

Detección de anomalías basada en técnicas de ensembles

Ignacio Aguilera Martos

8 de septiembre de 2019

Trabajo Fin de Grado

Código disponible en [GitHub](#)

1. Concepto de anomalía basado en distancias
2. Aprendizaje Automático
3. Probabilidad Multivariante
4. Concepto probabilístico de anomalía
5. Modelos implementados
6. Resultados
7. Conclusiones y Trabajo Futuro

Concepto de anomalía basado en distancias

Tukey's Fences

Tukey's Fences

Pensado para el caso uno-dimensional.

Tukey's Fences

Pensado para el caso uno-dimensional.

Tukey's Fences

Valores fuera del rango $[Q_1 - k(Q_3 - Q_1), Q_3 + k(Q_3 - Q_1)]$ con $k = 1,5$

Tukey's Fences

Pensado para el caso uno-dimensional.

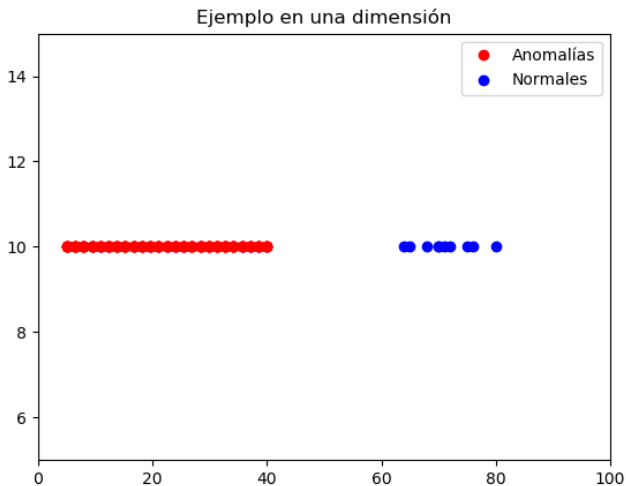
Tukey's Fences

Valores fuera del rango $[Q_1 - k(Q_3 - Q_1), Q_3 + k(Q_3 - Q_1)]$ con $k = 1,5$

La propuesta de $k = 1,5$ es arbitraria.

Tukey's Fences

Tukey's Fences



Extensión al caso de mayor dimensionalidad

Extensión al caso de mayor dimensionalidad

Criterio

Aplicar el criterio de Tukey a cada una de las características.

Extensión al caso de mayor dimensionalidad

Criterio

Aplicar el criterio de Tukey a cada una de las características. **Trivial**.

Extensión al caso de mayor dimensionalidad

Criterio

Aplicar el criterio de Tukey a cada una de las características. **Trivial**.

Criterio de clusters

1. Agrupamos los datos por clusters.

Extensión al caso de mayor dimensionalidad

Criterio

Aplicar el criterio de Tukey a cada una de las características. **Trivial**.

Criterio de clusters

1. Agrupamos los datos por clusters.
2. Encontramos el cluster más cercano para cada instancia.

Extensión al caso de mayor dimensionalidad

Criterio

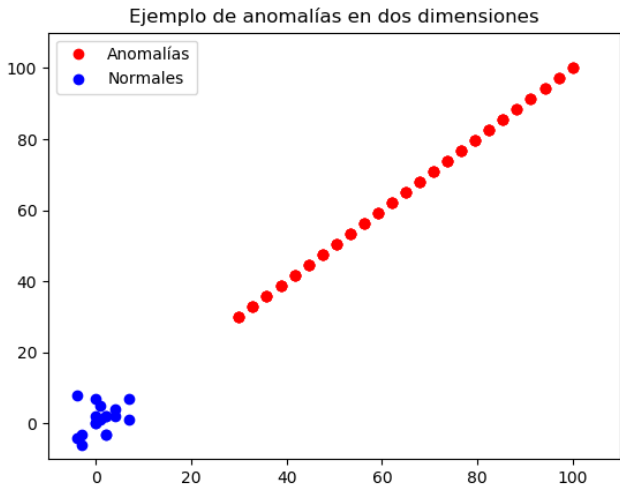
Aplicar el criterio de Tukey a cada una de las características. **Trivial**.

Criterio de clusters

1. Agrupamos los datos por clusters.
2. Encontramos el cluster más cercano para cada instancia.
3. Si la distancia del objeto al centroide del cluster es mayor que 1,5 veces la mayor distancia intercluster entonces es una anomalía.

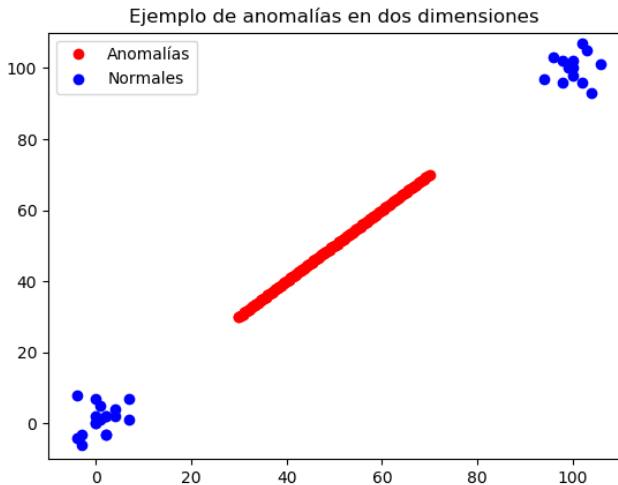
Ejemplo 1

Ejemplo 1



Ejemplo 2

Ejemplo 2



Aprendizaje Automático

Problema que abordamos

Problema que abordamos

Problema de detección de anomalías

Es un problema de aprendizaje no supervisado pues no disponemos de las etiquetas.

Problema que abordamos

Problema de detección de anomalías

Es un problema de aprendizaje no supervisado pues no disponemos de las etiquetas.

Problema no supervisado y acierto

Es un problema no supervisado. Usaremos conjuntos de datos etiquetados.

Propiedades de conjuntos de alta dimensionalidad

1. La densidad disminuye exponencialmente al aumentar la dimensionalidad.

Propiedades de conjuntos de alta dimensionalidad

1. La densidad disminuye exponencialmente al aumentar la dimensionalidad.
2. Cuanto mayor es la dimensionalidad mayor debe ser el radio de una bola para englobar el mismo porcentaje de datos.

Propiedades de conjuntos de alta dimensionalidad

1. La densidad disminuye exponencialmente al aumentar la dimensionalidad.
2. Cuanto mayor es la dimensionalidad mayor debe ser el radio de una bola para englobar el mismo porcentaje de datos.
3. Casi todo punto está más cerca del borde del conjunto que de otro punto.

$$D(d, n) = \left(1 - \frac{1}{2^{\frac{1}{n}}}\right)^{\frac{1}{d}}$$

Propiedades de conjuntos de alta dimensionalidad

1. La densidad disminuye exponencialmente al aumentar la dimensionalidad.
2. Cuanto mayor es la dimensionalidad mayor debe ser el radio de una bola para englobar el mismo porcentaje de datos.
3. Casi todo punto está más cerca del borde del conjunto que de otro punto.

$$D(d, n) = \left(1 - \frac{1}{2^{\frac{1}{n}}}\right)^{\frac{1}{d}}$$

4. Casi todo punto es una anomalía sobre su propia proyección.

Propiedades de conjuntos de alta dimensionalidad

1. La densidad disminuye exponencialmente al aumentar la dimensionalidad.
2. Cuanto mayor es la dimensionalidad mayor debe ser el radio de una bola para englobar el mismo porcentaje de datos.
3. Casi todo punto está más cerca del borde del conjunto que de otro punto.

$$D(d, n) = \left(1 - \frac{1}{2^{\frac{1}{n}}}\right)^{\frac{1}{d}}$$

4. Casi todo punto es una anomalía sobre su propia proyección.

Maldición de la alta dimensionalidad

A mayor dimensionalidad mayor número de puntos necesitamos para obtener una aproximación con funciones de igual regularidad.

Aproximación de funciones

Nuestro objetivo es aproximar la función de salida del sistema con los datos que tenemos.

Aproximación de funciones

Aproximación de funciones

Nuestro objetivo es aproximar la función de salida del sistema con los datos que tenemos.

Teoremas útiles

- Teorema de Aproximación de Weierstrass

Aproximación de funciones

Aproximación de funciones

Nuestro objetivo es aproximar la función de salida del sistema con los datos que tenemos.

Teoremas útiles

- Teorema de Aproximación de Weierstrass
- Serie de Fourier

Aproximación de funciones

Nuestro objetivo es aproximar la función de salida del sistema con los datos que tenemos.

Teoremas útiles

- Teorema de Aproximación de Weierstrass
- Serie de Fourier
- Teorema de Kolmogorov-Arnold

Teorema de Kolmogorov-Arnold

Teorema de Kolmogorov-Arnold

Teorema de Superposición Kolmogorov-Arnold

Sea f una función continua de varias variables $f : X_1 \times \dots \times X_n \rightarrow \mathbb{R}$, entonces existen funciones $\Phi_q : \mathbb{R} \rightarrow \mathbb{R}$ y $\phi_{q,p} : X_p \rightarrow [0, 1]$ tales que f se puede expresar como:

$$f(x) = f(x_1, \dots, x_n) = \sum_{q=0}^{2n} \Phi_q \left(\sum_{p=1}^n \phi_{q,p}(x_p) \right)$$

Teorema de Kolmogorov-Arnold

Teorema de Superposición Kolmogorov-Arnold

Sea f una función continua de varias variables $f : X_1 \times \dots \times X_n \rightarrow \mathbb{R}$, entonces existen funciones $\Phi_q : \mathbb{R} \rightarrow \mathbb{R}$ y $\phi_{q,p} : X_p \rightarrow [0, 1]$ tales que f se puede expresar como:

$$f(x) = f(x_1, \dots, x_n) = \sum_{q=0}^{2n} \Phi_q\left(\sum_{p=1}^n \phi_{q,p}(x_p)\right)$$

Dimensionalidad en datos y funciones

La maldición de la dimensionalidad es intrínseca a los datos. Este Teorema nos dice que la capacidad expresiva o complejidad de las funciones de una sola variable es la misma que las de varias variables.

Probabilidad Multivariante

Contenidos de la sección

Contenidos útiles

- Vectores aleatorios.

Contenidos útiles

- Vectores aleatorios.
- Independencia.

Contenidos útiles

- Vectores aleatorios.
- Independencia.
- Probabilidad condicionada (sucesos, variables y σ -álgebras) y sus propiedades.

Contenidos útiles

- Vectores aleatorios.
- Independencia.
- Probabilidad condicionada (sucesos, variables y σ -álgebras) y sus propiedades.
- Esperanza condicionada (sucesos, variables y σ -álgebras) y sus propiedades.

Contenidos útiles

- Vectores aleatorios.
- Independencia.
- Probabilidad condicionada (sucesos, variables y σ -álgebras) y sus propiedades.
- Esperanza condicionada (sucesos, variables y σ -álgebras) y sus propiedades.
- Desigualdades famosas.

Desigualdades estudiadas

- Desigualdad de Markov
- Desigualdad de Chebychev
- Cotas de Chernoff
- Desigualdad de Hoeffding

Concepto probabilístico de anomalía

Notación usada

$$X = \{x_1, \dots, x_n\}, \quad x_i = (x_{s_1}, \dots, x_{s_d})$$

Notación usada

$$X = \{x_1, \dots, x_n\}, \quad x_i = (x_{s_1}, \dots, x_{s_d})$$

$$S = \{s_i | s_i \in \{s_1, \dots, s_d\} \text{ con } i \in \Delta\}$$

Notación usada

$$X = \{x_1, \dots, x_n\}, \quad x_i = (x_{s_1}, \dots, x_{s_d})$$

$$S = \{s_i | s_i \in \{s_1, \dots, s_d\} \text{ con } i \in \Delta\}$$

X_S proyección de los datos en el subespacio S

Notación usada

$$X = \{x_1, \dots, x_n\}, \quad x_i = (x_{s_1}, \dots, x_{s_d})$$

$$S = \{s_i | s_i \in \{s_1, \dots, s_d\} \text{ con } i \in \Delta\}$$

X_S *proyección de los datos en el subespacio S*

$$p_{s_1, \dots, s_p}(x_{s_1}, \dots, x_{s_p})$$

Notación usada

$$X = \{x_1, \dots, x_n\}, \quad x_i = (x_{s_1}, \dots, x_{s_d})$$

$$S = \{s_i | s_i \in \{s_1, \dots, s_d\} \text{ con } i \in \Delta\}$$

X_S *proyección de los datos en el subespacio S*

$$p_{s_1, \dots, s_p}(x_{s_1}, \dots, x_{s_p})$$

$$p_{s_i}(x_{s_i})$$

Definición subespacio incorrelado

Decimos que un subespacio S es un subespacio incorrelado si y sólo si:

$$p_{s_1, \dots, s_p}(x_{s_1}, \dots, x_{s_p}) = \prod_{i=1}^p p_{s_i}(x_{s_i})$$

Definición subespacio incorrelado

Decimos que un subespacio S es un subespacio incorrelado si y sólo si:

$$p_{s_1, \dots, s_p}(x_{s_1}, \dots, x_{s_p}) = \prod_{i=1}^p p_{s_i}(x_{s_i})$$

Definición anomalía no trivial

Decimos que un objeto x_S es una anomalía no trivial respecto del subespacio S si:

$$p_{s_1, \dots, s_p}(x_{s_1}, \dots, x_{s_p}) \ll p_{esp}(x_{s_1}, \dots, x_{s_p})$$

Definición subespacio incorrelado

Decimos que un subespacio S es un subespacio incorrelado si y sólo si:

$$p_{s_1, \dots, s_p}(x_{s_1}, \dots, x_{s_p}) = \prod_{i=1}^p p_{s_i}(x_{s_i})$$

Definición anomalía no trivial

Decimos que un objeto x_S es una anomalía no trivial respecto del subespacio S si:

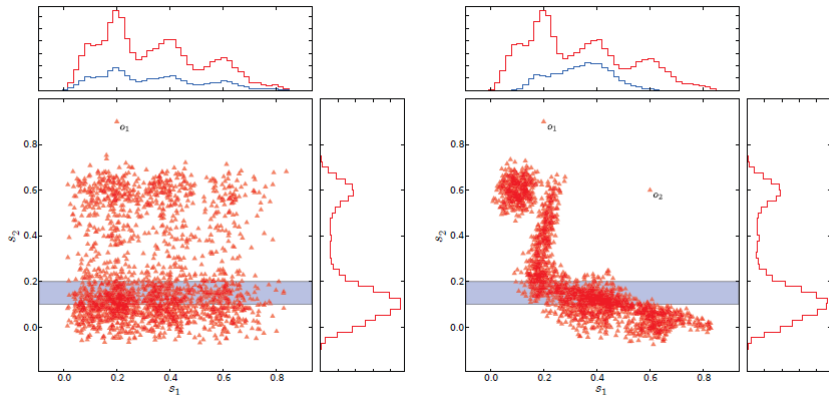
$$p_{s_1, \dots, s_p}(x_{s_1}, \dots, x_{s_p}) \ll p_{esp}(x_{s_1}, \dots, x_{s_p})$$

Relación entre conceptos de anomalía

Este concepto de anomalía es complementario.

Ejemplo de anomalía

Ejemplo de anomalía



Modelos implementados

Algoritmos de ensamblaje

- Algoritmos secuenciales

Algoritmos de ensamblaje

- Algoritmos secuenciales
- Algoritmos independientes

Algoritmos de ensamblaje

- Algoritmos secuenciales
- Algoritmos independientes

Ensamblaje secuencial:

Entrada: Conjunto de datos \mathcal{D} , Algoritmos base $\mathcal{A}_1, \dots, \mathcal{A}_r$

j=1

repetir

Tomamos el algoritmo \mathcal{A}_j según los resultados anteriores

Tomamos el conjunto de datos modificado $f_j(\mathcal{D})$ de anteriores ejecuciones

Ejecutamos el algoritmo \mathcal{A}_j sobre $f_j(\mathcal{D})$

j=j+1

hasta que *fin*;

Resultado: Combinación de los resultados

Algoritmos de ensamblaje

- Algoritmos secuenciales
- Algoritmos independientes

Ensamblaje independiente:

Entrada: Conjunto de datos \mathcal{D} , Algoritmos base $\mathcal{A}_1, \dots, \mathcal{A}_r$

j=1

repetir

 Tomamos el algoritmo \mathcal{A}_j

 Creamos el conjunto de datos modificado $f_j(\mathcal{D})$

 Ejecutamos el algoritmo \mathcal{A}_j sobre $f_j(\mathcal{D})$

 j=j+1

hasta que *fin*;

Resultado: Combinación de los resultados

HICS

Entrada: D: dataset

scores = []

sub = subespacios de alto contraste

para cada $S \in sub$ **hacer**

 Ajustamos un modelo con el algoritmo LOF con la proyección
 sobre S

scores = *scores* + *puntaje LOF*

fin

$scores = \frac{scores}{|sub|}$

Salida: scores: puntajes

CalcularContraste

Entrada: subespacio: subespacio, M : número de iteraciones del subsampling, α : valor para obtener el tamaño de la muestra, D : conjunto de datos

$$size = n \cdot \sqrt[|subespacio|]{\alpha}$$

$$dev = 0$$

para cada $i \in [1, M]$ **hacer**

$comp_atr = \text{aleatorio de subespacio}$

$sel_obj = \text{muestra aleatoria de } D \text{ de tamaño } size$

$dev = dev + \text{CalcularDev}(comp_atr, sel_obj, subespacio, D)$

fin

$$dev = \frac{dev}{M}$$

Salida: dev : contraste

CalcularDev

Entrada: *comp_atr*: atributo con el que comparar, *sel_obj*: muestra seleccionada aleatoriamente, *subespacio*: subespacio sobre el que calcular la desviación, *D*: conjunto de datos

max = 0

para cada $d \in D$ **hacer**

$cum_1 = \sum_{o \in D} o[comp_atr]$ si $o[comp_atr] < d[comp_atr]$

$cum_2 = \sum_{o \in sel_obj} o[comp_atr]$ si $o[comp_atr] < d[comp_atr]$

$f_a = \frac{cum_1}{|D|}$

$f_b = \frac{cum_2}{|D|}$

$subs = |f_a - f_b|$

si $subs > max$ **entonces**

$max = subs$

fin

fin

Salida: *max*: máxima desviación

OUTRES

Entrada: o : instancia, S : subespacio

para cada $i \in (D \setminus S)$ **hacer**

$S' = S \cup \{i\}$

si S' *es relevante* **entonces**

$$\text{den}(o, S') = \frac{1}{n} \sum_{p \in AN(o, S')} K_e\left(\frac{\text{dist}_{S'}(o, p)}{\epsilon(|S'|)}\right)$$

$$\text{dev}(o, S') = \frac{\mu - \text{den}(o, S')}{2\sigma}$$

si $\text{dev}(o, S') \geq 1$ **entonces**

$$\quad \quad r(o) = r(o) \cdot \frac{\text{den}(o, S')}{\text{dev}(o, S')}$$

fin

$OUTRES(o, S')$

fin

en otro caso

 Para recursividad

fin

fin

Salida: r : puntajes

Subespacio relevante

Decimos que un subespacio es relevante si no está distribuido uniformemente.

Subespacio relevante

Decimos que un subespacio es relevante si no está distribuido uniformemente.

Procedimiento

Si la proyección de los datos sobre un subespacio está distribuida según una uniforme entonces sus proyecciones en una dimensión también lo están.

Elementos involucrados

$$AN(o, S) = \{p | dist_S(o, p) \leq \epsilon(|S|)\}$$

Elementos involucrados

$$AN(o, S) = \{p | dist_S(o, p) \leq \epsilon(|S|)\}$$

$$\epsilon(|S|) = 0,5 \cdot \frac{h_{optimal}(|S|)}{h_{optimal}(2)}$$

Elementos involucrados

$$AN(o, S) = \{p | \text{dist}_S(o, p) \leq \epsilon(|S|)\}$$

$$\epsilon(|S|) = 0,5 \cdot \frac{h_{\text{optimal}}(|S|)}{h_{\text{optimal}}(2)}$$

$$h_{\text{optimal}}(d) = \left(\frac{8\Gamma(\frac{d}{2} + 1)}{\pi^{\frac{d}{2}}} \right) (d + 4)(2\sqrt{\pi})^d n^{\frac{-1}{d+4}}$$

Elementos involucrados

$$AN(o, S) = \{p | dist_S(o, p) \leq \epsilon(|S|)\}$$

$$\epsilon(|S|) = 0,5 \cdot \frac{h_{optimal}(|S|)}{h_{optimal}(2)}$$

$$h_{optimal}(d) = \left(\frac{8\Gamma(\frac{d}{2} + 1)}{\pi^{\frac{d}{2}}} \right) (d + 4)(2\sqrt{\pi})^d n^{\frac{-1}{d+4}}$$

$$K_{\epsilon}(x) = (1 - x^2) \quad \forall x < 1$$

Proyecciones Aleatorias

Entrada: d : dimension, D : dataset, k : número de histogramas y proyecciones

$no_neg = \lceil \sqrt{d} \rceil$

$proyecciones = []$

para cada $i \in [1, k]$ **hacer**

$ind = no_neg$ índices aleatorios en $[0, d]$

$proy =$ vector con ceros en todas las posiciones menos en ind ,
 donde hay valores sacados de una normal $\mathcal{N}(0, 1)$

$proyecciones = [proyecciones, proy]$

fin

Salida: $proyecciones$: proyecciones

ObtenerHistogramas

Entrada: D : dataset, $\{w_i\}_{i=1}^k$: vectores de proyecciones, k : numero de histogramas y vectores de proyección

Inicializamos los histogramas $\{h_i\}_{i=1}^k$

```

para  $j = 1 \rightarrow |D|$  hacer
    |
    | para  $i = 1 \rightarrow k$  hacer
    | |  $z_i = x_j^T w_i$ 
    | | Actualiza el histograma  $h_i$  con  $z_i$ 
    | fin
fin
    
```

Salida: $\{h_i\}_{i=1}^k$: histogramas

LODA

Entrada: x : instancia, $\{h_i\}_{i=1}^k$: histogramas, $\{w_i\}_{i=1}^k$: vectores de proyección

para $i = 1 \rightarrow k$ **hacer**

$$z_i = x^T w_i$$

Obtenemos $p_i = p_i(z_i)$ del histograma h_i

fin

$$f = \frac{-1}{k} \sum_{i=1}^k \log(p_i(z_i))$$

Salida: f : puntaje de anomalía de x

Mahalanobis Kernel

Mahalanobis Kernel Entrada: D

$$S = DD^T.$$

$$S = Q\Delta^2Q^T.$$

Almacenamos los vectores propios columna no negativos de $Q\Delta$ en una matriz D'

Normalizamos D' para que tenga media 0 y varianza 1.

vector_media = *media*(D')

puntuaciones = []

para cada fila en D' hacer

score = *distancia*(*vector_media*, *fila*)

puntuaciones = [*puntuaciones*, *score*]

fin

Salida: *puntuaciones*

Componentes

1. Componente basado en distancias: KNN con $k = 5$

Componentes

1. Componente basado en distancias: KNN con $k = 5$
2. Componente basado en dependencia: Mahalanobis Kernel

Componentes

1. Componente basado en distancias: KNN con $k = 5$
2. Componente basado en dependencia: Mahalanobis Kernel
3. Componente basado en densidad en subespacios: IForest

Nota

Aplicamos en todos los componentes la técnica de subsampling.

Trinity Entrada: D

Ejecutamos KNN con $k = 5$ y guardamos el resultado en E_1

Ejecutamos Mahalanobis Kernel y guardamos el resultado en E_2

Ejecutamos IForest y guardamos el resultado en E_3

Estandarizamos E_1 , E_2 y E_3 a media 0 y varianza 1

Hacemos la media para obtener las puntuaciones

Salida: puntuaciones

Recursos usados

- Implementación hecha en Python3

Recursos usados

- Implementación hecha en Python3
- Disponible en GitHub

Recursos usados

- Implementación hecha en Python3
- Disponible en GitHub
- Documentada con Sphinx

Resultados

Conversión del problema a semisupervisado

Conversión del problema a semisupervisado

Problema

Si queremos medir el rendimiento de los modelos necesitamos conjuntos de datos que tengan detectadas las anomalías.

Conversión del problema a semisupervisado

Problema

Si queremos medir el rendimiento de los modelos necesitamos conjuntos de datos que tengan detectadas las anomalías.

Solución

Vamos a utilizar los conjuntos de datos de ODDS de la Universidad de Stony Brooks.

Problema

Conjuntos de datos mantenidos por la Universidad Stony Brooks.
Etiquetados los datos con 0 si son normales y 1 si son anomalías.

Conjuntos de datos ODDS

| <u>Nombre</u> | <u>Dimensionalidad</u> | <u>Número de instancias</u> |
|---------------|------------------------|-----------------------------|
| annthyroid | 6 | 7200 |
| arrhythmia | 274 | 452 |
| breastw | 9 | 683 |
| cardio | 21 | 1831 |
| glass | 9 | 214 |
| ionosphere | 33 | 351 |
| letter | 32 | 1600 |
| lympho | 18 | 148 |
| mammography | 6 | 11183 |
| mnist | 100 | 7603 |
| musk | 166 | 3062 |
| optdigits | 64 | 5216 |
| pendigits | 16 | 6870 |
| pima | 8 | 768 |

Conjuntos de datos ODDS

| <u>Nombre</u> | <u>Dimensionalidad</u> | <u>Número de instancias</u> |
|---------------|------------------------|-----------------------------|
| satellite | 36 | 6435 |
| satimage-2 | 36 | 5803 |
| speech | 400 | 3686 |
| thyroid | 6 | 3772 |
| vertebral | 6 | 240 |
| vowels | 12 | 1456 |
| wbc | 30 | 378 |
| wine | 13 | 129 |

Experimentación

Los 5 modelos se han puesto frente a modelos clásicos implementados en la librería PyOD.

Experimentación

Los 5 modelos se han puesto frente a modelos clásicos implementados en la librería PyOD.

Angle-Based Outlier Detection (ABOD), Connectivity-Based Outlier Factor (COF), Histogram-Based Outlier Score (HBOS), K Nearest Neighbors (KNN), Local Outlier Factor (LOF), Minimum Covariance Determinant (MCD), One-Class Support Vector Machines (OCSVM), Principal Component Analysis (PCA), Subspace Outlier Detection (SOD) y Stochastic Outlier Selection (SOS)

Mejores modelos

Trinity, Mahalanobis Kernel y LODA

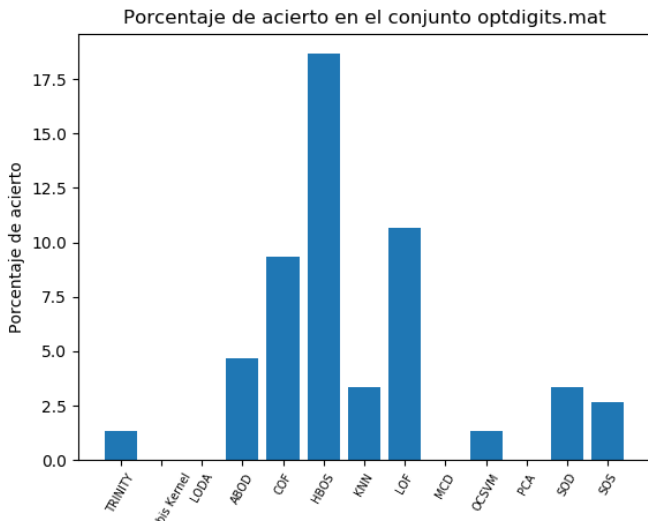
Mejores modelos

Trinity, Mahalanobis Kernel y LODA

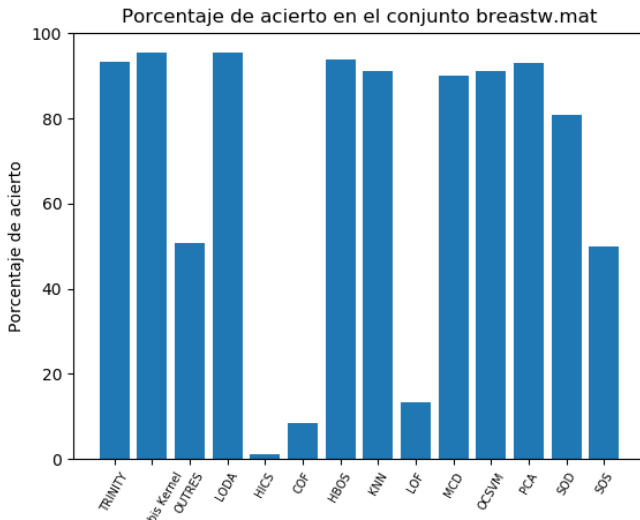
Peores modelos

HICS y OUTRES

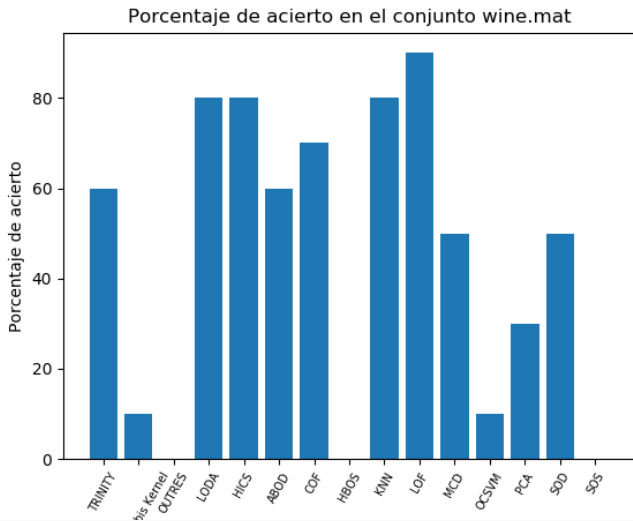
Dificultad del problema



Buenos resultados



Mejor resultado de HICS



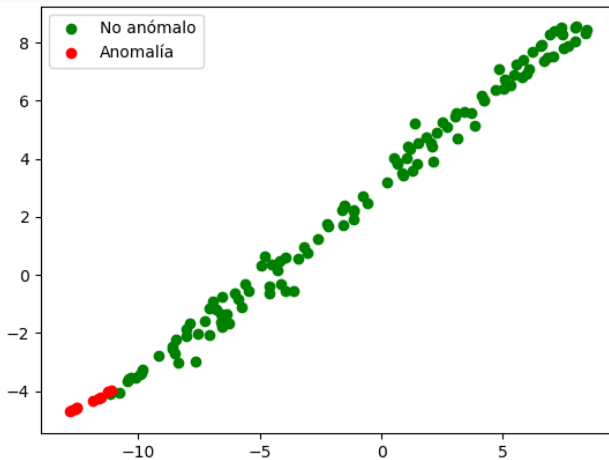
Porcentaje de acierto

| Data / Mod | Trinity | MK | OUTRES | LODA | HICS |
|-------------|-----------|------------------|-----------|------------------|-----------|
| annthyroid | 26.2172 % | 12.3595 % | 7.4906 % | 8.9887 % | 6.5543 % |
| arrhythmia | 48.4848 % | 51.5151 % | NAN | 37.8787 % | NAN |
| breastw | 93.3054 % | 95.3974 % | 50.6276 % | 95.3974 | 1.2552 % |
| cardio | 44.3181 % | 51.7045 % | NAN | 59.0909 % | NAN |
| glass | 11.1111 % | 0 % | 0 % | 0 % | 11.1111 % |
| ionosphere | 73.0158 % | 57.1428 % | NAN | 46.8253 % | NAN |
| letter | 16 % | 15 % | NAN | 4 % | NAN |
| lympho | 83.3333 % | 50 % | NAN | 0 % | NAN |
| mammography | 25.7692 % | 1.1538 % | 5.3846 % | 28.0769 % | 8.0769 % |
| mnist | 40.2857 % | 54.7142 % | NAN | 2 % | NAN |
| musk | 34.0206 % | 0 % | NAN | 32.9896 % | NAN |
| optdigits | 1.3333 % | 0 % | NAN | 0 % | NAN |
| pendigits | 25 % | 16.6666 % | 2.5641 % | 0 % | NAN |
| pima | 50.7462 % | 38.8059 % | 34.7014 % | 54.8507 % | 30.5970 % |

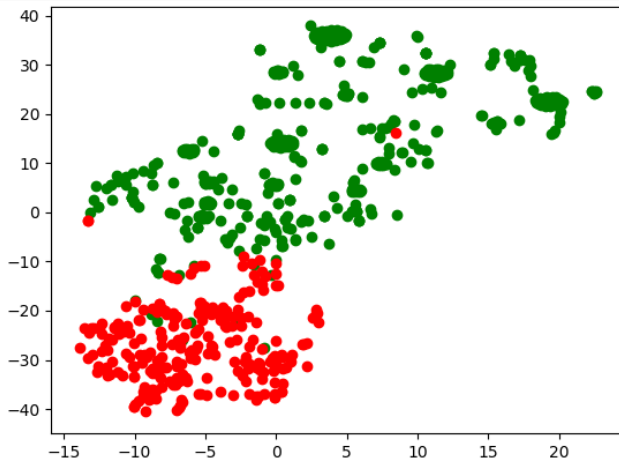
Porcentaje de acierto

| Data / Mod | Trinity | MK | OUTRES | LODA | HICS |
|------------|------------------|------------------|----------|------------------|----------|
| satellite | 54.3713 % | 32.8585 % | NAN | 12.6227 % | NAN |
| satimage-2 | 90.1408 % | 90.1408 % | NAN | 0 % | NAN |
| speech | 0 % | 4.9180 % | NAN | 3.2786 % | NAN |
| thyroid | 39.7849 % | 22.5806 % | 5.3763 % | 1.0752 % | 0 % |
| vertebral | 3.3333 % | 10 % | 3.3333 % | 0 % | 3.3333 % |
| vowels | 24 % | 0 % | 4 % | 8 % | 32 % |
| wbc | 42.8571 % | 0 % | NAN | 71.4285 % | NAN |
| wine | 60 % | 10 % | 0 % | 80 % | 80 % |

Proyecciones

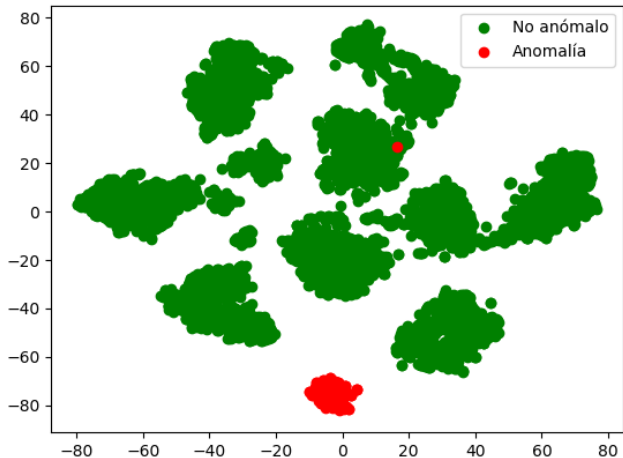


Proyecciones

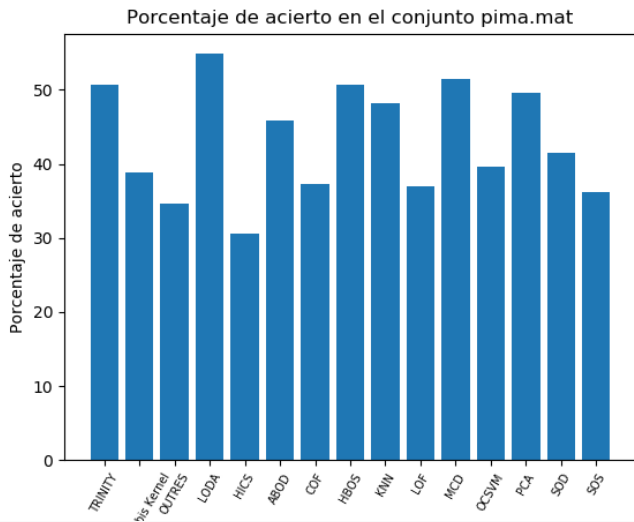


Proyección del conjunto breastw

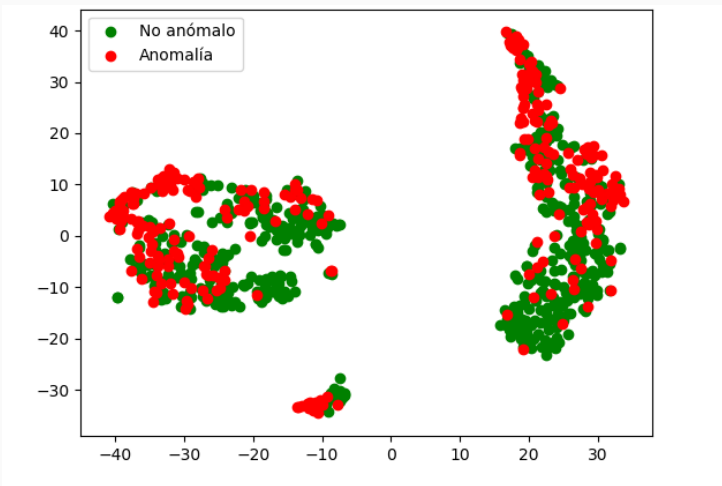
Proyecciones



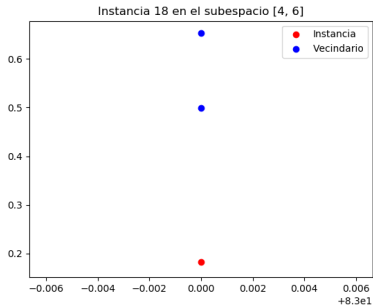
Proyección del conjunto optdigits

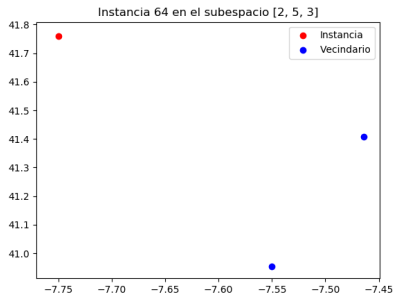
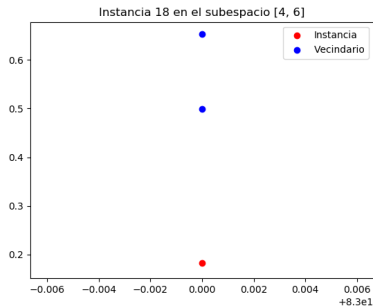


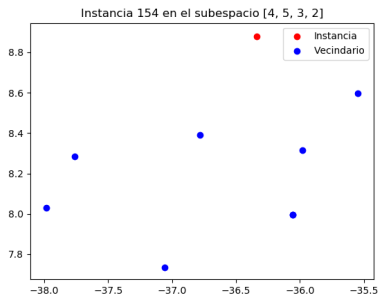
OUTRES



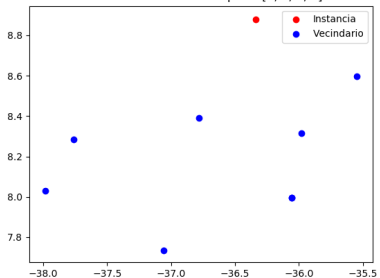
Proyección del conjunto pima



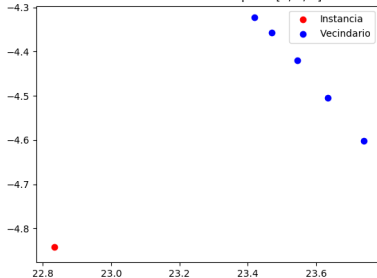




Instancia 154 en el subespacio [4, 5, 3, 2]



Instancia 170 en el subespacio [2, 6, 0]



Valores AUC

$$TPR = \frac{VP}{VP + FN}, \quad FPR = \frac{FP}{VN + FP}$$

Valores AUC

$$TPR = \frac{VP}{VP + FN}, \quad FPR = \frac{FP}{VN + FP}$$

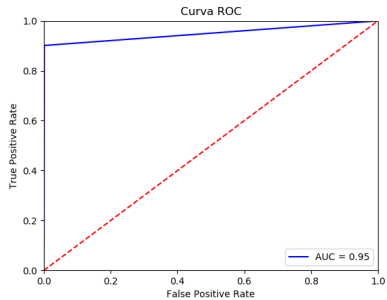
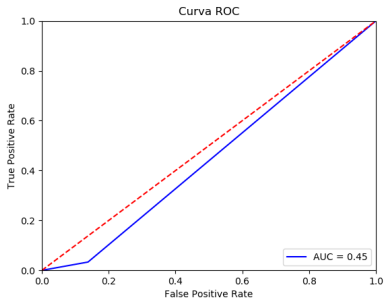
Área bajo la curva

AUC y ROC

| Data / Model | Tinity | MK | OUTRES | LODA | HICS |
|--------------|---------------|---------------|--------|---------------|--------|
| annthyroid | 0.6015 | 0.5266 | 0.5003 | 0.5084 | 0.4953 |
| arrhythmia | 0.6983 | 0.7161 | NAN | 0.6362 | NAN |
| breastw | 0.9485 | 0.9646 | 0.6202 | 0.9646 | 0.2405 |
| cardio | 0.6919 | 0.7328 | NAN | 0.7737 | NAN |
| glass | 0.5360 | 0.4780 | 0.4780 | 0.4780 | 0.5360 |
| ionosphere | 0.7895 | 0.6657 | NAN | 0.5852 | NAN |
| letter | 0.5519 | 0.5466 | NAN | 0.488 | NAN |
| lympho | 0.9131 | 0.7394 | NAN | 0.4788 | NAN |
| mammography | 0.6200 | 0.4940 | 0.5156 | 0.6318 | 0.5294 |
| mnist | 0.6711 | 0.7506 | NAN | 0.4603 | NAN |
| musk | 0.6593 | 0.4836 | NAN | 0.6539 | NAN |
| optdigits | 0.4920 | 0.4851 | NAN | 0.4851 | NAN |
| pendigits | 0.6162 | 0.5736 | 0.5015 | 0.4883 | NAN |
| pima | 0.6217 | 0.5300 | 0.4985 | 0.6532 | 0.4669 |

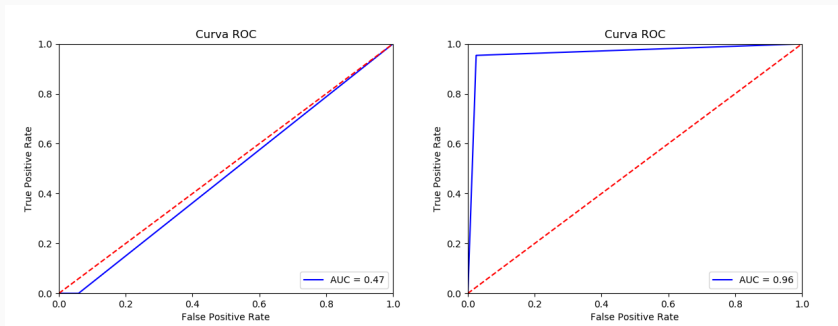
| Data / Model | Tinity | MK | OUTRES | LODA | HICS |
|--------------|---------------|---------------|--------|---------------|--------|
| satellite | 0.6662 | 0.5089 | NAN | 0.3609 | NAN |
| satimage-2 | 0.9500 | 0.9500 | NAN | 0.4938 | NAN |
| speech | 0.4915 | 0.5165 | NAN | 0.5082 | NAN |
| thyroid | 0.6913 | 0.6031 | 0.5149 | 0.4928 | 0.4873 |
| vertebral | 0.4476 | 0.4857 | 0.4476 | 0.4285 | 0.4476 |
| vowels | 0.6064 | 0.4822 | 0.5029 | 0.5236 | 0.6479 |
| wbc | 0.6974 | 0.4705 | NAN | 0.8487 | NAN |
| wine | 0.7831 | 0.5121 | 0.4579 | 0.8915 | 0.8915 |

AUC y ROC



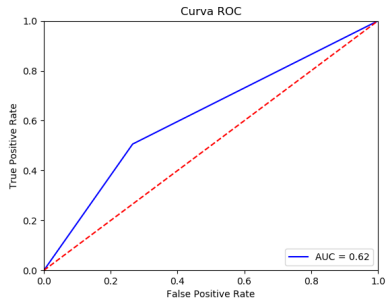
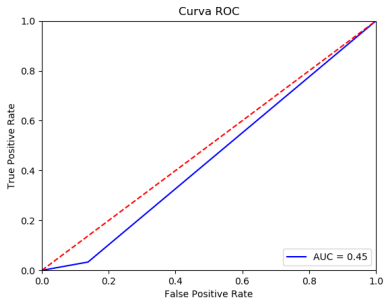
Curvas ROC para Trinity

AUC y ROC



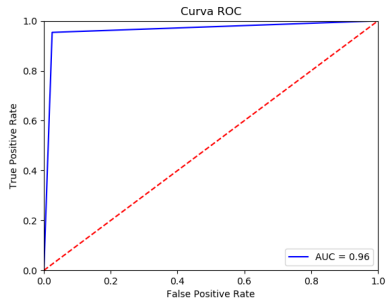
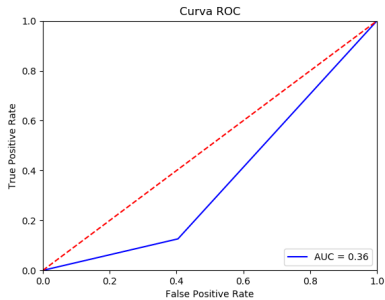
Curvas ROC para Mahalanobis Kernel

AUC y ROC



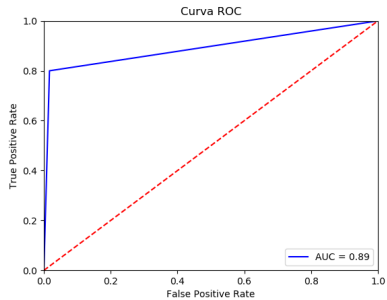
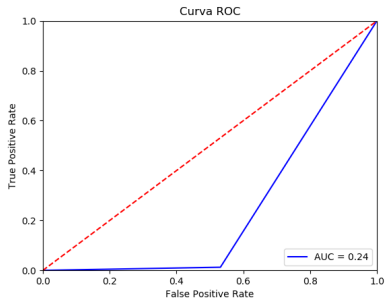
Curvas ROC para OUTRES

AUC y ROC



Curvas ROC para LODA

AUC y ROC



Curvas ROC para HICS

| Data / Mod | Trinity | MK | OUTRES | LODA | HICS |
|--------------------|----------------|-----------|---------------|-------------|-------------|
| annthyroid | 28.5212 | 390.5937 | 193.6110 | 137.8692 | 1169.3422 |
| arrhythmia | 28.8828 | 0.1243 | NAN | 9.0474 | NAN |
| breastw | 23.3296 | 0.3165 | 117.8225 | 13.2879 | 315.4130 |
| cardio | 33.0496 | 6.5851 | NAN | 36.6768 | NAN |
| glass | 12.5272 | 0.0220 | 11.5349 | 4.0798 | 79.2560 |
| ionosphere | 16.5319 | 0.0565 | NAN | 6.9664 | NAN |
| letter | 41.8301 | 5.1648 | NAN | 31.8875 | NAN |
| lympho | 12.2874 | 0.0187 | NAN | 2.8804 | NAN |
| mammography | 31.5619 | 1444.1323 | 27765.1978 | 224.8207 | 2502.0258 |
| mnist | 42.7623 | 464.8962 | NAN | 150.1434 | NAN |
| musk | 42.8683 | 34.4336 | NAN | 60.4207 | NAN |
| optdigits | 36.2723 | 145.6876 | NAN | 103.9115 | NAN |
| pendigits | 34.0999 | 340.2855 | 2757.0345 | 135.0839 | NAN |
| pima | 26.3576 | 0.3482 | 40.6379 | 15.2128 | 150.1283 |

| Data / Mod | Trinity | MK | OUTRES | LODA | HICS |
|-------------------|----------------|-----------|---------------|-------------|-------------|
| satellite | 48.2618 | 301.9960 | NAN | 129.6712 | NAN |
| satimage-2 | 39.7535 | 200.5037 | NAN | 113.2818 | NAN |
| speech | 76.6633 | 62.2470 | NAN | 73.0844 | NAN |
| thyroid | 32.6503 | 65.3469 | 64.2370 | 72.1607 | 376.2697 |
| vertebral | 13.4726 | 0.0301 | 1.7653 | 4.6250 | 9.1404 |
| vowels | 36.2093 | 3.1876 | 426.6157 | 28.2930 | 5126.5627 |
| wbc | 17.0011 | 0.0675 | NAN | 7.5109 | NAN |
| wine | 11.2756 | 0.0165 | 58.9642 | 2.4540 | 774.9212 |

Conclusiones y Trabajo Futuro

Conclusiones y trabajo futuro

- Rendimiento similar a los clásicos aunque los superan en algunos casos: potencial.

Conclusiones y trabajo futuro

- Rendimiento similar a los clásicos aunque los superan en algunos casos: potencial.
- Mal enfoque de los algoritmos HICS y OUTRES. Detección de anomalías no triviales.

Conclusiones y trabajo futuro

- Rendimiento similar a los clásicos aunque los superan en algunos casos: potencial.
- Mal enfoque de los algoritmos HICS y OUTRES. Detección de anomalías no triviales.
- Consumo de tiempo mucho mayor.

Conclusiones y trabajo futuro

- Rendimiento similar a los clásicos aunque los superan en algunos casos: potencial.
- Mal enfoque de los algoritmos HICS y OUTRES. Detección de anomalías no triviales.
- Consumo de tiempo mucho mayor.
- Valores perdidos y tratamiento.

Conclusiones y trabajo futuro

- Rendimiento similar a los clásicos aunque los superan en algunos casos: potencial.
- Mal enfoque de los algoritmos HICS y OUTRES. Detección de anomalías no triviales.
- Consumo de tiempo mucho mayor.
- Valores perdidos y tratamiento.
- Algoritmo nuevo.

Gracias por su atención.

¿Preguntas?