Regresión en variedades y aplicaciones a la inferencia conformal

Alejandro Cholaquidis

Seminario de Probabilidad y Estadística - 2024

Algunos datos en variedades

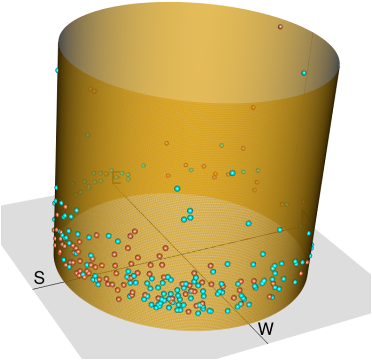
Regresión en $\mathbb{R}^d$

Regresión en variedades

Inferencia conforme

Datos en variedades - Mediciones de viento

Una medición de viento se puede ver como un dato en el cilindro: una dirección en S^1 y una intensidad en $\mathbb{R}^+$, Cholaquidis et al. (2022b)



Cono de matrices definidas positivas

- C_r matrices simétricas $r \times r$ definidas positivas.

[1] Recordar que M es def. positiva si $v^T M v > 0$ para todo v, donde v^T denote vector transpuesto

Cono de matrices definidas positivas

- C_r matrices simétricas $r \times r$ definidas positivas. [1].

- Incluye matrices de covarianza no degeneradas.

[1] Recordar que M es def. positiva si $v^T M v > 0$ para todo v, donde v^T denote vector transpuesto

Cono de matrices definidas positivas

- C_r matrices simétricas $r \times r$ definidas positivas. [1].

- Incluye matrices de covarianza no degeneradas.

- Se puede definir en C_r,

$$\rho(A, B) = || \log(A^{-1/2} B A^{-1/2}) ||.$$

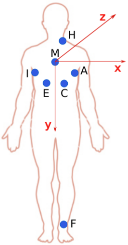
  donde es $||A||^2 = \text{traza}(A^T A)$, y resulta una variedad diferenciable, ver Bonnabel and Sepulchre (2010); Cholaquidis et al. (2022a),

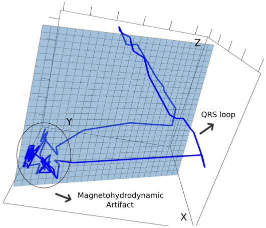
- Si queremos predecir una matriz de covarianzas no cualquier método de regresión va a resultar adecuado, ya que podría dar una matriz que no sea definida positiva.

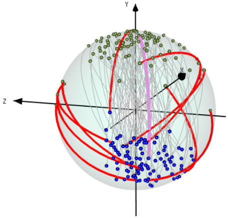
[1] Recordar que M es def. positiva si $v^T M v > 0$ para todo v, donde v^T denote vector transpuesto

Datos en variedades -Vectocardiogramas

Se obtiene de los electrodos una curva, que se resume en un elemento de $S(3, 2)$: bases ortonormales de planos en $\mathbb{R}^3$, ver Cholaquidis et al. (2022b)

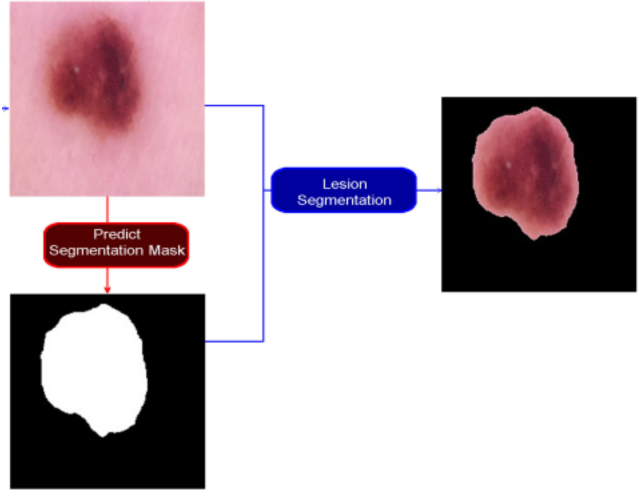






Imágenes como datos en espacios métricos

En este caso podemos trabajar en (M, d_H) con M los subconjuntos compactos de algún e.m.,



La esfera

Un modelo de regresión en la esfera (ver Mardia et al. (2000)):

$$Y_i \sim \mathrm{M}\left(\frac{\eta + \beta X_i}{\|\eta + \beta X_i\|}, \kappa\right), \; i = 1, 2 \ldots, 400$$

M es la distribución de Von-Misses en la esfera[2], $X_i \sim \mathrm{U}(-1, 1)$ iid.

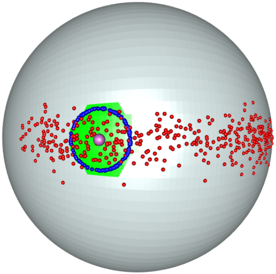
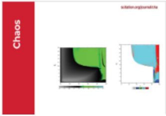


Figura: $\eta = (1, 0, 0)$, $\beta = (0, 0, 1)$, y $\kappa = 200$. Se representan 400 puntos Y_i en rojo. Los puntos en azul pertenecen al borde de la banda teórica de confianza 90 % para $X_i = 0$. Los puntos en verde pertenecen a la estimación de dicho conjunto.

[2] su densidad es $f(x; \mu, \kappa) = C \exp(\kappa \mu^T x), x \in S^2$

Algo de irrupción en la economía

Volume 32, Issue 8
August 2022



RESEARCH ARTICLE | AUGUST 08 2022

Time-series forecasting using manifold learning, radial basis function interpolation, and geometric harmonics FREE

Panagiotis G. Papaioannou; Ronen Talmon; Ioannis G. Kevrekidis; Constantinos Siettos

Check for updates

+ Author & Article Information

Chaos 32, 083113 (2022)

Detecting strange attractors in turbulence

Pasan de un proceso (o serie de tiempo) a una variedad, usando un Teo de Florins Takens (1981).

Theorem 1. Let M be a compact manifold of dimension m. For pairs (φ, y), $\varphi : M \to M$ a smooth diffeomorphism and $y : M \to \mathbb{R}$ a smooth function, it is a generic property that the map $\Phi_{(\varphi,y)} : M \to \mathbb{R}^{2m+1}$, defined by

$$\Phi_{(\varphi,y)}(x) = (y(x), y(\varphi(x)), \ldots, y(\varphi^{2m}(x))$$

is an embedding; by "smooth" we mean at least C^2.

Toman $M = \mathbb{R}$, donde $t_i = t_0 + i\delta$, para $i = 0, \ldots, m-1$ y $\varphi(t) = t + \delta$. $y(t) = X(t)$ el proceso (o serie de tiempo) en t.

Algo de irrupción en la economía

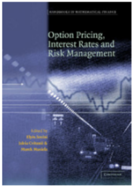
Econometrica, Vol. 84, No. 3 (May, 2016), 1249–1264

A GEOMETRIC APPROACH TO NONLINEAR ECONOMETRIC MODELS

ISAIAH ANDREWS
Harvard University, Cambridge, MA 02138, U.S.A.

ANNA MIKUSHEVA
M.I.T., Cambridge, MA 02142, U.S.A.

Algo de irrupción en la economía



Handbooks in Mathematical Finance

7 - A Geometric View of Interest Rate Theory

from Part two - Interest Rate Modeling

Published online by Cambridge University Press: **29 January 2010**

By T. Björk

Edited by E. Jouini, J. Cvitanic and Marek Musiela

Chapter

Get access | Share | Cite

Algo de irrupción en la economía

ELSEVIER

Decision Support Systems

Volume 64, August 2014, Pages 31-42

A kernel entropy manifold learning approach for financial data analysis ☆

Yan Huang [a], Gang Kou [b c]

Algo de irrupción en la economía

ELSEVIER

European Journal of Operational Research
Volume 258, Issue 2, 16 April 2017, Pages 692-702

Innovative Applications of O.R.

Nonlinear manifold learning for early warnings in financial markets

Huang Yan [a 1], Kou Gang [b 1], Peng Yi [c 1]

Algo de irrupción en la economía

10.1098/rspa.2000.0722

THE ROYAL SOCIETY

Interest rates and information geometry

By Dorje C. Brody[1] and Lane P. Hughston[2]

[1] *Blackett Laboratory, Imperial College of Science, Technology and Medicine, London SW7 2BZ, UK and*
Centre for Mathematical Science, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK (dorje@ic.ac.uk)

[2] *Department of Mathematics, King's College London, Strand, London WC2R 2LS, UK* (lane.hughston@kcl.ac.uk)

Algo de irrupción en la economía

J. Phys. A: Math. Gen. **33** (2000) L5–L14. Printed in the UK PII: S0305-4470(00)07603-4

LETTER TO THE EDITOR

Gauge geometry of financial markets

Kirill Ilinski†

IPhys Group, CAPE, 14th line of Vasilievskii's Island, 29 St Petersburg 199178, Russian Federation
and
School of Physics and Space Research, University of Birmingham, Edgbaston, B15 2TT Birmingham, UK

Algo de irrupción en la economía

mathematics

Article

Interest Rate Based on The Lie Group *SO(3)* in the Evidence of Chaos

Melike Bildirici [1], Yasemen Ucan [2] and Sérgio Lousada [3,4,5,6,7,*]

Algo de irrupción en la economía

Quantitative Finance ›
Volume 11, 2011 - Issue 4: Special Issue on Rates and FX

Submit an article | Journal homepage

Research Papers

Interest rate models on Lie groups

F. C. Park, C. M. Chun, C. W. Han & N. Webber

Pages 559-572 | Received 26 Jun 2008, Accepted 29 Oct 2009, Published online: 11 May 2010

Cite this article https://doi.org/10.1080/14697680903468963

Regresión en $\mathbb{R}^d$

- Dada $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ queremos saber cómo la variable de respuesta Y depende del vector de observaciones X.

Regresión en $\mathbb{R}^d$

- Dada $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ queremos saber cómo la variable de respuesta Y depende del vector de observaciones X.
- ...o un poco menos: una $f : \mathbb{R}^d \to \mathbb{R}$ tal que $f(X)$ sea una buena aproximación de Y.

Regresión en $\mathbb{R}^d$

- Dada $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ queremos saber cómo la variable de respuesta Y depende del vector de observaciones X.
- ...o un poco menos: una $f : \mathbb{R}^d \to \mathbb{R}$ tal que $f(X)$ sea una buena aproximación de Y.
- La f medible que minimiza $\mathbb{E}|f(X) - Y|^2$ es $\mathbb{E}(Y|X = x) = f(x)$.

Regresión en $\mathbb{R}^d$

- Dada $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ queremos saber cómo la variable de respuesta Y depende del vector de observaciones X.
- ...o un poco menos: una $f : \mathbb{R}^d \to \mathbb{R}$ tal que $f(X)$ sea una buena aproximación de Y.
- La f medible que minimiza $\mathbb{E}|f(X) - Y|^2$ es $\mathbb{E}(Y|X = x) = f(x)$.
- El modelo de regresión clásico con ruido aditivo es $Y = f(X) + \epsilon$ con $\mathbb{E}(\epsilon|X) = 0$.

Regresión en $\mathbb{R}^d$

- Dada $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ queremos saber cómo la variable de respuesta Y depende del vector de observaciones X.
- ...o un poco menos: una $f : \mathbb{R}^d \to \mathbb{R}$ tal que $f(X)$ sea una buena aproximación de Y.
- La f medible que minimiza $\mathbb{E}|f(X) - Y|^2$ es $\mathbb{E}(Y|X = x) = f(x)$.
- El modelo de regresión clásico con ruido aditivo es $Y = f(X) + \epsilon$ con $\mathbb{E}(\epsilon|X) = 0$.
- Los métodos clásicos de regresión (k-nn, Kernel, redes neuronales, RF, etc) aproximan $\mathbb{E}(Y|X = x)$ a partir de una muestra iid de (X, Y).

Regresión en $\mathbb{R}^d$

- Dada $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ queremos saber cómo la variable de respuesta Y depende del vector de observaciones X.
- ...o un poco menos: una $f : \mathbb{R}^d \to \mathbb{R}$ tal que $f(X)$ sea una buena aproximación de Y.
- La f medible que minimiza $\mathbb{E}|f(X) - Y|^2$ es $\mathbb{E}(Y|X = x) = f(x)$.
- El modelo de regresión clásico con ruido aditivo es $Y = f(X) + \epsilon$ con $\mathbb{E}(\epsilon|X) = 0$.
- Los métodos clásicos de regresión (k-nn, Kernel, redes neuronales, RF, etc) aproximan $\mathbb{E}(Y|X = x)$ a partir de una muestra iid de (X, Y).
- Se pueden extender a X y/o $Y \in H$ con H de Hilbert de manera natural y existen diversos resultados de consistencia.

Regresión en $\mathbb{R}^d$

- Dada $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ queremos saber cómo la variable de respuesta Y depende del vector de observaciones X.
- ...o un poco menos: una $f : \mathbb{R}^d \to \mathbb{R}$ tal que $f(X)$ sea una buena aproximación de Y.
- La f medible que minimiza $\mathbb{E}|f(X) - Y|^2$ es $\mathbb{E}(Y|X = x) = f(x)$.
- El modelo de regresión clásico con ruido aditivo es $Y = f(X) + \epsilon$ con $\mathbb{E}(\epsilon|X) = 0$.
- Los métodos clásicos de regresión (k-nn, Kernel, redes neuronales, RF, etc) aproximan $\mathbb{E}(Y|X = x)$ a partir de una muestra iid de (X, Y).
- Se pueden extender a X y/o $Y \in H$ con H de Hilbert de manera natural y existen diversos resultados de consistencia.
- El caso $Y \in \{1, \ldots, K\}$ se llama clasificación, se busca minimizar $\mathbb{P}(f(X) \neq Y)$.

Algunas preguntas...

- ¿Tiene interés el caso $Y \in M$ con M una variedad?

Regresión en $\mathbb{R}^d$

- Dada $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ queremos saber cómo la variable de respuesta Y depende del vector de observaciones X.
- ...o un poco menos: una $f : \mathbb{R}^d \to \mathbb{R}$ tal que $f(X)$ sea una buena aproximación de Y.
- La f medible que minimiza $\mathbb{E}|f(X) - Y|^2$ es $\mathbb{E}(Y|X = x) = f(x)$.
- El modelo de regresión clásico con ruido aditivo es $Y = f(X) + \epsilon$ con $\mathbb{E}(\epsilon|X) = 0$.
- Los métodos clásicos de regresión (k-nn, Kernel, redes neuronales, RF, etc) aproximan $\mathbb{E}(Y|X = x)$ a partir de una muestra iid de (X, Y).
- Se pueden extender a X y/o $Y \in H$ con H de Hilbert de manera natural y existen diversos resultados de consistencia.
- El caso $Y \in \{1, \ldots, K\}$ se llama clasificación, se busca minimizar $\mathbb{P}(f(X) \neq Y)$.

Algunas preguntas...

- ¿Tiene interés el caso $Y \in M$ con M una variedad?
- Más general ¿Y en un espacio métrico ?

Regresión en $\mathbb{R}^d$

- Dada $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ queremos saber cómo la variable de respuesta Y depende del vector de observaciones X.
- ...o un poco menos: una $f : \mathbb{R}^d \to \mathbb{R}$ tal que $f(X)$ sea una buena aproximación de Y.
- La f medible que minimiza $\mathbb{E}|f(X) - Y|^2$ es $\mathbb{E}(Y|X = x) = f(x)$.
- El modelo de regresión clásico con ruido aditivo es $Y = f(X) + \epsilon$ con $\mathbb{E}(\epsilon|X) = 0$.
- Los métodos clásicos de regresión (k-nn, Kernel, redes neuronales, RF, etc) aproximan $\mathbb{E}(Y|X = x)$ a partir de una muestra iid de (X, Y).
- Se pueden extender a X y/o $Y \in H$ con H de Hilbert de manera natural y existen diversos resultados de consistencia.
- El caso $Y \in \{1, \ldots, K\}$ se llama clasificación, se busca minimizar $\mathbb{P}(f(X) \neq Y)$.

Algunas preguntas...

- ¿Tiene interés el caso $Y \in M$ con M una variedad?
- Más general ¿Y en un espacio métrico ?
- ¿Que sería $\mathbb{E}(Y|X = x)$ cuando $Y \in M$?

Regresión en $\mathbb{R}^d$

- Dada $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ queremos saber cómo la variable de respuesta Y depende del vector de observaciones X.
- ...o un poco menos: una $f : \mathbb{R}^d \to \mathbb{R}$ tal que $f(X)$ sea una buena aproximación de Y.
- La f medible que minimiza $\mathbb{E}|f(X) - Y|^2$ es $\mathbb{E}(Y|X = x) = f(x)$.
- El modelo de regresión clásico con ruido aditivo es $Y = f(X) + \epsilon$ con $\mathbb{E}(\epsilon|X) = 0$.
- Los métodos clásicos de regresión (k-nn, Kernel, redes neuronales, RF, etc) aproximan $\mathbb{E}(Y|X = x)$ a partir de una muestra iid de (X, Y).
- Se pueden extender a X y/o $Y \in H$ con H de Hilbert de manera natural y existen diversos resultados de consistencia.
- El caso $Y \in \{1, \dots, K\}$ se llama clasificación, se busca minimizar $\mathbb{P}(f(X) \neq Y)$.

Algunas preguntas...

- ¿Tiene interés el caso $Y \in M$ con M una variedad?
- Más general ¿Y en un espacio métrico ?
- ¿Que sería $\mathbb{E}(Y|X = x)$ cuando $Y \in M$?
- ¿Que es $\mathbb{E}(Y)$ si $Y \in M$?...

Regresión en $\mathbb{R}^d$

- Dada $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ queremos saber cómo la variable de respuesta Y depende del vector de observaciones X.
- ...o un poco menos: una $f : \mathbb{R}^d \to \mathbb{R}$ tal que $f(X)$ sea una buena aproximación de Y.
- La f medible que minimiza $\mathbb{E}|f(X) - Y|^2$ es $\mathbb{E}(Y|X = x) = f(x)$.
- El modelo de regresión clásico con ruido aditivo es $Y = f(X) + \epsilon$ con $\mathbb{E}(\epsilon|X) = 0$.
- Los métodos clásicos de regresión (k-nn, Kernel, redes neuronales, RF, etc) aproximan $\mathbb{E}(Y|X = x)$ a partir de una muestra iid de (X, Y).
- Se pueden extender a X y/o $Y \in H$ con H de Hilbert de manera natural y existen diversos resultados de consistencia.
- El caso $Y \in \{1, \dots, K\}$ se llama clasificación, se busca minimizar $\mathbb{P}(f(X) \neq Y)$.

Algunas preguntas...

- ¿Tiene interés el caso $Y \in M$ con M una variedad?
- Más general ¿Y en un espacio métrico ?
- ¿Que sería $\mathbb{E}(Y|X = x)$ cuando $Y \in M$?
- ¿Que es $\mathbb{E}(Y)$ si $Y \in M$?... media de Fréchet.

Algunas problemas técnicos

Problemas si (M, d_M) es un e.m

- Definir $\mathbb{E}(Y|X = x)$ con R.N requiere poder definir $\mathbb{E}(Y\mathbf{I}_{X\in B})$ para B boreliano de $\mathbb{R}^d$.

Algunas problemas técnicos

Problemas si (M, d_M) es un e.m

- Definir $\mathbb{E}(Y|X = x)$ con R.N requiere poder definir $\mathbb{E}(Y\mathbf{I}_{X\in B})$ para B boreliano de $\mathbb{R}^d$.
- La media de Fréchet: $\mathbb{E}(Y) = \arg\min_{z\in M} \mathbb{E}(d_M(Y, z)^2)$ no tiene por qué existir o ser única.

Algunas problemas técnicos

Problemas si (M, d_M) es un e.m

- Definir $\mathbb{E}(Y|X=x)$ con R.N requiere poder definir $\mathbb{E}(Y\mathbf{I}_{X\in B})$ para B boreliano de $\mathbb{R}^d$.
- La media de Fréchet: $\mathbb{E}(Y) = \arg\min_{z\in M} \mathbb{E}(d_M(Y,z)^2)$ no tiene por qué existir o ser única.

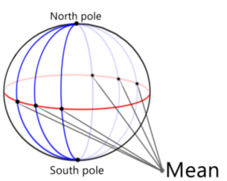


Figura: La media de Fréchet en la esfera es cualquier punto del Ecuador, si los datos son el polo norte y el polo sur. Si P_Y es uniforme en S^2 su media es toda la esfera.

- Si $f(x) = \mathbb{E}(Y|X=x)$ minimiza $\mathbb{E}|Y-g(X)|^2$ entre las $g:\mathbb{R}^d \to \mathbb{R}$ medibles, es razonable proponer.

$$\mathbb{E}(Y|X=x) = \arg\min_{g:\mathbb{R}^d\to M} \mathbb{E}d_M(Y,g(X))^2$$

No está claro en que casos existe o es única...por lo tanto ¿Qué queremos estimar?

Algunas soluciones técnicas

- Si M es compacta y $C^2 \Rightarrow \mathbb{E}(Y|X = x)$ existe, no tiene por qué ser única.

Algunas soluciones técnicas

- Si M es compacta y $C^2 \Rightarrow \mathbb{E}(Y|X=x)$ existe, no tiene por qué ser única.

- Si suponemos que existe $f_{Y|X} := f_{Y,X}/f_X$ la densidad de $P_{Y|X}$, definimos

$$\mathbb{E}(Y|X=x) = \arg\min_{y \in M} \int_M d_M(y,z)^2 f_{Y|X=x}(z)dz.$$

Algunas soluciones técnicas

- Si M es compacta y $C^2 \Rightarrow \mathbb{E}(Y|X=x)$ existe, no tiene por qué ser única.

- Si suponemos que existe $f_{Y|X} := f_{Y,X}/f_X$ la densidad de $P_{Y|X}$, definimos

$$\mathbb{E}(Y|X=x) = \arg\min_{y \in M} \int_M d_M(y,z)^2 f_{Y|X=x}(z)dz.$$

- Si M tiene curvatura ≤ 0 o $\mathbb{P}_{Y|X=x}$ esta incluida en $B(p,\ell)$

$$\ell < \frac{1}{2} \text{mín} \left\{ \text{inj}(M), \frac{\pi}{2\sqrt{\Delta}} \right\},$$

Δ : cota superior de la curvatura seccional, $\text{inj}(M)$ radio de inyectividad de M.

Algunas preguntas y problemas

1. ¿Qué tan restrictivo es suponer $\exists! \ \mathbb{E}(Y|X=x)$?

Algunas soluciones técnicas

- Si M es compacta y $C^2 \Rightarrow \mathbb{E}(Y|X=x)$ existe, no tiene por qué ser única.

- Si suponemos que existe $f_{Y|X} := f_{Y,X}/f_X$ la densidad de $P_{Y|X}$, definimos

$$\mathbb{E}(Y|X=x) = \arg\min_{y \in M} \int_M d_M(y,z)^2 f_{Y|X=x}(z)dz.$$

- Si M tiene curvatura ≤ 0 o $\mathbb{P}_{Y|X=x}$ esta incluida en $B(p,\ell)$

$$\ell < \frac{1}{2} \text{mín} \left\{ \text{inj}(M), \frac{\pi}{2\sqrt{\Delta}} \right\},$$

Δ : cota superior de la curvatura seccional, $\text{inj}(M)$ radio de inyectividad de M.

Algunas preguntas y problemas

1. ¿Qué tan restrictivo es suponer $\exists!\ \mathbb{E}(Y|X=x)$?
2. Estudiar casos particulares de M y $P_{Y|X}$ donde $\exists!\ \mathbb{E}(Y|X=x)$

Algunas soluciones técnicas

- Si M es compacta y $C^2 \Rightarrow \mathbb{E}(Y|X=x)$ existe, no tiene por qué ser única.

- Si suponemos que existe $f_{Y|X} := f_{Y,X}/f_X$ la densidad de $P_{Y|X}$, definimos

$$\mathbb{E}(Y|X=x) = \arg\min_{y\in M} \int_M d_M(y,z)^2 f_{Y|X=x}(z)dz.$$

- Si M tiene curvatura ≤ 0 o $\mathbb{P}_{Y|X=x}$ esta incluida en $B(p,\ell)$

$$\ell < \frac{1}{2}\,\text{mín}\left\{\text{inj}(M), \frac{\pi}{2\sqrt{\Delta}}\right\},$$

Δ : cota superior de la curvatura seccional, $\text{inj}(M)$ radio de inyectividad de M.

Algunas preguntas y problemas

1. ¿Qué tan restrictivo es suponer $\exists!\ \mathbb{E}(Y|X=x)$?
2. Estudiar casos particulares de M y $P_{Y|X}$ donde $\exists!\ \mathbb{E}(Y|X=x)$
3. Se puede hacer regresión sin estimar $\mathbb{E}(Y|X=x)$

Algunas soluciones técnicas

- Si M es compacta y $C^2 \Rightarrow \mathbb{E}(Y|X=x)$ existe, no tiene por qué ser única.

- Si suponemos que existe $f_{Y|X} := f_{Y,X}/f_X$ la densidad de $P_{Y|X}$, definimos

$$\mathbb{E}(Y|X=x) = \arg\min_{y\in M} \int_M d_M(y,z)^2 f_{Y|X=x}(z)dz.$$

- Si M tiene curvatura ≤ 0 o $\mathbb{P}_{Y|X=x}$ esta incluida en $B(p,\ell)$

$$\ell < \frac{1}{2} \text{mín} \left\{ \text{inj}(M), \frac{\pi}{2\sqrt{\Delta}} \right\},$$

Δ : cota superior de la curvatura seccional, $\text{inj}(M)$ radio de inyectividad de M.

Algunas preguntas y problemas

1. ¿Qué tan restrictivo es suponer $\exists! \ \mathbb{E}(Y|X=x)$?
2. Estudiar casos particulares de M y $P_{Y|X}$ donde $\exists! \ \mathbb{E}(Y|X=x)$
3. Se puede hacer regresión sin estimar $\mathbb{E}(Y|X=x)$...si

k-nn

Sea $\mathcal{D}_n = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ iid de (X, Y) y $k \in \mathbb{N}^+$,

1. Ordenamos las X_i: $d_N(x, X_{1(x)}) < \cdots < d_N(x, X_{n(x)})$

k-nn

Sea $\mathcal{D}_n = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ iid de (X, Y) y $k \in \mathbb{N}^+$,

1. Ordenamos las X_i: $d_N(x, X_{1(x)}) < \cdots < d_N(x, X_{n(x)})$

2. Denotamos $Y_{1(x)}, \ldots, Y_{k(x)}$ las etiquetas de los k vecinos de x

k-nn

Sea $\mathcal{D}_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$ iid de (X, Y) y $k \in \mathbb{N}^+$,

1. Ordenamos las X_i: $d_N(x, X_{1(x)}) < \cdots < d_N(x, X_{n(x)})$

2. Denotamos $Y_{1(x)}, \dots, Y_{k(x)}$ las etiquetas de los k vecinos de x

3. Definimos el estimador

$$\hat{\varphi}_k(x) = \arg\min_{y \in M} \frac{1}{k} \sum_{i=1}^{k} d_M^2(y, Y_{i(x)}) = \arg\min_{y \in M} \hat{\mathcal{C}}_{k,x}(y)$$

   donde

$$\hat{\mathcal{C}}_{k,x}(y) = \frac{1}{k} \sum_{i=1}^{k} d_M^2(y, Y_{i(x)}).$$

k-nn

Sea $\mathcal{D}_n = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ iid de (X, Y) y $k \in \mathbb{N}^+$,

1. Ordenamos las X_i: $d_N(x, X_{1(x)}) < \cdots < d_N(x, X_{n(x)})$

2. Denotamos $Y_{1(x)}, \ldots, Y_{k(x)}$ las etiquetas de los k vecinos de x

3. Definimos el estimador

$$\hat{\varphi}_k(x) = \arg\min_{y \in M} \frac{1}{k} \sum_{i=1}^{k} d_M^2(y, Y_{i(x)}) = \arg\min_{y \in M} \hat{\mathcal{C}}_{k,x}(y)$$

   donde

$$\hat{\mathcal{C}}_{k,x}(y) = \frac{1}{k} \sum_{i=1}^{k} d_M^2(y, Y_{i(x)}).$$

El objetivo es probar la consistencia de k-n.n cuando $(X, Y) \in N \times M$, N, M variedades.

Una propuesta análoga se puede hacer con Kernel...

Algunas hipótesis sobre las distribuciones

H1

$(X, Y) \subset N \times M$ satisface **H1** si para todo $x \in \text{supp}(\mathbb{P}_X)$ existe y es única $\mathbb{E}(Y|X = x)$.

Algunas hipótesis sobre las distribuciones

H1

$(X, Y) \subset N \times M$ satisface **H1** si para todo $x \in \text{supp}(\mathbb{P}_X)$ existe y es única $\mathbb{E}(Y|X = x)$.

H2

Si para todo $x \in \text{supp}(\mathbb{P}_X)$, $y \in M$,

$$\mathcal{C}_x(y) = \int_{z \in M} d_M(y, z)^2 f_{Y|X=x}(z) dz.$$

$(X, Y) \subset N \times M$ satisface **H2** si existe $a > 0$, tal que, en un entorno U de $y^* = \mathbb{E}(Y|X = x)$,

$$\mathcal{C}_x(y) \geq \mathcal{C}_x(y^*) + a d_M(y, y^*)^2 \quad \text{para todo } y \in U$$

Algunas hipótesis sobre las distribuciones

H1

$(X, Y) \subset N \times M$ satisface **H1** si para todo $x \in \text{supp}(\mathbb{P}_X)$ existe y es única $\mathbb{E}(Y|X = x)$.

H2

Si para todo $x \in \text{supp}(\mathbb{P}_X)$, $y \in M$,

$$\mathcal{C}_x(y) = \int_{z \in M} d_M(y, z)^2 f_{Y|X=x}(z) dz.$$

$(X, Y) \subset N \times M$ satisface **H2** si existe $a > 0$, tal que, en un entorno U de $y^* = \mathbb{E}(Y|X = x)$,

$$\mathcal{C}_x(y) \geq \mathcal{C}_x(y^*) + a d_M(y, y^*)^2 \quad \text{para todo } y \in U$$

H3

$(X, Y) \subset N \times M$ satisface **H3** si existen $r_0 > 0$ y $L > 0$, tal que para todo $x, x' \in \text{supp}(\mathbb{P}_X)$ con $\|x - x'\| < r_0$ se cumple

$$W_2(\mathbb{P}_{Y|X=x}, \mathbb{P}_{Y|X=x'}) \leq L\|x - x'\| \tag{1}$$

Si existe $T_{x' \to x} : M \mapsto M$ tal que $T_{x' \to x} \#(\mathbb{P}_{Y|X=x'}) = \mathbb{P}_{Y|X=x}$ y $\forall z \in M$, $d_M(z, T_{x' \to x}(z)) \leq L_0\|x - x'\|$ $\Rightarrow$ se verifica (1) con $L = L_0$.

Ejemplos

Generaliza el caso clásico

$M = \mathbb{R}$ y $d_M(a, b) = |a - b|$.

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

donde $Y \in \mathbb{R}, X \in \mathbb{R}^{D_X}, \varepsilon | X = x \sim \mathcal{N}(0, \sigma_x^2)$.

Si $\varphi(x)$ es Lipschitz y $x \mapsto \sigma_x$ es Lipschitz $\Rightarrow$ se cumplen H1, H2, H3 y además

$$\mathbb{E}(Y|X = x) = \varphi(x).$$

Algunos resultados

Teorema

Sean N y M C^2 y compactas. Supongamos soporte$(X, Y) \subset N \times M$. Denotamos d_X la dimensión de N.

Supongamos que $\mathbb{P}_X$ es estándar respecto de ν la medida de Hausdorff en N[a].

Supongamos H1, H2 y H3, si $k_n = cn^{\frac{d_X}{d_X+2}}$ entonces

$$\mathbb{E}\left(\left\|\hat{\varphi}_{k_n}(x) - \mathbb{E}(Y|X = x)\right\|^2\right) = O\left(n^{-\frac{1}{2}\frac{d_X}{d_X+2}}\right).$$

[a] existe $\epsilon_0 > 0$ tal que $\mathbb{P}_X(B(x, \epsilon)) \geq c\nu(B(x, \epsilon))$ para todo $x \in \mathrm{supp}(\mathbb{P}_X), 0 \leq \epsilon < \epsilon_0$

Por donde va la prueba

Sea $x \in N, \varepsilon > 0$ y ϕ una densidad con soporte$(\phi) \subset B(x, \varepsilon)$.

Entonces,

$$(\sqrt{\mathcal{C}_x(y)} - L\varepsilon)^2 \leq \overline{\mathcal{C}}^{\phi}_{x,\varepsilon}(y) \leq (\sqrt{\mathcal{C}_x(y)} + L\varepsilon)^2,$$

donde $\overline{\mathcal{C}}^{\phi}_{x,\varepsilon}(y) = \int_{B(x,\varepsilon)} \phi(u)\mathcal{C}_u(y)du$.

Lo vamos a aplicar a

$$\phi_\varepsilon(u) = \frac{f_X(u)}{\mathbb{P}_X(B(x,\varepsilon))}\mathbb{I}_{B(x,\varepsilon)}(u),$$

Hoeffding mediante, si $\Lambda = diam(M)$, se puede probar

$$\mathbb{P}\left(\left|\hat{\mathcal{C}}_{k_n,x}(y) - \overline{\mathcal{C}}^{\phi_{r_n(x)}}_{x,r_n(x)}(y)\right| \geq \Delta\sqrt{\frac{\log(k_n)}{2k_n}}\right) \leq \frac{2}{k_n} \qquad (2)$$

donde $r_n \leq (k_n/n)^{D_X}$, y $N \subset \mathbb{R}^{D_X}$.

Algunos datos en variedades

Regresión en $\mathbb{R}^d$

Regresión en variedades

Inferencia conforme

El abc

- **Intercambiabilidad:** para toda permutación π de $\{1, \dots, n\}$,

$$(X_1, \dots, X_n) = (X_{\pi(1)}, \dots, X_{\pi(n)})$$

  en distribución, donde $X_i \in \mathbf{Z}$.

- **Medida de no conformidad:**

$$A(\mathcal{B}_n, z) : \mathbf{Z}^n \times \mathbf{Z} \to \mathbb{R}$$

  A invariante por permutaciones. Evalúa cuán diferente es un ejemplo z de un conjunto $\mathcal{B}_n = \{X_1, \dots, X_n\}$

- Definimos

$$\begin{aligned} R_i &:= A\left(\mathcal{B}_n \cup \{z\} \setminus X_i, X_i\right) \quad i = 1, \dots, n \\ R_{n+1} &:= A\left(\mathcal{B}_n, z\right) \\ \pi(z) &:= \frac{1}{n+1} \# \left\{ i = 1, \dots, n+1 : R_i \geq R_{n+1} \right\} \end{aligned}$$

- Para $\alpha \in [0, 1]$

$$\gamma^\alpha(X_1, \dots, X_n) := \{z \in \mathbf{Z} : \pi(z) > \alpha\}.$$

Proposición [Vovk et al. (2005)]

$$\mathbb{P}\Big(X_{n+1} \notin \gamma^\alpha(X_1, \dots, X_n)\Big) \leq \alpha.$$

En regresión

Dada $\mathcal{B}_n = \{Z_i = (X_i, Y_i)\}$ iid de $Z = (X, Y)$ buscamos $C_n(x)$, **condicionalmente válidos:**

$$\mathbb{P}\big(Y \in C_n(x) \mid X = x\big) \geq 1 - \alpha.$$

Si existe $p(y|x)$ estamos estimando

$$C_P(x) = \{y : p(y|x) \geq t^{\alpha}(x)\} \tag{3}$$

donde $t^{\alpha}(x)$ es tal que

$$\int \mathbf{1}_{\{p(y|x) \geq t^{\alpha}(x)\}} p(y|x) dy = 1 - \alpha.$$

Medidas de no conformidad

- $\hat{\eta}(x)$ un estimador de $\mathbb{E}(Y|X = x)$, en cuyo caso $R_i := \hat{\eta}_{\mathcal{B}_n \cup \{(x,y)\} \setminus Z_i}(X_i)$
- $\hat{p}(x)$ un estimador de $p_{Y|X}(x)$, en cuyo caso $R_i = \hat{p}_{\mathcal{B}_n \cup \{(x,y)\} \setminus Z_i}(X_i)$. Cholaquidis et al. (2024)

$$\gamma^{\alpha}(Z_1, \dots, Z_n) := \{y : \pi(x, y) \geq \alpha\}.$$

Bonnabel, S. and Sepulchre, R. (2010). Riemannian metric and geometric mean for positive semidefinite matrices of fixed rank. *SIAM Journal on Matrix Analysis and Applications*, 31(3):1055–1070.

Cholaquidis, A., Fraiman, R., Gamboa, F., and Moreno, L. (2022a). Weighted lens depth: Some applications to supervised classification. *The Canadian Journal of Statistics*.

Cholaquidis, A., Fraiman, R., and Moreno, L. (2022b). Level sets of depth measures in abstract spaces. *arxiv*.

Cholaquidis, A., Gamboa, F., and Moreno, L. (2024). Conformal inference for regression on riemannian manifolds.

Mardia, K. V., Jupp, P. E., and Mardia, K. (2000). *Directional statistics*, volume 2. Wiley Online Library.

Vovk, V., Gammerman, A., and Shafer, G. (2005). *Algorithmic learning in a random world*, volume 29. Springer.