t_1, f(t_1)), \dots, (t_n, f(t_n))$  y  $(t_1, g(t_1)), \dots, (t_n, g(t_n))$  y asumimos que las funciones  $f$  y  $g$  son continuas, calculamos:

$$\tilde{d}(f, g) = \max \left\{ \max_{\{i: g(t_i) \geq f(t_i)\}} \hat{d}((t_i, g(t_i)), \partial H_f), \max_{\{i: f(t_i) \geq g(t_i)\}} \hat{d}((t_i, f(t_i)), \partial H_g) \right\}$$

donde para cada  $i : g(t_i) > f(t_i)$ :

$$\hat{d}((t_i, g(t_i)), \partial H_f) = \min \left\{ d((t_i, g(t_i)), (t_1, f(t_1))), \dots, d((t_i, g(t_i)), (t_n, f(t_n))) \right\},$$

y para cada  $i : f(t_i) > g(t_i)$ :

$$\hat{d}((t_i, f(t_i)), \partial H_g) = \min \left\{ d((t_i, f(t_i)), (t_1, g(t_1))), \dots, d((t_i, f(t_i)), (t_n, g(t_n))) \right\}.$$

# Compacidad local de $(\mathcal{E}, d)$

## Proposición

*El espacio  $(\mathcal{E}, d)$  es localmente compacto.*

## Demostración.

- Sea  $\{f_n\} \subset \mathcal{E}$  una sucesión acotada con la métrica  $d$ , existe  $C = \lim H_{f_n}$ .
- Hay que ver:  $C = H_f$  para alguna  $f : [0, 1] \rightarrow \mathbb{R}$  USC.
- Sea  $f : [0, 1] \rightarrow \mathbb{R}^+$  definida como

$$x \mapsto \sup_{\{x_n\} : x_n \rightarrow x} \limsup f_n(x_n).$$

*Entonces  $f$  es USC y  $H_f = C$ .*





## 1 Introducción

## 2 Métricas en espacios de funciones

## 3 Aplicación: Clasificación de espectrometrías

# Aplicación: Clasificación de espectrometrías

## Espectrometría:

El espectrómetro de masas es un instrumento que permite analizar con gran precisión la composición de diferentes elementos químicos e isótopos atómicos, separando los núcleos atómicos en función de su relación masa/carga. Una espectrometría es una función que asigna a cada cociente masa/carga el número de moléculas encontradas correspondiente a dicho cociente.

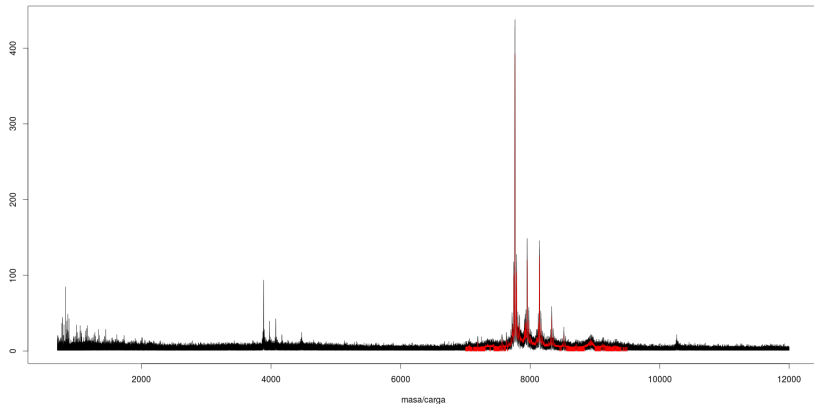
## Datos:

216 espectrometrías hechas a un grupo de 216 mujeres, con el fin de determinar la presencia o no de cáncer de ovario. Los valores de masa/carga de nuestra muestra varían entre 700 y 12.000, y tenemos, para dicho rango, entre 320.000 y 360.000 valores de la función, *dependiendo del paciente estudiado*.

## Tratamiento de los datos

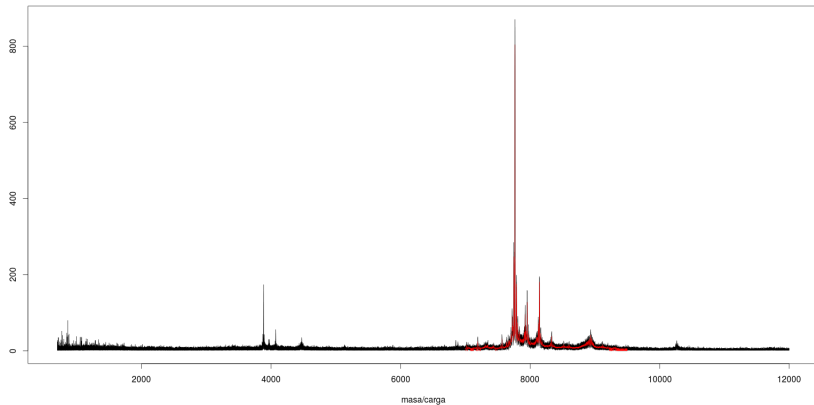
- 1) Construimos el estimador de Nadaraya-Watson de la muestra, (usando núcleos gaussianos) y lo evaluamos en una grilla equiespaciada de  $p$  valores en  $[7.000, 9.500]$ .
- 2) Asignamos el valor 0 a los cocientes masa/carga con valor funcional menor que 5.

# Una paciente sin la enfermedad



**Figura:** En negro la espectrometría de una paciente sin la enfermedad y en rojo la función que tomamos como aproximación, antes del reescalado.

# Una paciente con la enfermedad



**Figura:** En negro la espectrometría de una paciente con la enfermedad y en rojo la función que tomamos como aproximación, antes del reescalado.








# Resultados

## Con “Leave One Out”

k=5		
	$\hat{C}$	$\hat{N}$
C	113	8
N	9	86

k=7		
	$\hat{C}$	$\hat{N}$
C	110	11
N	7	88

Si bien LOO fue el método que dio mejores resultados (luego de probar diferentes valores de ruido, y variar  $p$ ). Usamos otros métodos como ponerle pesos bajos a las observaciones “malas” (aquellas de una clase, que son mal clasificadas usando la muestra de entrenamiento). También probamos tomando muestras de entrenamiento de tamaño 50, quitando outliers, etc.

-  Bañllo, A.; Cuevas, A. y Fraiman, R. (2010) - *Classification methods for functional data*. The Oxford Handbook of Functional Data Analysis.
-  Alt, H.; Brab, P.; Godau, M.; Knauer, C. y Wenk, C. (2003) *Computing the Hausdorff Distance of Geometric Patterns and Shapes*. Discrete and Computational Geometry. Springer-Verlag Berlin Heidelberg
-  Billingsley, P. (1999) *Convergence of probability measures, s.e.*. Wiley.
-  Cérou, F. y Guyader, A. (2005) *Nearest neighbor classification in infinite dimension*. Rapport de Recherche 5536, INRIA.
-  Cuevas, A. y Fraiman, R. (1998) *On visual distances in dencity estimation: The Hausdorff choice*. Statistics and probability letters.
-  Devroye, L.; Györfi, L. y Lugosi, G. (1996) *A probabilistic theory of pattern recognition..* Springer-Verlag, New York.
-  Forzani, L.; Fraiman, R. y Llop, P. (2012) *Consistent nonparametric regression for functional data under the Stone-Besicovitch conditions*
-  Marron, J.S. y Tsybakov, A.B. (1995) *Visual error criteria for qualitative smoothing*. Journal of the American Statistical Association. Volumen 90.



Nutanong, S.; Jacox, E. y Samet, H (2011) *An incremental Hausdorff distance calculation algorithm*. Proceedings of the VLDB Endowment. Volumen 4.



Sendov, B. (1990) - *Hausdorff approximations*. Kluwer, Dodrecht.



Scharf, L. (2003) *Computing the Hausdorff distance between sets of curves*. Freire Universität Berlin.



Stone, C. (1977) *Consistent Nonparametric Regression*. The Annals of Statistics.