Algunas Medidas de Profundidad y Aplicaciones

Marcela Svarc[1]
msvarc@udesa.edu.ar

[1]Universidad de San Andrés - CONICET

XV Congreso Dr. Antonio Monteiro

Resumen

Resumen

Introducción

Para una distribución P en $\mathbb{R}^d$, una función de profundidad, $D(x, P)$, provee un orden del centro hacia afuera para puntos $x \in \mathbb{R}^d$ basado en P.

Introducción

Para una distribución P en $\mathbb{R}^d$, una función de profundidad, $D(x, P)$, provee un orden del centro hacia afuera para puntos $x \in \mathbb{R}^d$ basado en P.

Interpretar "orden del centro hacia afuera" sugiere dos cosas

Introducción

Para una distribución P en $\mathbb{R}^d$, una función de profundidad, $D(x, P)$, provee un orden del centro hacia afuera para puntos $x \in \mathbb{R}^d$ basado en P.

Interpretar "orden del centro hacia afuera"sugiere dos cosas

- Una noción relevante de centro.

Introducción

Para una distribución P en $\mathbb{R}^d$, una función de profundidad, $D(x, P)$, provee un orden del centro hacia afuera para puntos $x \in \mathbb{R}^d$ basado en P.

Interpretar "orden del centro hacia afuera" sugiere dos cosas

- Una noción relevante de centro.
- Los puntos cerca del centro deberían tener mayor profundidad que los más lejanos.

Introducción

Para una distribución P en $\mathbb{R}^d$, una función de profundidad, $D(x, P)$, provee un orden del centro hacia afuera para puntos $x \in \mathbb{R}^d$ basado en P.

Interpretar "orden del centro hacia afuera" sugiere dos cosas

- Una noción relevante de centro.
- Los puntos cerca del centro deberían tener mayor profundidad que los más lejanos.

De estas consideraciones se desprende que el "centro" debe constar de los puntos que maximizen globalmente la función de profundidad.

Algunas definiciones

Profundidad del semiespacio (Tukey, 1975)

Sea $x \in \mathbb{R}^p$, la profundidad de x con respecto a una medida de probabilidad P en $\mathbb{R}^p$ se define como la mínima probabilidad que queda determinada por cualquier semiespacio cerrado que contenga x, es decir,

$$HD(x, P) = \text{ínf}\{P(H) : H \text{ es un semiespacio cerrado que contenga a } x\},$$

$x \in \mathbb{R}^p$.

Algunas definiciones

Profundidad del semiespacio (Tukey, 1975)

La versión muestral es la siguiente, sea $x \in \mathbb{R}^d$, la profundidad de x respecto de una muestra $X_1, X_2, \ldots, X_n$ también $\mathbb{R}^d$ es la menor fracción de puntos de la muestra que hay en un semiespacio cerrado que contenga a x, es decir,

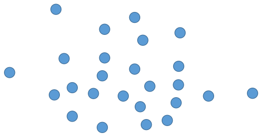
$$HD_n(x) = \frac{\min_{u \in \mathbb{R}^p} \sharp\{i : X_i' u \geq x' u\}}{n}.$$

En el caso real, tiene una formulación explícita y sencilla. Si consideramos P una medida de probabilidad en $\mathbb{R}$ y X una variable aletoria con función de distribución acumulada F, entonces

$$HD(x, P) = \min\{F(x), 1 - F(x^-)\}.$$

Algunas definiciones

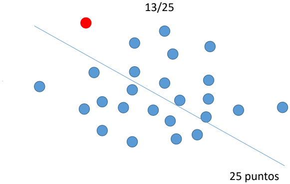
Profundidad del semiespacio (Tukey, 1975)



25 puntos

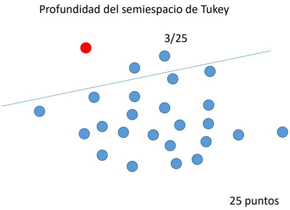
Algunas definiciones

Profundidad del semiespacio (Tukey, 1975)



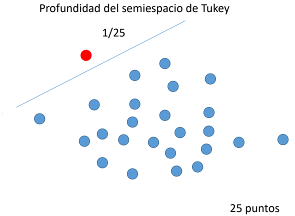
Algunas definiciones

Profundidad del semiespacio (Tukey, 1975)



Algunas definiciones

Profundidad del semiespacio (Tukey, 1975)



De este modo habría que probar con todos los hiperplanos que contengan al punto señalado

Algunas definiciones

Profundiad Simplicial (Liu, 1990)

Sea $x \in \mathbf{R}^p$, la profundidad de x con respecto a una medida de probabilidad P en $\mathbf{R}^p$ se define la probabilidad de que x pertenezca a un símplice aleatorio en $\mathbb{R}^p$, es decir,

$$SD(x, P) = P(x \in S[X_1, \ldots, X_{p+1}]), x \in \mathbf{R}^p,$$

donde $X_1, \ldots, X_{p+1}$ es una muestra aleatoria de con distribución P y $S[X_1, \ldots, X_{p+1}]$ denota un símplice aleatorio d-dimensional con vertices en $x_l, \ldots, x_{p+l}$, es decir, es el conjunto de todos los puntos de $\mathbb{R}^p$ que son combinaciones convexas de $x_1, \ldots, x_{d+l}$.

Algunas definiciones

Profundiad Simplicial (Liu, 1990)

Definición muestral, sea $X_1, \ldots, X_n$ una muestra aleatoria en $\mathbb{R}^p$, la profundidad simplicial de un punto $x \in \mathbb{R}^p$ es

$$SD_n(x) = \binom{n}{p+1}^{-1} \sum_{1 \le i_1 < i_2 < \cdots < i_{p+1} \le n} \mathcal{I}\{x \in S[X_{i_1}, \ldots, X_{i_{p+1}}]\},$$

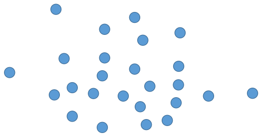
donde $\mathcal{I}$ es la función indicadora.

En el caso univariado, la función de profundidad simplicial es $SD(x, P) = F(x)(1 - F(x^-))$, donde F es la función de distribución acumulada.

Algunas definiciones

Profundiad Simplicial (Liu, 1990)

Profundidad del simplicial

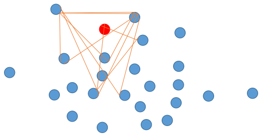


25 puntos

Algunas definiciones

Profundiad Simplicial (Liu, 1990)

Profundidad del simplicial



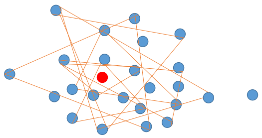
25 puntos

Y así continuar con todos los símplices posibles

Algunas definiciones

Profundiad Simplicial (Liu, 1990)

Profundidad del simplicial



25 puntos

Y así continuar con todos los símplices posibles

Propiedades

P.1. Invarianza Afín
La profundidad de un punto $x \in \mathbb{R}^d$ no debería depender del sistema de coordenadas, en particular de la escala del sistema de medición.

Propiedades

P.1. Invarianza Afín
La profundidad de un punto $x \in \mathbb{R}^d$ no debería depender del sistema de coordenadas, en particular de la escala del sistema de medición.

P.2. Maximalidad al centro
Para una distribución que tenga un único *centro* definido, por ejemplo, un punto de simetría, la profundidad se maximizaría en dicho punto.

Propiedades

P.3. Monotonía respecto del punto de máxima profundidad
A medida que $x \in \mathbb{R}^d$ se aleja del punto *más profundo* a lo largo de cualquier rayo fijo la profundidad de x debería decrecer.

Propiedades

P.3. Monotonía respecto del punto de máxima profundidad
A medida que $x \in \mathbb{R}^d$ se aleja del punto *más profundo* a lo largo de cualquier rayo fijo la profundidad de x debería decrecer.

P.4. Se anula en el infinito
La profundidad de un punto x debería aproximarse a cero si $\|x\|$ es demasiado grande.

Propiedades

Existen medidas de profundidad que no satisfacen **P.3.** y que en su lugar cumplen una condición de **cuasi-concavidad como función de x**, es decir que el conjunto $\{x : D(x; P) \geq c\}$ es convexo para cualquier número real c.

Propiedades

Existen medidas de profundidad que no satisfacen **P.3.** y que en su lugar cumplen una condición de **cuasi-concavidad como función de x**, es decir que el conjunto $\{x : D(x; P) \geq c\}$ es convexo para cualquier número real c.

Por otra parte, es bueno que las medidas de profundidad sean "continuas", es decir, que cumplan

P.5. Continuidad como función de x
Es esperable pensar que si dos puntos se encuentran lo suficientemente cerca las profundidades no sean demasiado distintas

Propiedades

Existen medidas de profundidad que no satisfacen **P.3.** y que en su lugar cumplen una condición de **cuasi-concavidad como función de x**, es decir que el conjunto $\{x : D(x; P) \geq c\}$ es convexo para cualquier número real c.

Por otra parte, es bueno que las medidas de profundidad sean "continuas", es decir, que cumplan

P.5. Continuidad como función de x
Es esperable pensar que si dos puntos se encuentran lo suficientemente cerca las profundidades no sean demasiado distintas

P.6. Continuidad como función de P.
Análogo a la propiedad anterior para distribuciones.

Definición

Definición (Función de Profundidad)

Sea $D(\cdot,\cdot) : \mathbb{R}^d \times \mathbb{P} \to \mathbb{R}$ *una función acotada y no negativa que satisface:*

1. $D(Ax+b, P_{AX+b}) = D(x, P_X)$ *para todo vector aleatorio* $X \in \mathbb{R}^d$ *vector aleatorio, para todo matriz* $A \in R^{dxd}$ *no singular y para cualquier vector* $b \in \mathbb{R}^d$.
2. $D(\theta, P) = \sup_{x \in R^d} D(x, P),$ *para cualquier* $P \in \mathbb{P}$ *con centro* θ.
3. *Para toda* $P \in \mathbb{P}$ *con punto más profundo* θ $D(x, P) \leq D(\theta + \alpha(x-\theta), P)$, *para cualquier* $\alpha \in [0,1]$.
4. $\lim_{\|x\| \to +\infty} D(x, P) = 0,$ *para todo* $P \in \mathbb{P}$.

Definición

Si además quisieramos que se cumplieran **P.5.** y **P.6.**, deberíamos pedir que

5. Fijada $P \in \mathbb{P}$ $D(\cdot, P) : \mathbb{R}^d \to \mathbb{R}$ es continua.
6. Fijado $x \in \mathbb{R}^d$ $D(x, \cdot) : \mathbb{P} \to \mathbb{R}$ es continua.

Definición

Definición (Simetría)

Sea $X \in \mathbb{R}^d$ *un vector aleatorio decimos que:*

- *Es* **Centralmente simétrico** *(C-simétrico) respecto de* θ *si la distribución de* $X - \theta$ *es la misma que la de* $\theta - X$*.*

Definición

Definición (Simetría)

Sea $X \in \mathbb{R}^d$ *un vector aleatorio decimos que:*

- *Es* **Centralmente simétrico** *(C-simétrico) respecto de* θ *si la distribución de* $X - \theta$ *es la misma que la de* $\theta - X$.
- *Es* **Angularmente simétrico** *(A-simétrico) respecto de* θ *si* $(X - \theta)/\|X - \theta\|$ *es centralmente simétrico respecto del origen.*

Definición

Definición (Simetría)

Sea $X \in \mathbb{R}^d$ *un vector aleatorio decimos que:*

- *Es* **Centralmente simétrico** *(C-simétrico) respecto de* θ *si la distribución de* $X - \theta$ *es la misma que la de* $\theta - X$*.*
- *Es* **Angularmente simétrico** *(A-simétrico) respecto de* θ *si* $(X - \theta)/\|X - \theta\|$ *es centralmente simétrico respecto del origen.*
- *Es* **Semiespacialmente simétrico** *(H-simétrico) respecto de* θ *si* $P(X \in H) \geq \frac{1}{2}$ *para todo* H *semiespacio cerrado tal que* $\theta \in H$*.*

Definición

Definición (Simetría)

Sea $X \in \mathbb{R}^d$ un vector aleatorio decimos que:

- *Es* **Centralmente simétrico** *(C-simétrico) respecto de θ si la distribución de $X - \theta$ es la misma que la de $\theta - X$.*
- *Es* **Angularmente simétrico** *(A-simétrico) respecto de θ si $(X - \theta)/\|X - \theta\|$ es centralmente simétrico respecto del origen.*
- *Es* **Semiespacialmente simétrico** *(H-simétrico) respecto de θ si $P(X \in H) \geq \frac{1}{2}$ para todo H semiespacio cerrado tal que $\theta \in H$.*

Observación

Es claro que C-simétrico $\Rightarrow$ A-simétrico $\Rightarrow$ H-simétrico.

Algunas definiciones

Funciones de Profundidad Tipo A

Dados $x_1, \ldots, x_p$ puntos en $\mathbb{R}^p$, y una función de proximidad entre un punto $x \in \mathbb{R}^p$ y a los puntos $x_1, \ldots, x_r$, $h(x; x_1, \ldots, x_r)$, donde esta función es nonegativa y acotada, se define la profundidad como

$$D_A(x, P) = E(h(x; X_1, \ldots, X_r)).$$

Algunos ejemplos son la profundidad simplicial donde $h(x; x_1, \ldots, x_r) = I_{x \in S[x_1, \ldots, x_p + 1}$.

Algunas definiciones

Funciones de Profundidad Tipo A

La profundidad mayorante (Singh, 1991) dados $x_1, \ldots, x_p$ puntos en $\mathbb{R}^p$ que determinan un único hiperplano que lo contengan, a partir de el mismo quedan determinados dos semi-espacios cerrados con este hiperplano como límite. Denotamos por $H^P_{x_1,\ldots,x_p}$ al semiespacio que tenga probabilidad (P) mayor o igual que 0.5, luego la profundidad mayorante está dada por

$$MJD(x, P) = P(x \in H^P_{X_1,\ldots,X_p}), x \in \mathbb{R}^p,$$

donde $X_1, \ldots, X_p$ es una muestra aleatoria de la distribución P. En este caso $h(x; x_1, \ldots, x_r) = I_{x \in H^P_{x_1,\ldots,x_p}}$.

Algunas definiciones

Funciones de profundidad tipo B

Sea $h(x; x_1, \dots, x_r)$ una función no negativa y no acotada que mide en cierto modo distancia entre el punto $x \in \mathbb{R}^p$ y $x_1, \dots, x_r$, se definen las funciones de profundidad tipo B como

$$D_b(x, P) = (1 + E(h(x; X_1, \dots, X_r)))^{-1},$$

donde $X_1, \dots, X_r$ es una muestra aleatoria de P. Por ejemplo, consideramos la *Profundidad simplicial de volumen*, sea $h(x; x_1, \dots, x_p) = \left(\frac{\Delta(S[x; X_1, \dots, X_p])}{det(\sqrt{\Sigma})}\right)^{\alpha}$, donde $X_1, \dots, X_p$ es una muestra aleatoria de P, Σ es la matriz de covarianza de P, $\Delta(S[x; X_1, \dots, X_p])$ es el volumen del simplex p dimensional y $\alpha > 0$.

Algunas definiciones

Medidas de profundidad de tipo C

Sea $O(x, P)$ una medida de atipicidad de x con respecto al centro o al punto más profundo de P, se puede definir una medida de profundidad del siguiente modo

$$D_C(x, P) = (1 + O(x, P))^{-1}$$

Específicamente, si $O(x, P) = \sup_{\|u\|=1} \frac{|u'x - med(u'X)|}{MAD(u'X)}$, donde X tiene distribución P, *med* es la mediana unidimensional y *MAD* es la desviación absoluta mediana entonces queda definida la *profundidad de proyección*

Algunas definiciones

Medidas de profundidad de tipo D

Sea $\mathcal{C}$ una clase de subconjuntos cerrados de $\mathbb{R}^p$ y P una medida de probabilidad en $\mathbb{R}^p$, se define una medida de probabilidad tipo D como

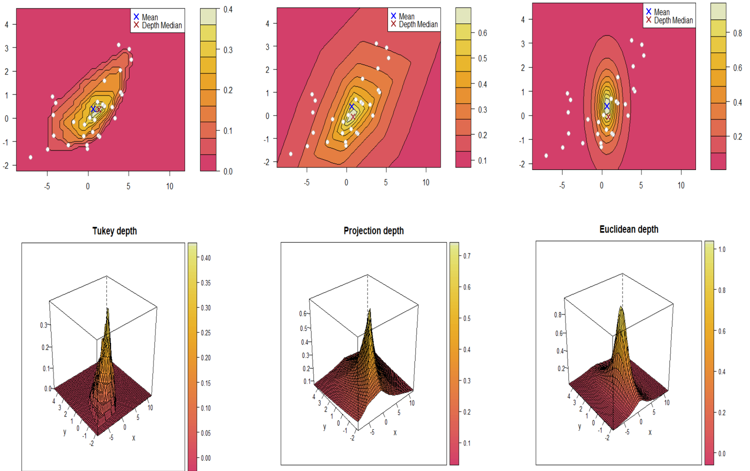
$$D_D(x, P; \mathcal{C}) = \inf_{\mathcal{C}} \{P(C) \text{ tal que } x \in C \in \mathcal{C}\},$$

la profundidad de un punto x es la mínima probabilidad que puede tener un conjunto $C \in \mathcal{C}$ que lo contenga. La clase de conjuntos $\mathcal{C}$ debe satisfacer las siguientes propiedades

- Si $C \in \mathcal{C}$ luego $\overline{C^c} \in \mathcal{C}$.
- Para $C \in \mathcal{C}$ y $x \in C^\circ$ existe $C_1 \in \mathcal{C}$ tal que $x \in \partial C_1 \subset C^\circ$.

Específicamente, si $\mathcal{C}$ es la clase de los subespacios cerrados entonces la profundidad del semiespacio pertenece a esta familia.

Algunos ejemplos



Profundidad de Tukey Aleatorizada

Un problema que presentan las profundidades es su alto costo computacional, posibles soluciones son computarlas en forma aproximada, esto suele hacerse aleatoriazando. Por ejemplo, Cuesta-Albertos y Nieto-Reyes (2008) definen la Profundidad de Tukey Aleatorizada para datos multivariados.

Definición

Sea $P \in \mathbb{P}$. Sea $\upsilon \in \mathbb{P}$ absolutamente continua y sean $\upsilon_1, \ldots, \upsilon_k$ vectores aleatorios i.i.d. con distribución υ. La Profundidad de Tukey Aleatorizada de $x \in \mathbb{R}^p$ con respecto a P basada en k vectores aleatorios elegidos con υ es

$$D_{T,k,\upsilon}(x, P) = min\{D_1(\Pi_{\upsilon_i}(x), P \circ \Pi_{\upsilon_i}^{-1}) : 1 \leq i \leq k\}, \ x \in \mathbb{R}^p.$$

Propiedades

	Hiperplano	**Simplicial**	**Mayorante**	**Proyecciones**	**Hiperplano aleatorizada**
P. 1. Inv. Afín	✓	✓	✓	✓	✓
P. 2. Max. En en centro	✓	✓	✓	✓	✓
P. 3. Monoton. centro	✓	✓	✓	✓	✓
P. 4. Se anula en ∞	✓	✓	✗	✓	✓

Propiedades

	Hiperplano	**Simplicial**	**Mayorante**	**Proyecciones**	**Hiperplano aleatorizada**
P. 1. Inv. Afín	✓	✓	✓	✓	✓
P. 2. Max. En en centro	✓	✓ Cont A-sim	✓	✓	✓
P. 3. Monoton. centro	✓	✓ Cont A-sim	✓	✓	✓ En prob
P. 4. Se anula en ∞	✓	✓	✗	✓	✓

Propiedades

	Hiperplano	Simplicial	Mayorante	Proyecciones	Hiperplano aleatorizada
P. 1. Inv. Afín	✓	✓	✓	✓	✓
P. 2. Max. En en centro	✓	✗ Disc. C-sim	✓	✓	✓
P. 3. Monoton. centro	✓	✗ Disc. C-sim	✓	✓	✓ En prob
P. 4. Se anula en ∞	✓	✓	✗	✓	✓

Resumen

Aplicaciones

- Regiones de profundidad: Serfling (2004).
- Tests de posición y diferencias de escala: Liu et al. (1999) y Li y Liu (2004).
- Diagnóstico de no normalidad: Liu et al. (1999).
- Detección de outliers: Chen et al. (2009).
- Clasificación: Ghosh y Chaudhuri (2005), Li et al. (2012), Cuevas y Fraiman (2009), Dutta y Ghosh (2012), Paindaveine and Van Bever (2012), Cuesta-Albertos et al. (2017), entre otros.

Marco general

- Muestra de entrenamiento
  - observaciones, variables, características registradas.
  - etiquetas indicando la naturaleza de las observaciones, es decir, a que categoría pertenecen.

Marco general

- Muestra de entrenamiento
  - observaciones, variables, características registradas.
  - etiquetas indicando la naturaleza de las observaciones, es decir, a que categoría pertenecen.
- Muestra de clasificación
  - observaciones, variables, características registradas.
  - etiquetas desconocidas.

El clasificador

- $x \in E$,
- $y \in \{1, ..., g\}$

Clasificador

$g : E \to \{1, ..., g\}$

Reglas de clasificación

- Análisis Discriminante de Fisher
- k-Vecinos más cercanos
- CART
- Support Vector Machine

Reglas de clasificación

- Análisis Discriminante de Fisher
- k-Vecinos más cercanos
- CART
- Support Vector Machine
- DD-clasificador

Clasificador de Máxima Profundidad

La idea es la siguiente:

Dadas dos probabilidades P, Q y D una función de profundidad se clasifica a un punto x como proveniente de P si $D(x, P) > D(x, Q)$.

El procedimiento se encuentra desarrollado en su totalidad en Ghosh y Chaudhuri (2005).

El DD-plot es una herramienta para comparar dos distribuciones o muestras multivariadas; introducido por Liu et al (1999).

Consideramos dos conjuntos de datos $\{x_1, \ldots, x_n\}$ e $\{y_1, \ldots, y_m\}$ generados por las distribuciones P y Q en $\mathbb{R}^p$ respectivamente.

Un DD-plot es un gráfico de dispersión entre las profundidades de las observaciones $\{x_1, \ldots, x_n, y_1, \ldots, y_m\}$ respecto de la profundidad P versus las profundidades del mismo conjunto de observaciones respecto de la profundidad Q, donde las profundidades provenientes de cada muestra deben ser distinguidas, por ejemplo utilizando distinto color.

Sin importar la dimensión del espacio al cual pertenezcan las observaciones, el DD-plot siempre es un gráfico bidimensional.

Ejemplo

Generamos un conjunto de datos de 400 puntos a partir de dos distribuciones distintas:

- $P \sim N_2(\mu_1, I_2)$, donde $\mu_1 = (0, 0)$ y I_2 es la matriz identidad de 2×2.
- $Q \sim N_2(\mu_2, I_2)$, donde $\mu_2 = (2, 2)$.

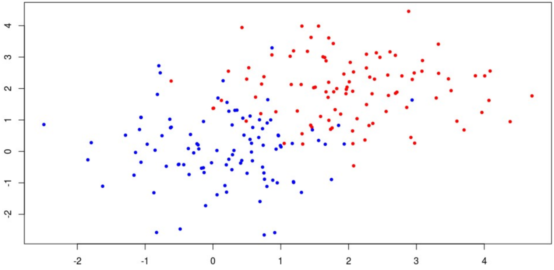


Figura: Rojo: P. Azul: Q

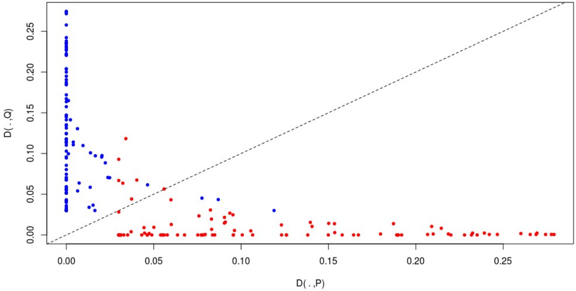


Figura: Rojo: P. Azul: Q

Ejemplo

Generamos un conjunto de datos de 200 puntos a partir de dos distribuciones distintas:

- $P \sim N_2(\mu, I_1)$, donde $\mu = (0, 0)$ y I_1 es la matriz identidad de 2×2.
- $Q \sim N_2(\mu, I_2)$, donde $I_2 = 4I_1$.

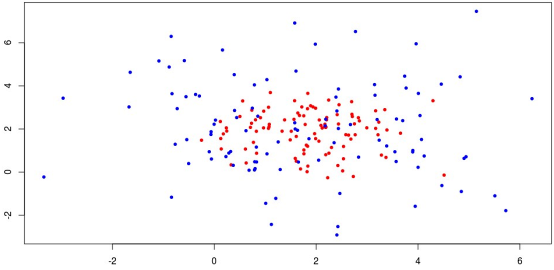


Figura: Rojo: P. Azul: Q

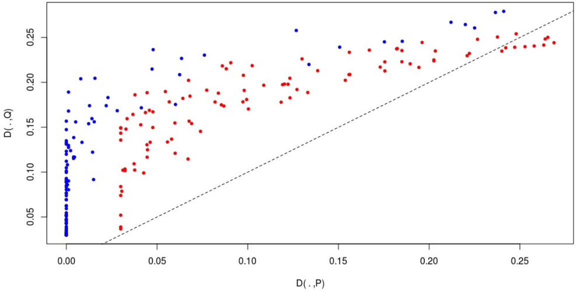


Figura: Rojo: P. Azul: Q

DD^G – clasificador

Cuesta-Albertos et al (2015) extienden la clasificación via profundidades a más de un grupo. Supongamos que tenemos un espacio χ donde tenemos g clases o grupos y D una medida de profundidad, queremos clasificar $x \in \chi$, luego

$$\chi \to \mathbb{R}^g$$

$$x \to (D^1(x), \ldots, D^g(x)),$$

ahora podemos aplicar algún clasificador en el espacio g-dimensional.

- Cómo elegir la profundidad?
- Elegir una clasificador, para el problema en dimensión baja.

Ejemplos

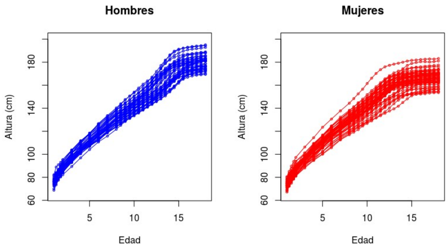


Figura: Berkeley Growth Study

Ejemplo

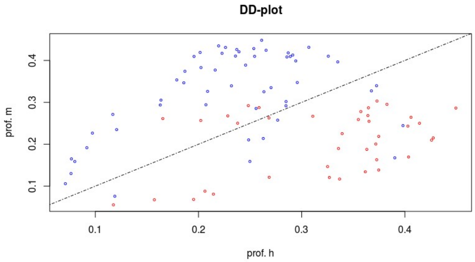


Figura: Berkeley Growth Study DD-plot

Ejemplo

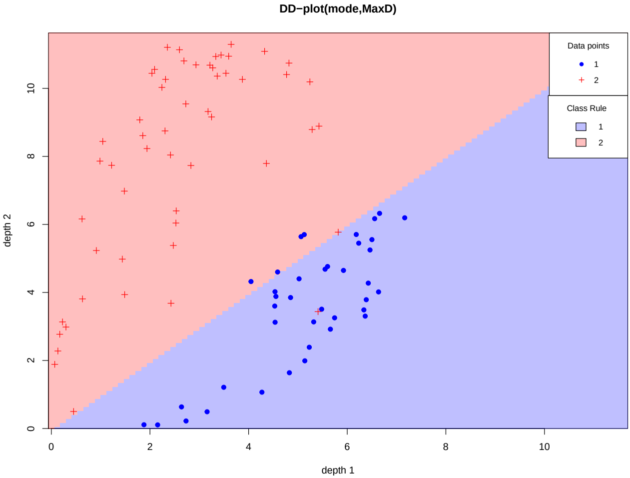


Figura: Berkeley Growth Study DD-plot

Aplicación

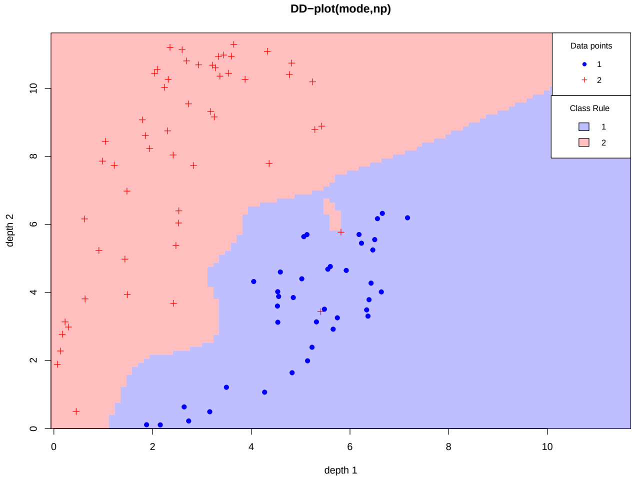


Figura: Berkeley Growth Study DD-plot

DD^G – *clasificador*

En el mismo trabajo, se extiende esta idea para poder utilizar varios clasificadores a la vez. Por ejemplo, si tenemos tres grupos ($g = 3$) y dos clasificadores, entonces

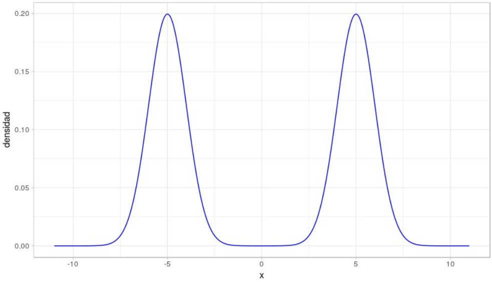
$$x \to (D_1^1(x), D_1^2(x), D_1^3(x), D_2^1(x), D_2^2(x), D_2^3(x)),$$

y realizamos la clasificación en el espacio de dimensión $3 \times 2 = 6$, sin importar la dimensión del espacio original.

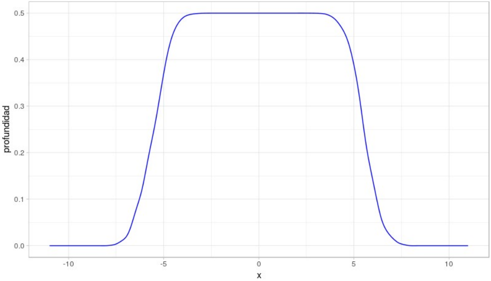
Resumen

Profundidades Locales

Las profundidades globales no pueden captar la estructura en datos con soporte que no sea convexo ni cuando hay multiples centros.



Las profundidades globales no pueden captar la estructura en datos con soporte que no sea convexo ni cuando hay multiples centros.



Con el espíritu de captar multiples centros, su localización e importancia relativa; en los últimos años se han definido profundidades locales:

- **2011:** Agostinelli, C. and Romanazzi, M., "Local Depth."
- **2013:** Paindavaine, D. and Van Bever, "From depth to local depth: A focus in centrality."
- **2018:** Agostinelli, C., "Local half-region depth for functional data."
- **2018:** Fernandez-Piana, L. and Svarc, M. "A Local Delpth Measure for General Data."

Profundidad local de Paindavaine-Van Bever

Sea X variable aleatoria continua con función de distribución acumulada F, sea $x \in \mathbb{R}$ y sea $\beta \in (0, 1]$.

Sea $\lambda_x^\beta = inf\{\lambda > 0 : F(x+\lambda) - F(x-\lambda) \geq \beta\}$.

Definición

Profundidad Local Simplicial Univariada,

$$LD_S^\beta(x, F) = \frac{2}{\beta^2}(F(x+\lambda_x^\beta) - F(x))(F(x) - F(x-\lambda_x^\beta)).$$

Profundidad local de Paindavaine-Van Bever

Sea X variable aleatoria continua con función de distribución acumulada F, sea $x \in \mathbb{R}$ y sea $\beta \in (0, 1]$.

Sea $\lambda_x^\beta = inf\{\lambda > 0 : F(x + \lambda) - F(x - \lambda) \geq \beta\}$.

Definición

Profundidad Local Simplicial Univariada,

$$LD_S^\beta(x, F) = \frac{2}{\beta^2}(F(x + \lambda_x^\beta) - F(x))(F(x) - F(x - \lambda_x^\beta)).$$

Análogamente podemos tener una versión local para la Profundidad del Semiespacio,

$$LHD^\beta(x, F) = \frac{1}{\beta} min(F(x + \lambda_x^\beta) - F(x), F(x) - F(x - \lambda_x^\beta)),$$

Profundidad Local Dual Integrada

Definición

Sea Ω un espacio de probabilidad y $\mathbb{E}$ un espacio de Banach separable, notemos $\mathbb{E}'$ al dual del Banach. Sea $X : \Omega \to \mathbb{E}$ un elemento aleatorio en $\mathbb{E}$ con distribución P y Q una medida de probabilidad sobre $\mathbb{E}'$ independiente de P, $\beta \in (0, 1]$ y $x \in \mathbb{E}$. Definimos la **Profundidad Local Dual Integrada** *(IDLD),*

$$IDLD^{\beta}(x, P) = \int LD_S^{\beta}(f(x), P_f)dQ(f),$$

donde LD_S^{β} es la profundidad local simplicial univariada, $f \in \mathbb{E}'$ y P_f es la distribución univariada de $f(X)$.

IDLD

Propiedades

P.1. Sea $\mathbb{E} = \mathbb{R}^d$ con la norma euclídea y $X \in \mathbb{E}$ un vector aleatorio, sea Q la medida uniforme en la esfera unidad de $\mathbb{E}'$ independiente de P_X. Sea $A : \mathbb{E} \to \mathbb{E}$ una transformación ortogonal y $\beta \in (0, 1]$. Entonces $IDLD^{\beta}(Ax, P_{AX}) = IDLD^{\beta}(x, P_X)$.

IDLD

Propiedades

P.1. Sea $\mathbb{E} = \mathbb{R}^d$ con la norma euclídea y $X \in \mathbb{E}$ un vector aleatorio, sea Q la medida uniforme en la esfera unidad de $\mathbb{E}'$ independiente de P_X. Sea $A : \mathbb{E} \to \mathbb{E}$ una transformación ortogonal y $\beta \in (0, 1]$. Entonces $IDLD^{\beta}(Ax, P_{AX}) = IDLD^{\beta}(x, P_X)$.

Se puede definir una versión de la IDLD que sea invariante afín si la X es un vector aleatorio con matriz de covarianza Σ_X no singular.

$$IDLD^{\beta}_{af}(x, P_X) = IDLD^{\beta}\left(\Sigma_X^{-1/2}x, P_{\Sigma_X^{-1/2}X}\right).$$

IDLD

Propiedades

Definición

Sea X una variable aleatoria en la recta real y $\beta \in (0,1]$, decimos que X es β-simétrico en θ si la distribución acumulada F satisface que,

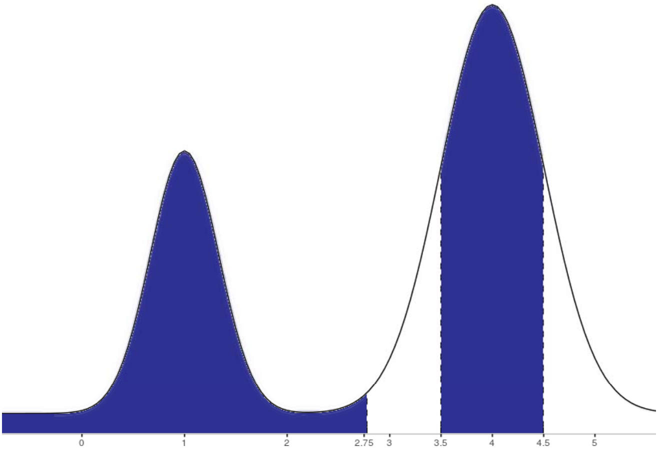
$$F\left(\theta + \lambda_{\theta}^{\beta'}\right) - F(\theta) = \frac{\beta'}{2}, \text{ para todo } 0 < \beta' \leq \beta.$$

Un elemento aleatorio X en un espacio de Banach $\mathbb{E}$ es β-simétrico en θ si para todo $f \in \mathbb{E}'$ no nulo,

$$F_f\left(f(\theta) + \lambda_{f(\theta)}^{\beta'}\right) - F_f\left(f(\theta)\right) = \frac{\beta'}{2}, \text{ para todo } 0 < \beta' \leq \beta.$$

IDLD

Propiedades



IDLD

Propiedades

P.2. Sea $X \in \mathbb{E}$ un elemento aleatorio continuo y β-simétrico en θ. Para $\beta \in (0,1]$ tenemos que,

$$IDLD^{\beta'}(\theta, P_X) = \max_{x\in\mathbb{E}} IDLD^{\beta'}(x, P_X), \text{ para todo } 0 < \beta' \leq \beta.$$

IDLD

Propiedades

P.2. Sea $X \in \mathbb{E}$ un elemento aleatorio continuo y β-simétrico en θ. Para $\beta \in (0,1]$ tenemos que,

$$IDLD^{\beta'}(\theta, P_X) = \max_{x \in \mathbb{E}} IDLD^{\beta'}(x, P_X), \text{ para todo } 0 < \beta' \leq \beta.$$

Propiedad: Sea $X \in \mathbb{E}$ un elemento aleatorio continuo C-simétrico en θ, entonces X es β-simétrico en θ para cada $\beta \in (0,1]$.

IDLD

Propiedades

P.3. Sea $\mathbb{E}$ un espacio de Banach separable con dual $\mathbb{E}'$. Sea X un elemento aleatorio continuo C-simétrico en $\theta \in \mathbb{E}$ con medida de probabilidad asociada P. Sea Q una medida de probabilidad en $\mathbb{E}'$ independiente de P y supongamos que para todo $f \in \mathbb{E}'$ se cumple que $f(X)$ tiene función de densidad unimodal en $f(\theta)$. Entonces, bajo condiciones de regularidad, para todo $x \in \mathbb{E}$ y $\beta \in (0,1]$,

$$IDLD^{\beta}(x,P) \leq IDLD^{\beta}((1-t)\theta + xt, P) \quad \text{para todo } t \in [0,1].$$

IDLD

Propiedades

P.4. Supongamos que

$$\sup_{\|u\|=1} \{f : f(u) \leq \epsilon\} = O(\epsilon),$$

donde $O(\epsilon)$ es una función tal que $O(\epsilon) \xrightarrow[\epsilon \to 0]{} 0$ luego,

$$IDLD^{\beta}(x, P) \xrightarrow[||x|| \to +\infty]{} 0.$$

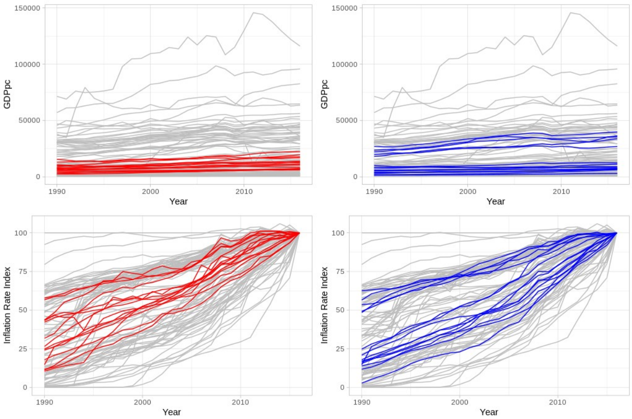
IDLD

Propiedades

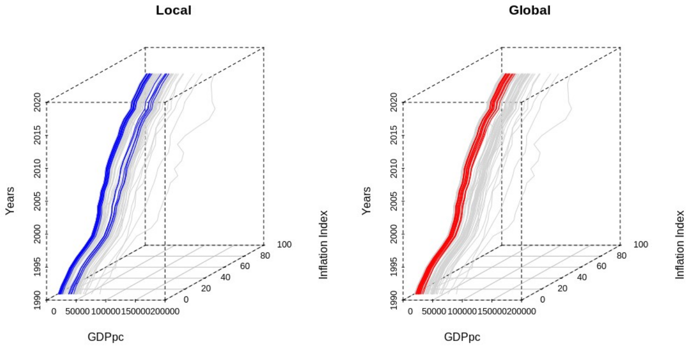
P.5. Sea $X \in \mathbb{E}$ (separable) un elemento aleatorio continuo con medida de probabilidad asociada P y $\beta \in (0,1]$, entonces $IDLD^{\beta}(\cdot, P) : \mathbb{E} \to \mathbb{R}$ es continua.

P.6. Sea E un Banach separable y sea $(P_n)_{n \geq 1}$ una sucesión de medidas de probabilidad definidas en E tales que P_n converge debilmente a P. Entonces, para todo $\beta \in (0,1]$, $IDLD^{\beta}(x, P_n) \xrightarrow[n \to +\infty]{} IDLD^{\beta}(x, P)$.

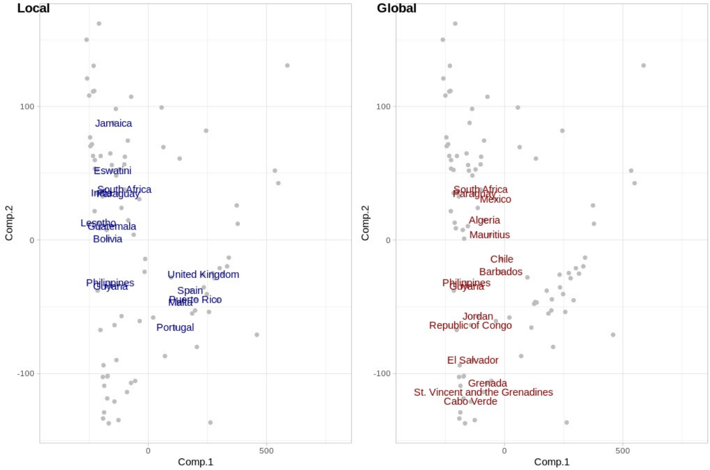
Ejemplo: Inflación y PBI



Ejemplo: Inflación y PBI



Ejemplo: Inflación y PBI



Growth data set, continuación