DECSAI
Departamento de Ciencias de la Computación e I.A.
Universidad de Granada

Detección de anomalías

© Fernando Berzal, berzal@acm.org

Detección de anomalías

- Introducción a la detección de anomalías
- Técnicas de detección de anomalías
  - Técnicas estadísticas.
  - Técnicas basadas en Teoría de la Información.
  - Técnicas basadas en proximidad.
  - Técnicas basadas en clustering.
  - Técnicas basadas en clasificación.
  - Técnicas basadas en asociación.
  - Técnicas basadas en reconstrucción.
  - Ensembles de detectores de anomalías
- Evaluación de resultados

Detección de anomalías

Una anomalía es una observación que no se ajusta a la distribución del resto de los datos (i.e. improbable dada la distribución del resto de observaciones).

El objetivo de la detección de anomalías es identificar esos objetos que no se ajustan a los patrones habituales.

a.k.a.

- **Outlier analysis** (análisis de valores atípicos)
- **Deviation detection** (detección de desviaciones)
- **Exception mining** (minería de excepciones)

Detección de anomalías

- Las anomalías son ejemplos (puntos de datos) considerablemente diferentes al resto de los datos.
- Por definición, son relativamente poco frecuentes: caso extremo de clasificación con clases muy poco balanceadas (pocas anomalías en muchos datos).
- A menudo, sólo se encuentran cuando tenemos muchos datos.
- El contexto puede resultar extremadamente importante (p.ej. heladas en agosto).

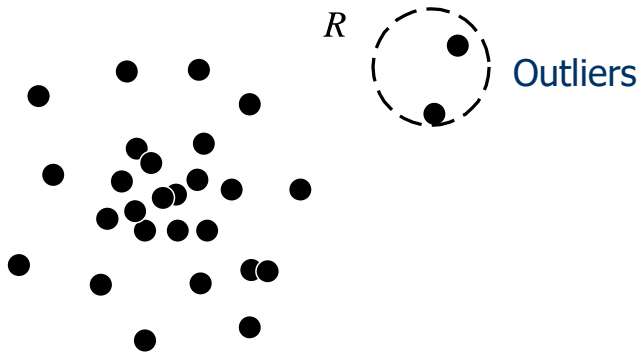
Detección de anomalías

Aplicaciones

- **Seguridad**: Patrones de actividad anormales en sistemas de detección de intrusiones [IDS].
- **Medicina**: Excepciones útiles en el diagnóstico, la evaluación de medicamentos o la detección de contraindicaciones.
- **Finanzas**: Detección de fraude en el uso de tarjetas de crédito o seguros, casos de "insider trading"…
- **Industria**: Monitorización de los dispositivos de una aeronave (seguridad aérea), control de calidad en fabricación, monitorización de ordenadores en un centro de datos…

Detección de anomalías

Causas de las anomalías



- Datos de diferentes clases
- Variación natural
- Errores en los datos (i.e. ruido)
- Aparición de nuevos fenómenos (p.ej. noticias, trending topics…), que inicialmente son anómalos y posteriormente se incorporan al modelo de comportamiento normal [novelty detection].

Detección de anomalías

Anomalías
vs. Ruido

- El ruido no produce necesariamente valores inusuales.
- El ruido introduce un error aleatorio o varianza en los datos, por lo que, idealmente, debería ser eliminado antes de detectar las verdaderas anomalías.
- El ruido no resulta interesante, las anomalías sí: los outliers no se generan por el mismo mecanismo que genera el resto de los datos.

Detección de anomalías

Detección de anomalías
vs. Aprendizaje supervisado (clasificación)

- En detección de anomalías tenemos un número muy pequeño (p.ej. 0-20) de ejemplos positivos y un número enorme de ejemplos negativos (en clasificación solemos tener muchos de ambos).
- En clasificación, tenemos suficientes ejemplos positivos como para caracterizarlos y saber qué aspecto pueden tener futuros ejemplos (similares a los ya disponibles), p.ej. filtro anti-spam.
- En detección de anomalías, podemos tener distintos "tipos" de anomalías, difíciles de caracterizar, y futuras anomalías puede que no se parezcan en nada a las ya conocidas.

Detección de anomalías

Detección de anomalías
vs. Aprendizaje no supervisado (clustering)

- Diferente propósito: Mientras que en clustering buscamos patrones mayoritarios en los datos para agruparlos de acuerdo a ellos, en detección de anomalías intentamos identificar aquellos casos excepcionales que se desvían sustancialmente de los patrones mayoritarios.

- Diferente metodología: Mientras que el clustering es no supervisado, la detección de anomalías puede utilizar distintos grados de supervisión.

Técnicas de detección de anomalías

- Se basan en asumir que la mayoría de los datos son normales.

- Algunas técnicas proporcionan una clasificación binaria (etiqueta sí/no), si bien suele ser recomendable poder medir el grado con el que un objeto es anómalo.

- El grado de anomalía [anomaly score] nos permite realizar un ranking (y, en ocasiones, puede tener un significado estadístico).

Técnicas de detección de anomalías

Estrategias de detección de anomalías

- Los **enfoques basados en modelos** [model-based] construyen un modelo (paramétrico o no paramétrico) y las anomalías son los datos que no se ajustan al modelo (o lo distorsionan).

- Los **enfoques sin modelo** [model-free] detectan directamente las anomalías sin construir un modelo explícito.

Técnicas de detección de anomalías

- Técnicas estadísticas.
- Técnicas basadas en Teoría de la Información.
- Técnicas basadas en proximidad (distancia & densidad).
- Técnicas basadas en clustering.
- Técnicas basadas en clasificación.
- Técnicas basadas en asociación.
- Técnicas basadas en reconstrucción.
- Ensembles

Detección de anomalías
Técnicas estadísticas

Definición probabilística de anomalía

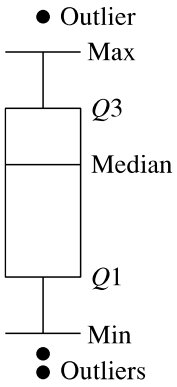
Objeto que tiene una probabilidad baja de ocurrencia con respecto a una distribución de probabilidad que modela los datos.

- Se asume un modelo paramétrico que describe la distribución de los datos (p.ej. normal).
- Se aplica un test estadístico que depende de los parámetros de la distribución de los datos (p.ej. media y varianza) y del número esperado de anomalías (límite de confianza).

Detección de anomalías
Técnicas estadísticas

Rango intercuartil

IQR = Q3 - Q1



Outliers en diagramas de cajas:

Un valor se considera anómalo si…

- … es 1.5*IQR menor que Q1
- … es 1.5*IQR mayor que Q3

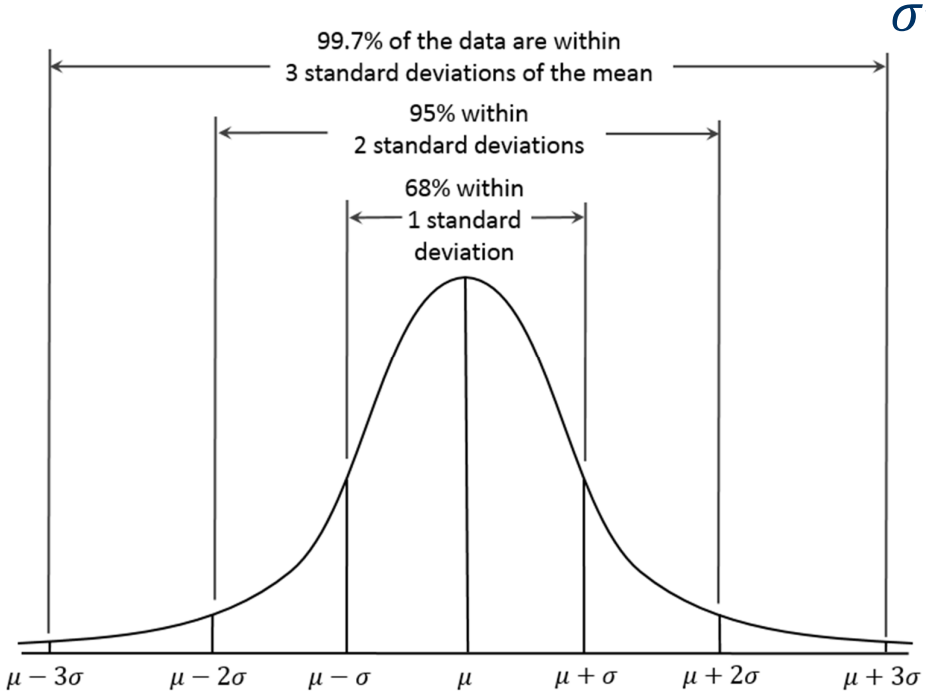
¿Por qué?

En una distribución normal,
el intervalo [Q1-1.5*IQR, Q3+1.5*IQR]
contiene el 99.3% de los casos.

Detección de anomalías
Técnicas estadísticas

La distribución normal

$$N(\mu, \sigma) \qquad f(\vec{x}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



Detección de anomalías
Técnicas estadísticas

La distribución normal

Test de Grubbs (z-score)

Para detectar anomalías en una variable, asumimos que sigue una distribución normal $N(\mu, \sigma)$:

- Detectamos un outlier en cada iteración, lo eliminamos y volvemos a repetir el test.
- Estadístico del test de Grubbs: $G = \frac{\max |x-\mu|}{\sigma}$
- Rechazamos la hipótesis nula (H_0: no hay outliers) si

$$G > \frac{n-1}{\sqrt{n}} \sqrt{\frac{t^2_{\alpha/2n,n-2}}{n-2+t^2_{\alpha/2n,n-2}}}$$

Detección de anomalías
Técnicas estadísticas

La distribución normal

Algoritmo de detección de anomalías

- Seleccionamos un conjunto de características que creamos indicativo de posibles anomalías.
- Ajustamos una distribución de probabilidad normal $N(\mu_j, \sigma_j)$ para cada una de las variables seleccionadas.
- Asumiendo que las características son independientes, calculamos la probabilidad de un ejemplo como

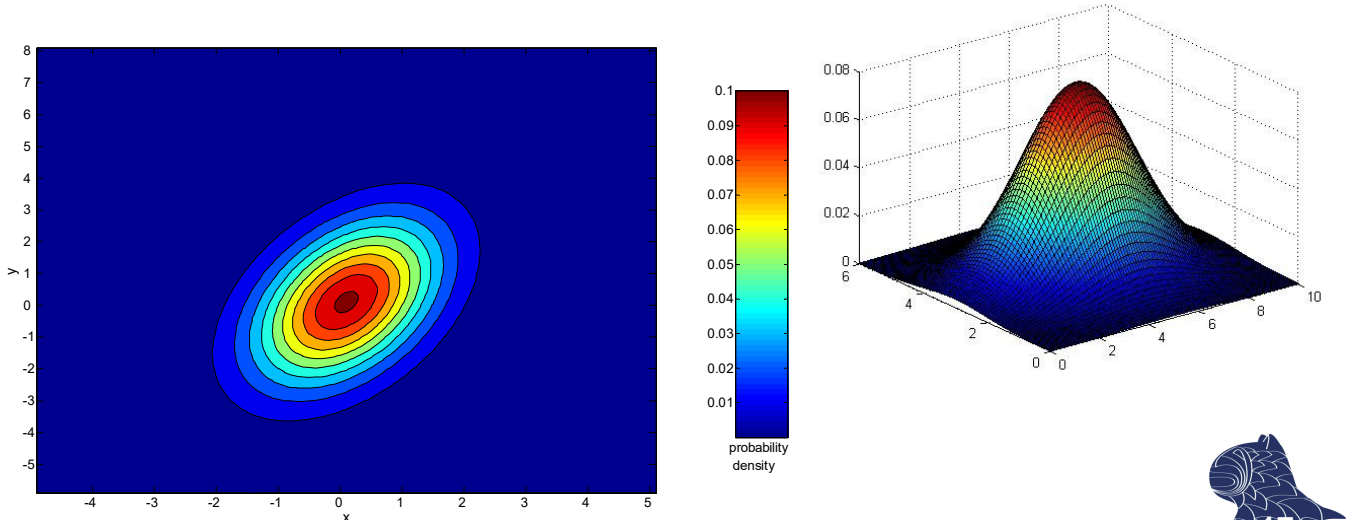
$$p(\vec{x}) = \prod p(x_j; \mu_j, \sigma_j^2) = \prod \frac{1}{\sigma_j \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_j - \mu_j}{\sigma_j}\right)^2}$$

- Decimos que es una anomalía si $p(\vec{x}) < \varepsilon$

Detección de anomalías
Técnicas estadísticas

La distribución normal multivariante

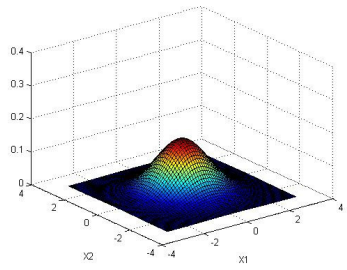
$$N(\vec{\mu}, \Sigma) \qquad f(\vec{x}) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu})}$$

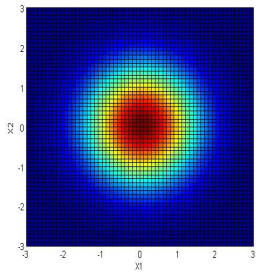


Detección de anomalías
Técnicas estadísticas

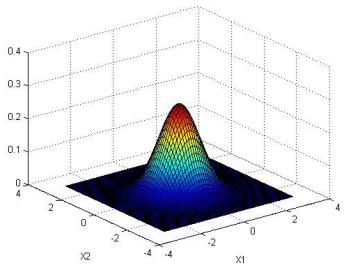
La distribución normal multivariante

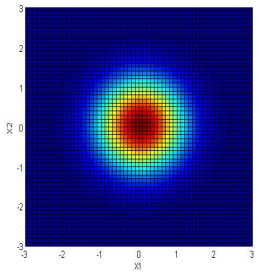
$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



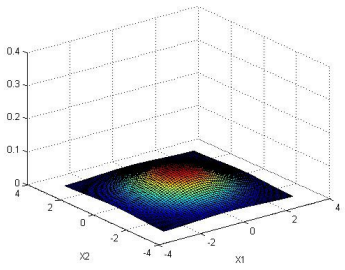


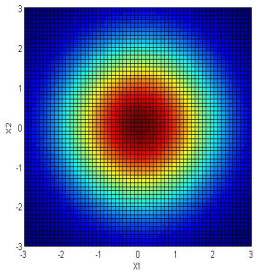
$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Sigma = \begin{bmatrix} 0.6 & 0 \\ 0 & 0.6 \end{bmatrix}$$





$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

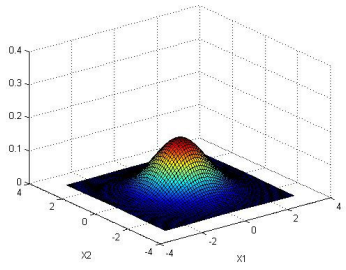


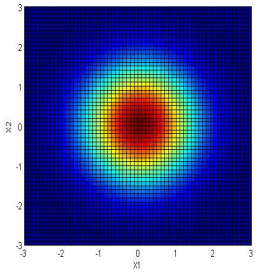


Detección de anomalías
Técnicas estadísticas

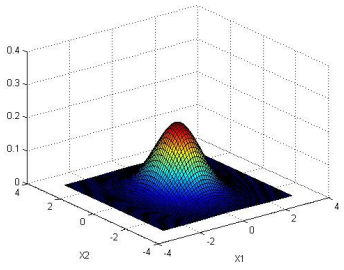
La distribución normal multivariante

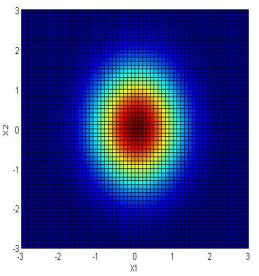
$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



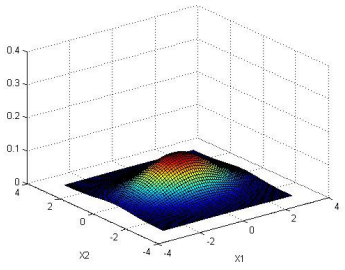


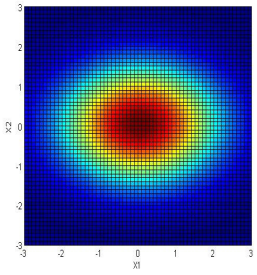
$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Sigma = \begin{bmatrix} 0.6 & 0 \\ 0 & 1 \end{bmatrix}$$





$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

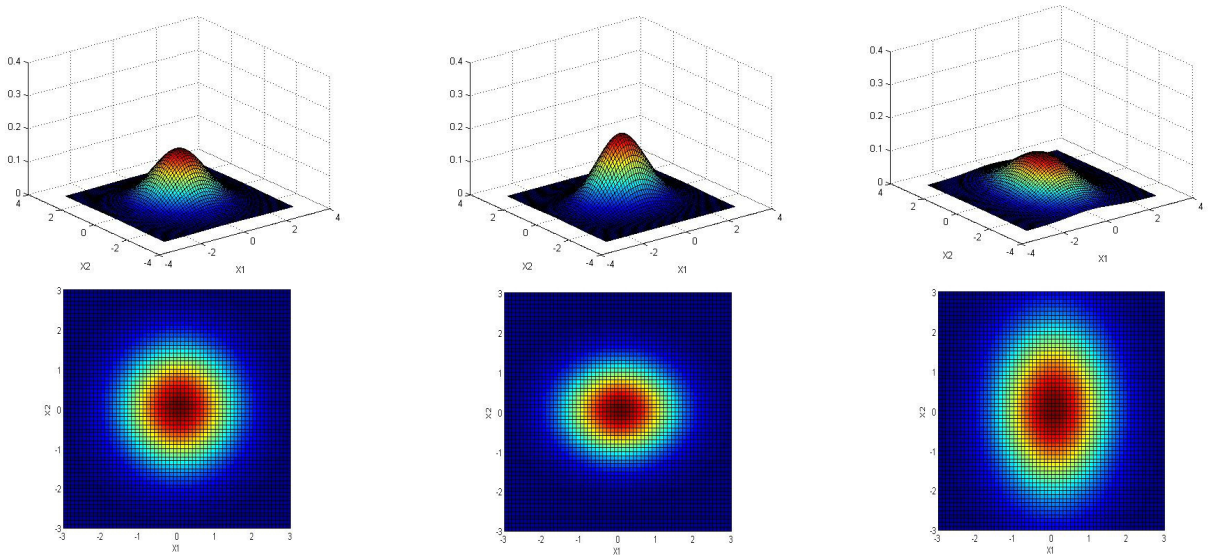




Detección de anomalías
Técnicas estadísticas

La distribución normal multivariante

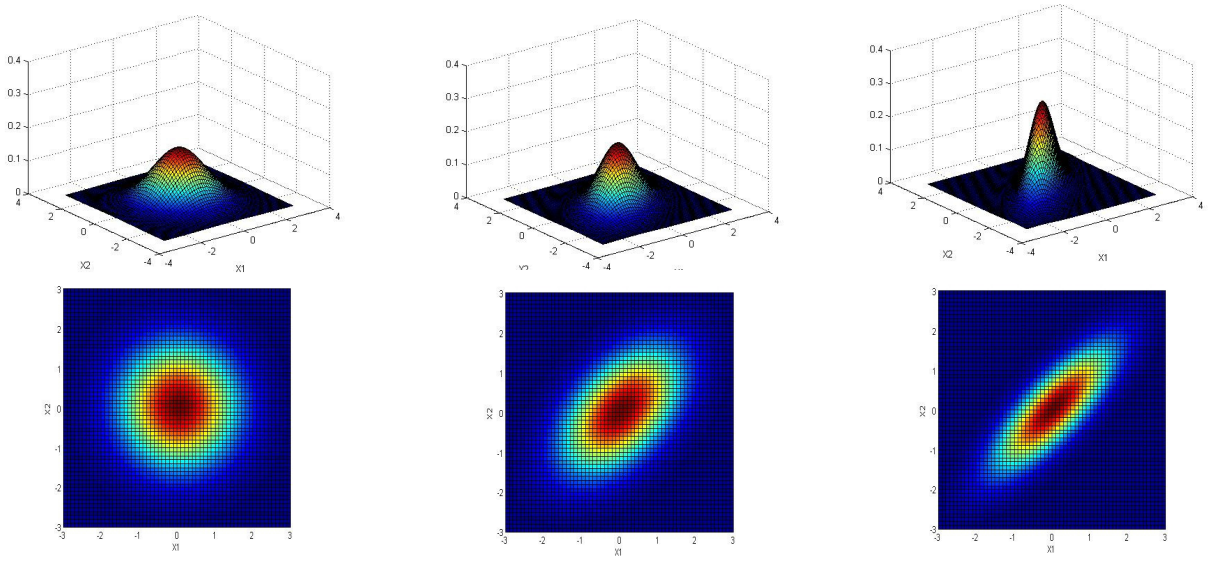
$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ $\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 0.6 \end{bmatrix}$ $\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$



Detección de anomalías
Técnicas estadísticas

La distribución normal multivariante

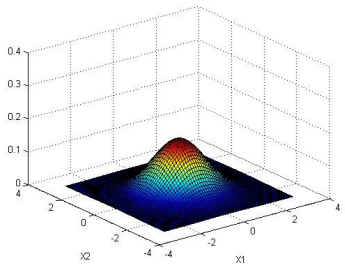
$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ $\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$ $\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Sigma = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$

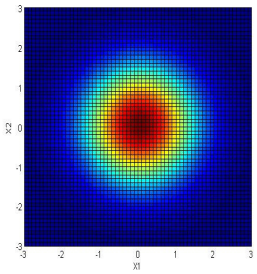


Detección de anomalías
Técnicas estadísticas

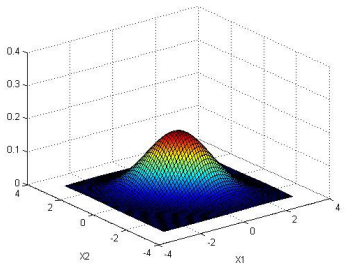
La distribución normal multivariante

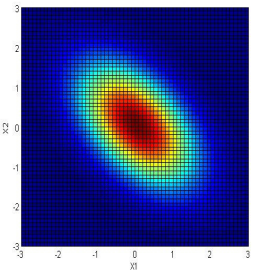
$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



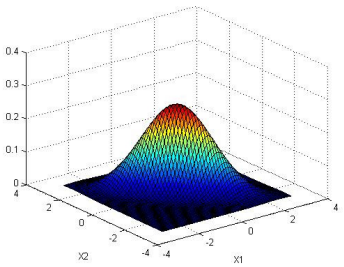


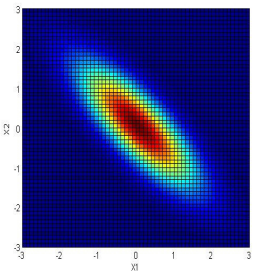
$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Sigma = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$$





$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Sigma = \begin{bmatrix} 1 & -0.8 \\ -0.8 & 1 \end{bmatrix}$$

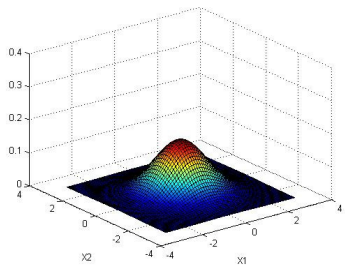


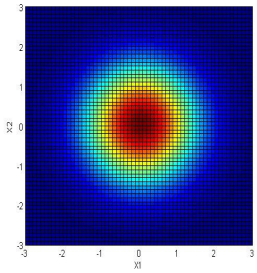


Detección de anomalías
Técnicas estadísticas

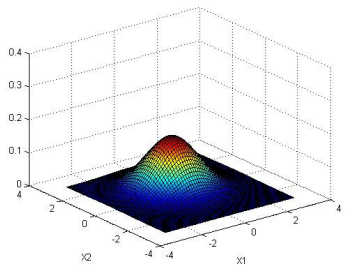
La distribución normal multivariante

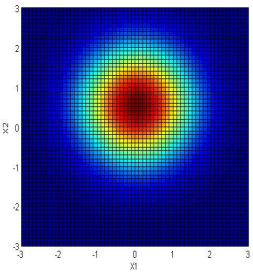
$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



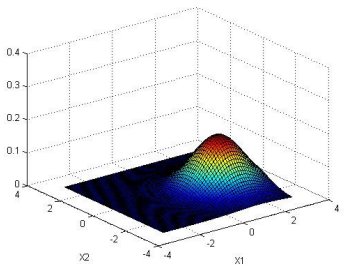


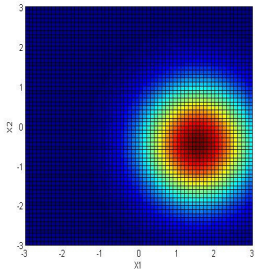
$$\mu = \begin{bmatrix} 0 \\ 0.5 \end{bmatrix} \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$





$$\mu = \begin{bmatrix} 1.5 \\ -0.5 \end{bmatrix} \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$





Detección de anomalías
Técnicas estadísticas

La distribución normal multivariante

Algoritmo de detección de anomalías

- Ajustamos los parámetros de la distribución $N(\vec{\mu}, \Sigma)$ a partir de los datos disponibles (vector de medias y matriz de covarianzas).

- Dado un ejemplo x, calculamos su probabilidad:

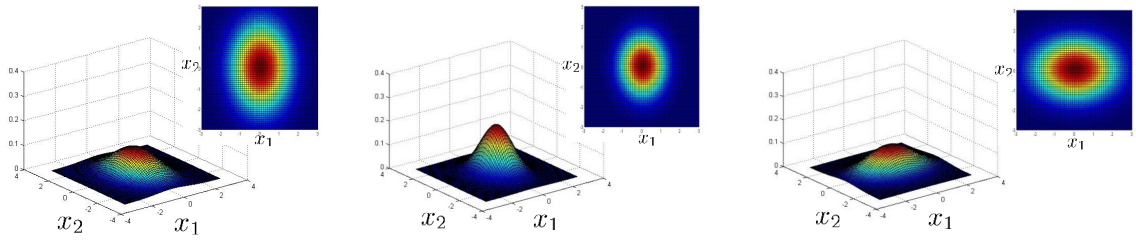
$$p(\vec{x}) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu})}$$

- Decimos que es una anomalía si $p(\vec{x}) < \varepsilon$

Detección de anomalías
Técnicas estadísticas

La distribución normal multivariante

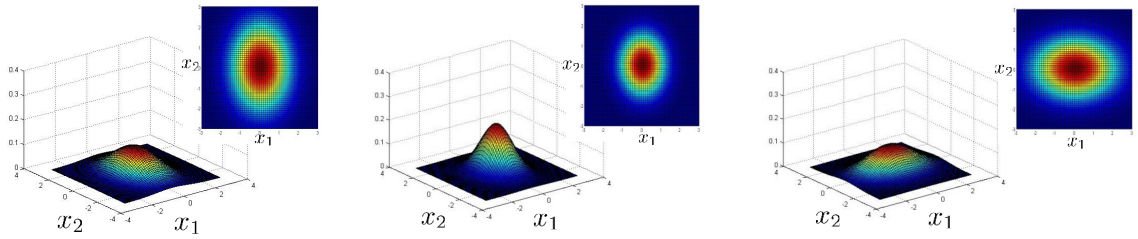
El modelo basado en distribuciones normales independientes corresponde a una distribución multivariante con una matriz de covarianza es diagonal (todos los elementos de la matriz de covarianza son 0 salvo los de su diagonal, que representan la varianza de cada variable por separado).



La distribución normal multivariante

El test de Grubbs puede adaptarse para el caso multivariante si utilizamos la distancia de Mahalanobis:

$$d^2_{Mahalanobis}(\vec{x},\vec{\mu}) = (\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu})$$



Detección de anomalías
Técnicas estadísticas

Test χ^2

Detección de anomalías multivariante

$$\chi^2 = \sum_{i=1}^{n} \frac{(o_i - E_i)^2}{E_i}$$

- o_i: Valor observado en la i-ésima dimensión.
- E_i: Valor esperado en la i-ésima dimensión (media).

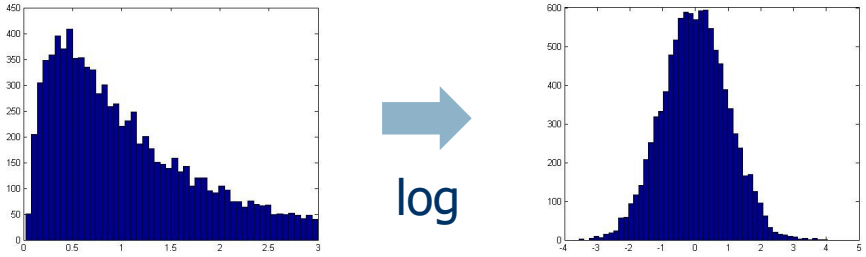
Los valores que más contribuyen al valor de χ^2 son aquéllos cuyo valor difiere más del esperado:

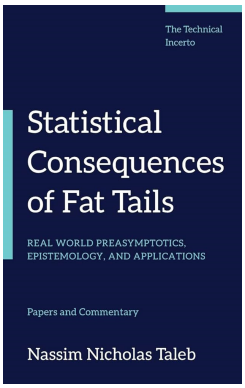
Cuanto mayor sea el valor de χ^2, más anómala es la observación.

Detección de anomalías
Técnicas estadísticas

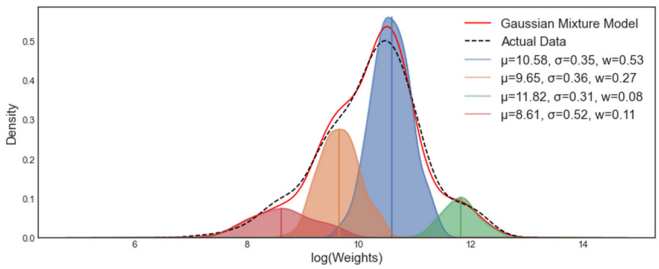
Problemas

- Identificación correcta de la distribución de probabilidad (p.ej. "fat tails")





- ¿Son los datos una mezcla [mixture] de múltiples distribuciones de probabilidad?



Detección de anomalías
Técnicas estadísticas

Mezcla de distribuciones de probabilidad

Asumimos que el conjunto de datos D contiene muestras de una mezcla de dos distribuciones de probabilidad:

- M (distribución mayoritaria)
- A (distribución anómala)

Estrategia

- Inicialmente, asumimos que todas las muestras son M.
- Para cada muestra x de M, probamos a moverla a A: Calculamos la diferencia de verosimilitud Δ entre la distribución actual y la distribución con el ejemplo cambiado. Si Δ es superior a un umbral, x se considera anómalo y el cambio se mantiene.

Detección de anomalías
Técnicas estadísticas

Mezcla de distribuciones de probabilidad

Distribución de los datos D = $(1 - \lambda)$ M + λ A

- M se estima a partir de los datos.
- A se asume inicialmente que es una distribución uniforme.

Verosimilitud [likelihood]

$$\mathrm{L(D)} = \prod P_D(x_i) = \left((1-\lambda)^{|M|} \prod_{x_i \in M} P_M(x_i)\right)\left(\lambda^{|A|} \prod_{x_i \in A} P_A(x_i)\right)$$

Log-likelihood

$$\mathrm{LL(D)} = |M| \log(1-\lambda) + \sum_{x_i \in M} \log P_M(x_i) + |A| \log \lambda + \sum_{x_i \in A} \log P_A(x_i)$$

Detección de anomalías
Técnicas estadísticas

KDE [Kernel Density Estimation]

Estimación de la densidad del kernel
= Estimación del kernel
= Método de Parzen-Rosenblatt

- Método no paramétrico para estimar la densidad de probabilidad de una variable aleatoria.

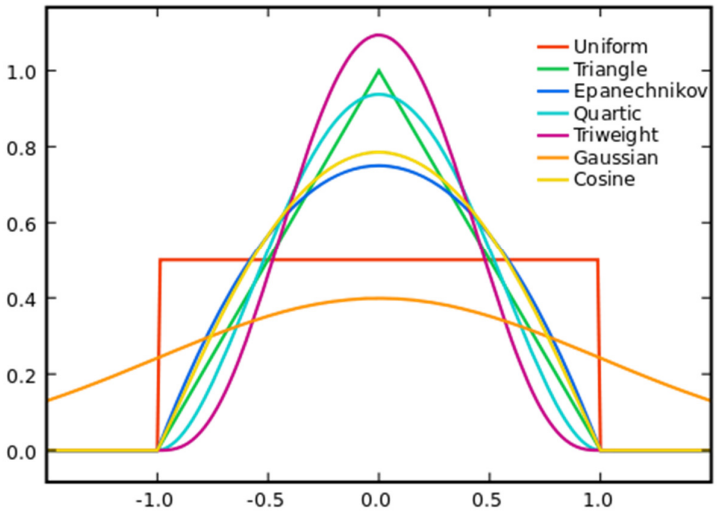
- Un kernel es una función real, integrable y no negativa que satisface dos condiciones:

$$\int_{-\infty}^{\infty} K(u)du = 1$$

$$K(-u) = K(u)$$

KDE [Kernel Density Estimation]

Ejemplos de kernels



https://en.wikipedia.org/wiki/Kernel_%28statistics%29

Detección de anomalías
Técnicas estadísticas

KDE [Kernel Density Estimation]

Kernel gaussiano N(0,1):

$$K(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$$

La estimación del kernel de una función de densidad de probabilidad es

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

donde K() es un kernel y h es su ancho de banda (hiperparámetro de suavizado).

Detección de anomalías
Técnicas estadísticas

Fortalezas y debilidades

- Fundamento matemático firme
  (resultados estadísticamente justificables)
- Buenos resultados si se conoce la distribución de los datos (puede que no la conozcamos).
- Para datos de dimensionalidad elevada, puede ser difícil (o imposible) estimar la distribución.
- Las anomalías pueden distorsionar los parámetros estimados de la distribución.

Detección de anomalías
Teoría de la Información

Idea

Medir cuánta información
se pierde al eliminar una observación

$$Gain(x) = Info(D) - Info(D \setminus x)$$

- Las anomalías
  deberían mostrar una mayor ganancia.
- Los puntos normales
  deberían proporcionar una menor ganancia.

Detección de anomalías
Proximidad

Idea

Objetos separados de los demás
pueden considerar anómalos [outliers]

Dos tipos de métodos de detección de anomalías basados en proximidad:

- Métodos basados en distancia: Un objeto es un outlier si su vecindario no incluye suficientes vecinos.
- Métodos basados en densidad: Un objeto es un outlier si su densidad es mucho menor que la de sus vecinos.

Detección de anomalías
Proximidad: Distancia

Detección de anomalías basada en distancias

Hiperparámetros

- Umbral de distancia r≥0
- Fracción $0 \leq \pi \leq 1$

Algoritmo

Un objeto o es un outlier DB(r,π) si

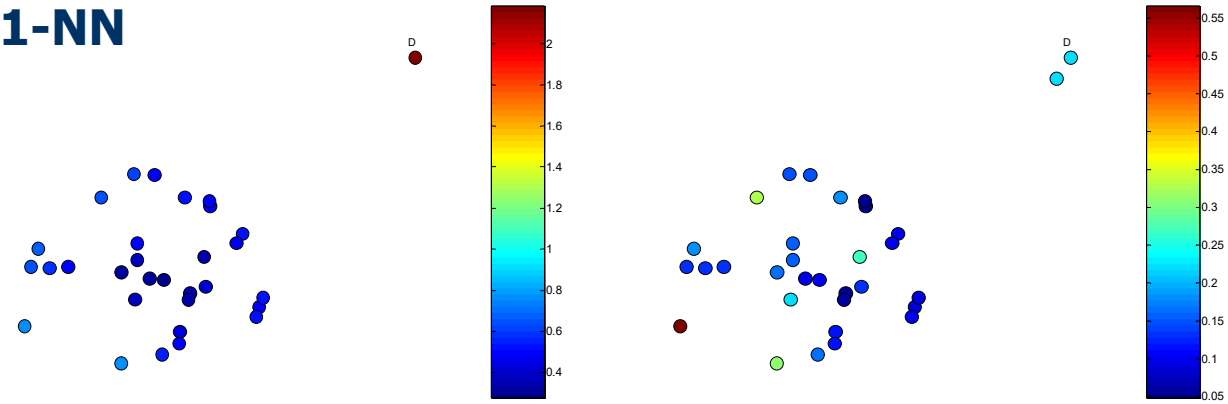
$$\frac{|\{o' | dist(o, o') \leq r\}|}{|D|} \leq \pi$$

Detección de anomalías
Proximidad: Distancia

Detección de anomalías basada en distancias

Otra posibilidad: Distancia al k-ésimo vecino más cercano

1-NN



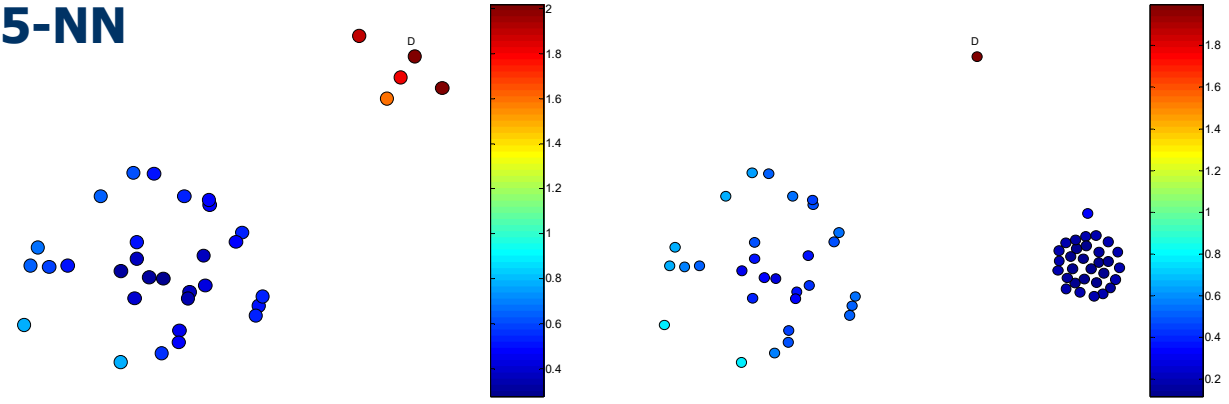
No funciona con dos outliers

Detección de anomalías
Proximidad: Distancia

Detección de anomalías basada en distancias

Otra posibilidad: Distancia al k-ésimo vecino más cercano

5-NN



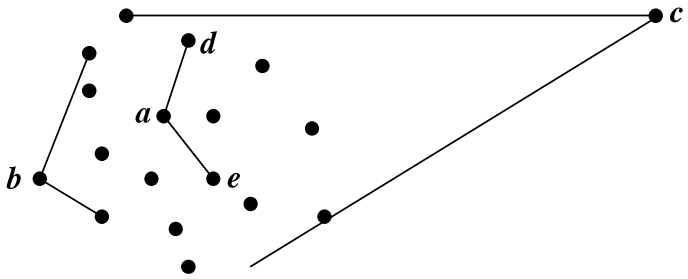
Pequeño cluster

Clusters de distinta densidad

Detección de anomalías
Proximidad: Distancia

ABOD [Angle-Based Outlier Detection]

En espacios de
elevada dimensionalidad…



ABOF
[Angle-Based Outlier Factor]

$$ABOF(o) = var_{x,y \in D \setminus \{o\}} \left(\frac{\langle \overrightarrow{ox}, \overrightarrow{oy} \rangle}{d(o,x)^2 d(o,y)^2} \right)$$

Cuanto más lejos esté un punto de otros, menor será la varianza de sus ángulos y menor será su ABOF.

En este caso,
las anomalías se ordenan de menor a mayor ABOF.

Detección de anomalías
Proximidad: Distancia

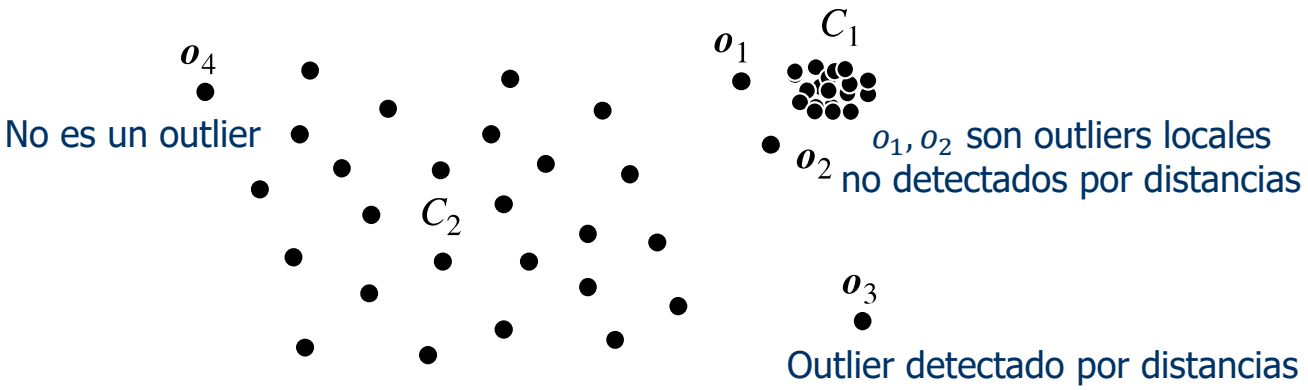
Fortalezas y debilidades

- Sencillez.
- Ineficiencia: $O(n^2)$.
- Sensible a los parámetros escogidos.
- Sensible a variaciones en densidad.
- Muchas métricas de distancia son problemáticas en espacios de alta dimensionalidad.

Detección de anomalías
Proximidad: Densidad

Detección de anomalías basada en densidades

Para detectar outliers locales:



Problema

Cómo medir la densidad relativa de un objeto

Detección de anomalías
Proximidad: Densidad

LOF [Local Outlier Factor]

k-distancia de un objeto: $d_k(o)$

Distancia entre o y su k-ésimo vecino más cercano.

Distancia $d(o,p)$, entre o y otro objeto $p \in D$, tal que

- Al menos k objetos $o' \in D\backslash\{o\}$, $d(o,o') \leq d(o,p)$
- Como mucho k−1 objetos $o'' \in D\backslash\{o\}$, $d(o,o'') < d(o,p)$

El vecindario $N_k(o)$ contiene los vecinos de o a distancia no mayor que $d_k(o)$: $N_k(o) = \{o' | o' \in D, d(o,o') \leq d_k(o)\}$

NOTA: $N_k(o)$ puede contener más de k objetos.

Detección de anomalías
Proximidad: Densidad

LOF [Local Outlier Factor]

Distancia de alcance [reachability distance]: $rd_k(o \leftarrow o')$

$$rd_k(o \leftarrow o') = \max\{d_k(o), d(o, o')\}$$

NOTA: La distancia de alcance no es simétrica
$rd_k(o \leftarrow o') \neq rd_k(o' \leftarrow o)$

Densidad de alcance local: $lrd_k(o)$

$$lrd_k(o) = \frac{|N_k(o)|}{\sum_{o' \in N_k(o)} rd_k(o' \leftarrow o)}$$

Detección de anomalías
Proximidad: Densidad

LOF [Local Outlier Factor]

Promedio de la razón entre las densidades locales de los vecinos de un objeto y la densidad local del objeto.

$$LOF_k(o) = \frac{1}{|N_{k(o)}|} \sum_{o' \in N_k(o)} \frac{lrd_k(o')}{lrd_k(o)}$$

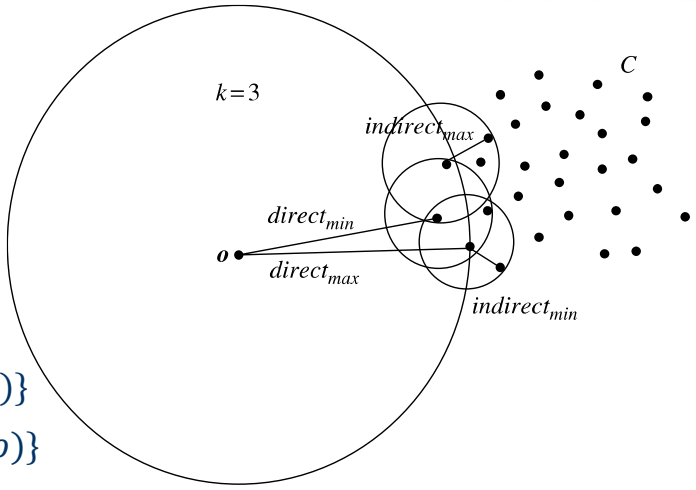
$$= \sum_{o' \in N_k(o)} lrd_k(o') \cdot \sum_{o' \in N_k(o)} rd_k(o' \leftarrow o)$$

- Para un objeto en un clúster consistente, su LOF es 1.
- Para un outlier, su LOF será elevado.

Detección de anomalías
Proximidad: Densidad

LOF [Local Outlier Factor]

Interpretación geométrica



$$direct_{min}(o) = \min\{rd_k(o' \leftarrow o) | o' \in N_k(o)\}$$
$$direct_{max}(o) = \max\{rd_k(o' \leftarrow o) | o' \in N_k(o)\}$$

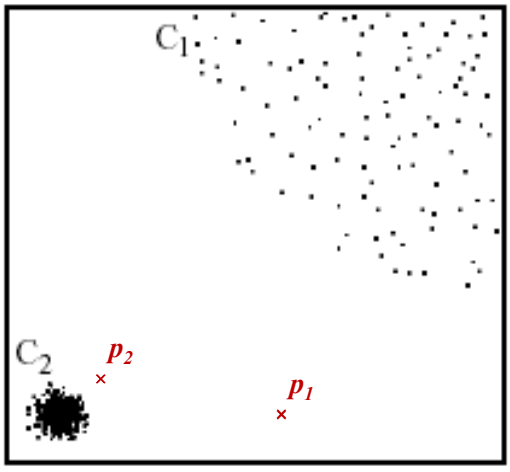
$$indirect_{min}(o) = \min\{rd_k(o'' \leftarrow o') | o' \in N_k(o) \; and \; o'' \in N_k(o')\}$$
$$indirect_{max}(o) = \max\{rd_k(o'' \leftarrow o') | o' \in N_k(o) \; and \; o'' \in N_k(o')\}$$

LOF(o) está acotado:

$$\frac{direct_{min}(o)}{indirect_{max}(o)} \leq LOF(o) \leq \frac{direct_{max}(o)}{indirect_{min}(o)}$$

Detección de anomalías
Proximidad: Densidad

Fortalezas y debilidades



- Sencillez.
- Ineficiencia: $O(n^2)$.
- Sensible a los parámetros escogidos.
- Las métricas de densidad son problemáticas en espacios de alta dimensionalidad.

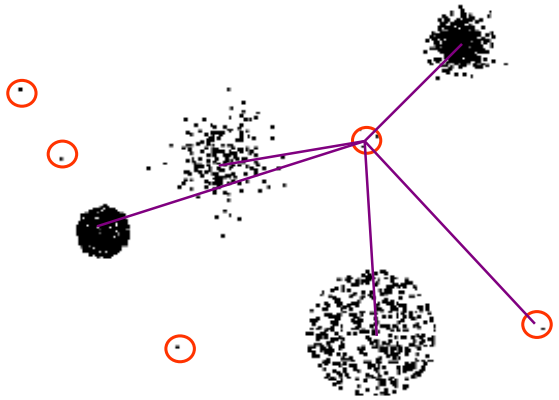
Detección de anomalías
Clustering

Idea

Un objeto es un outlier
si no pertenece claramente a ningún clúster.

- Clustering basado en prototipos:
  Outlier si no queda cerca de ningún centroide.
- Clustering basado en densidad:
  Outlier si su densidad es demasiado baja.
- Clustering basados en grafos:
  Outlier si no está bien conectado.

Detección de anomalías
Clustering

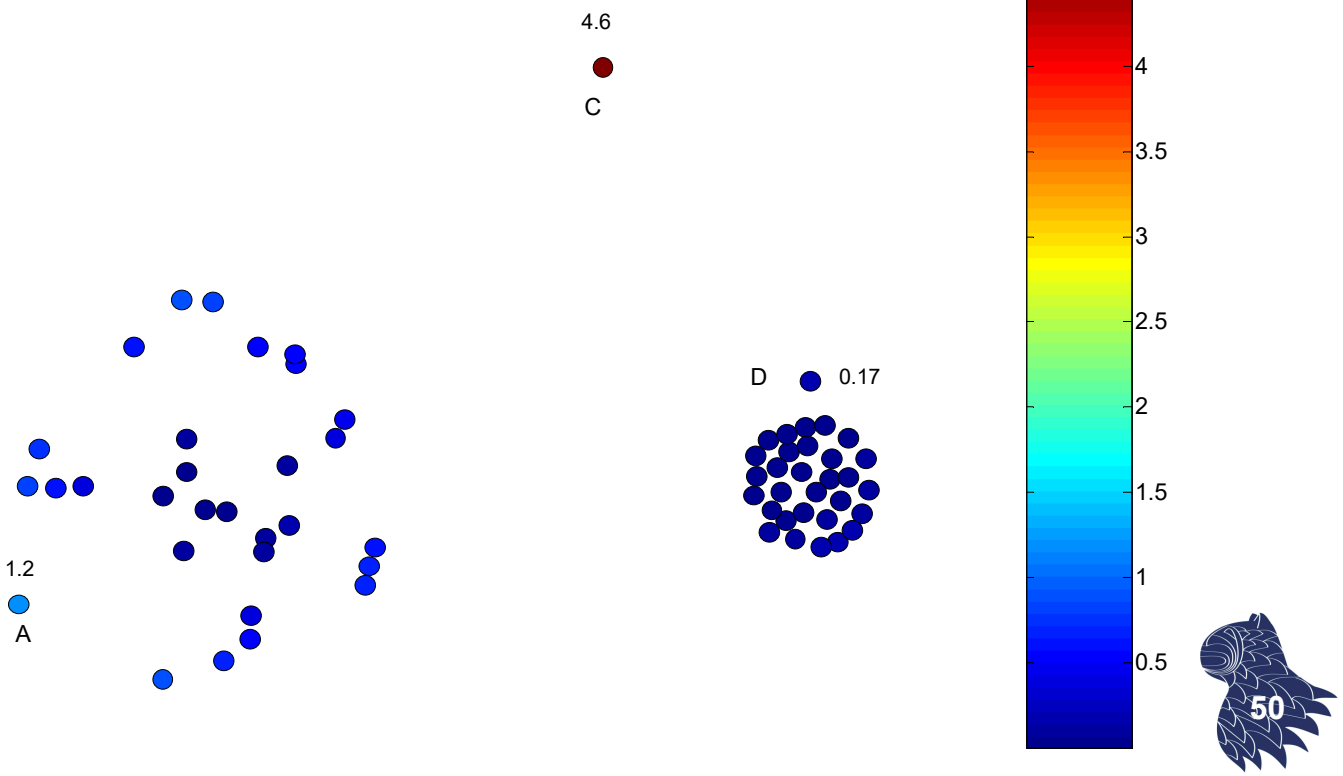


Posibles problemas

- Los outliers pueden afectar al algoritmo de clustering.
- Pueden confundirse ruido y outliers (p.ej. densidad).

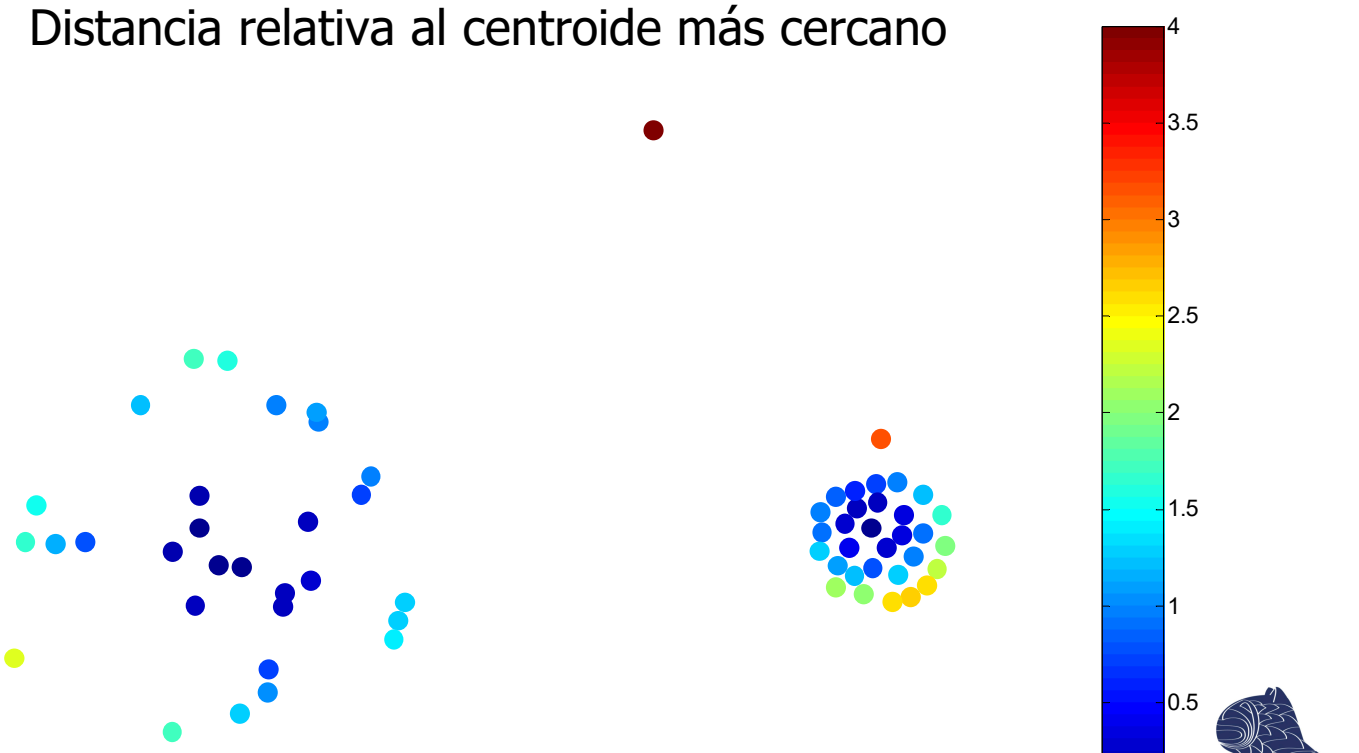
Detección de anomalías
Clustering

Distancia al centroide más cercano



Detección de anomalías
Clustering

Distancia relativa al centroide más cercano



Detección de anomalías
Clustering

Fortalezas y debilidades

- Sencillez.
- Múltiples algoritmos de clustering disponibles (puede ser difícil encontrar el más adecuado).
- Puede ser difícil ajustar los hiperparámetros del algoritmo de clustering (p.ej. número de clusters).
- La presencia de outliers distorsiona los clusters (y los resultados obtenidos por el algoritmo de clustering).

Detección de anomalías
Clasificación

Idea

Se resuelve la detección de anomalías como un problema de clasificación.

Problema

El conjunto de datos está muy sesgado (clases extremadamente poco balanceadas)

Por este motivo, se usan modelos de una clase:

- El clasificador sólo describe la clase normal.
- Cualquier ejemplo que no pertenezca a la clase normal se considera una anomalía/outlier.

Detección de anomalías
Clasificación

One class SVM

"origin" trick

Con un kernel gaussiano

$$\kappa(\mathbf{x},\mathbf{y}) = \exp(-\frac{||\mathbf{x}-\mathbf{y}||^2}{2\sigma^2})$$

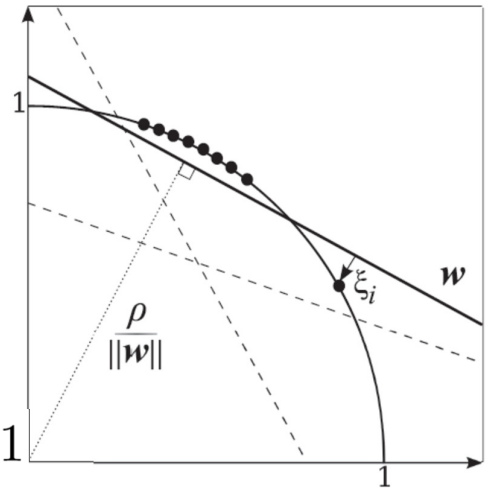
cada punto se proyecta

en una hiperesfera unitaria

$$\kappa(\mathbf{x},\mathbf{x}) = \langle\phi(\mathbf{x}),\phi(\mathbf{x})\rangle = ||\phi(\mathbf{x})||^2 = 1$$

en el mismo ortante (cuadrante)

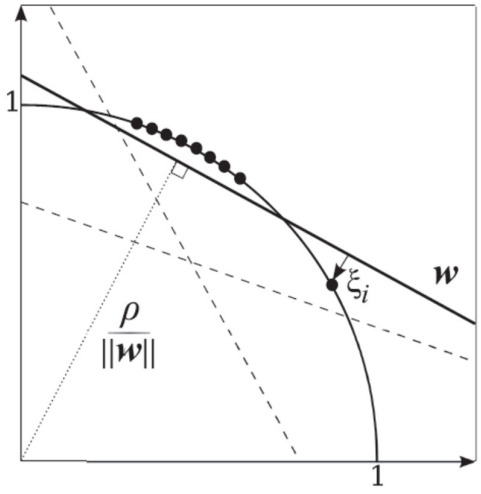
$$\kappa(\mathbf{x},\mathbf{y}) = \langle\phi(\mathbf{x}),\phi(\mathbf{y})\rangle \geq 0$$



Se maximiza la distancia
del hiperplano separador con el origen

Detección de anomalías
Clasificación

One class SVM



Ecuación del hiperplano

$$\langle\mathbf{w},\phi(\mathbf{x})\rangle = \rho$$

Vector de pesos

$$\mathbf{w} = \sum_{i=1}^{n} \alpha_i \phi(\mathbf{x_i})$$

Problema de optimización

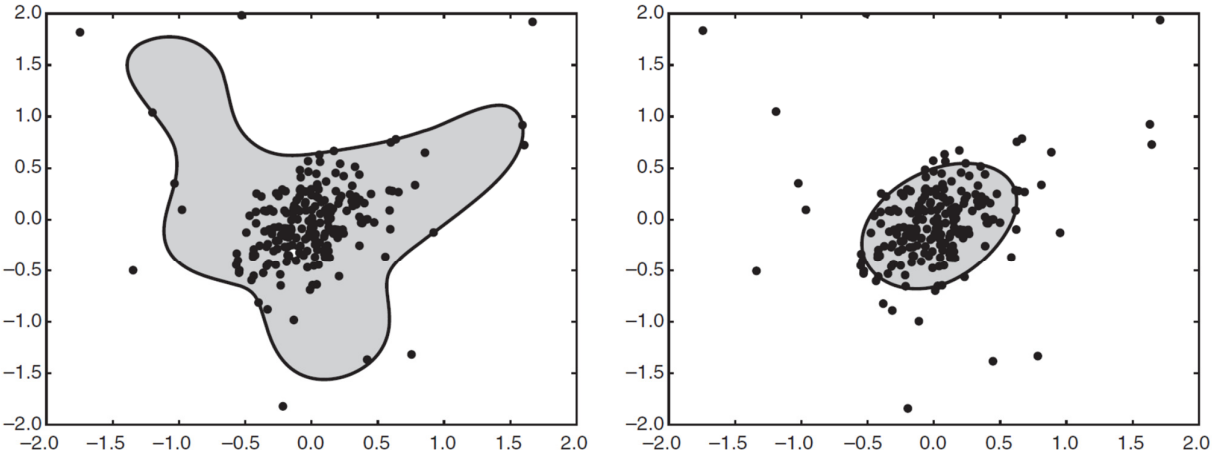
$$\min_{\mathbf{w},\,\rho,\,\xi} \frac{1}{2}||\mathbf{w}||^2 - \rho + \frac{1}{n\nu}\sum_{i=1}^{n}\xi_i,$$

$$\text{subject to: } \langle\mathbf{w},\phi(\mathbf{x_i})\rangle \geq \rho - \xi_i,\ \xi_i \geq 0$$

Detección de anomalías
Clasificación

One class SVM

con distintas fracciones ν de outliers



(a) $\nu = 0.05$. (b) $\nu = 0.2$.

Detección de anomalías
Clasificación

Deep Learning

De forma natural, las salidas de la última capa oculta de una red neuronal representan los datos de entrada de forma compacta (extrayendo características potencialmente útiles [embeddings])

- **OC-NN** [One-Class Neural Network] reemplaza el hiperplano del OC-SVM por una red neuronal @ 2018
- **DevNet** [Deviation Network] @ KDD'2019

R. Chalapathy & S. Chawla: **Deep learning for anomaly detection: A survey**. arXiv, 2019

Detección de anomalías
Clasificación

Estrategia semisupervisada
(datos parcialmente etiquetados)

- Se aplica un algoritmo de clustering.
- Si un cluster grande contiene principalmente objetos normales y no etiquetados, se tratan esos objetos no etiquetados como normales y se construye un modelo de una clase para identificar nuevos objetos normales.
- Si un cluster pequeño contiene algunos outliers y otros ejemplos no etiquetados, se consideran outliers.

Detección de anomalías
Asociación

Idea
Los algoritmos de extracción de reglas de asociación generan montones de reglas con asociaciones espúreas.

Se pueden diseñar medidas (objetivas o subjetivas) para medir el interés de una regla.

La clave es caracterizar correctamente las situaciones que realmente nos interesan,
p.ej. eventos sorprendentes poco frecuentes.

Detección de anomalías
Asociación

Reglas de excepción [exception rules]

Suzuki et al., PAKDD'2000 & IJPRAI'2002

$X \rightarrow Y$ is an association rule

$X\ I \rightarrow \neg Y$ is the exception rule

I is the "Interacting" itemset

$X \not\rightarrow I$ is the reference rule

Demasiadas excepciones

Detección de anomalías
Asociación

Reglas de asociación anómalas [anomalous association rules]

Berzal, Cubero et al., ICDM'2004 & KDD'2005

Reglas de confianza elevada que representan desviaciones homogéneas del comportamiento habitual.

Algoritmo eficiente de detección **ATBAR** basado en TBAR [Tree-Based Association Rules].

Detección de anomalías
Asociación

Reglas de asociación anómalas

X usually implies Y (dominant rule)

X ➔ Y frequent and confident

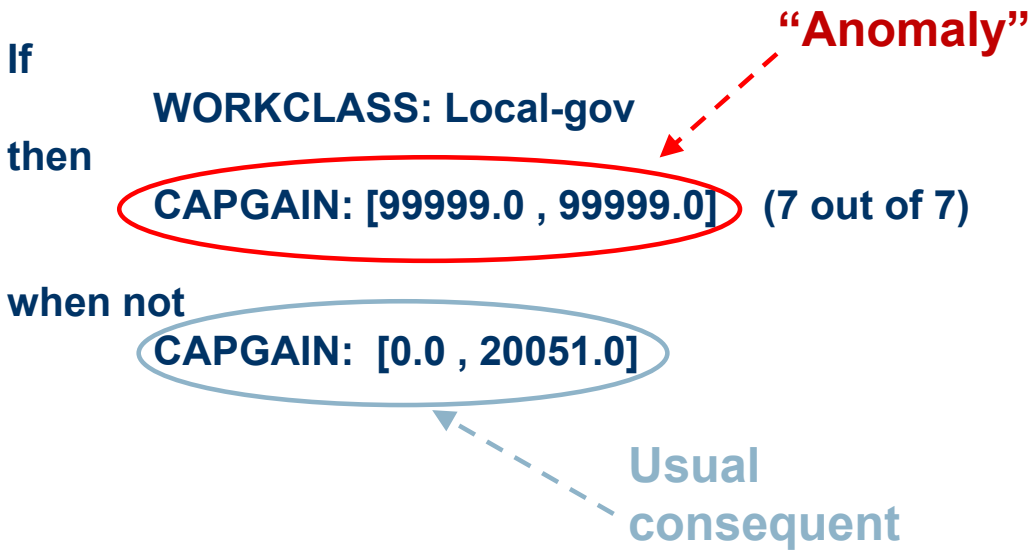
When X does not imply Y, then it usually implies A (the Anomaly)

X ¬Y ➔ A **Anomalous association rule**
confident

X Y ➔ ¬A confident

Detección de anomalías
Asociación

Reglas de asociación anómalas



Detección de anomalías
Reconstrucción

Idea

Existen patrones en la distribución de la clase normal que pueden capturarse con representaciones de menor dimensionalidad.

Estrategia

- Reducción de dimensionalidad, p.ej. PCA o autoencoders
- Medida del error de reconstrucción para cada objeto (diferencia entre el original y el derivado de la representación de menor dimensionalidad).

Detección de anomalías
Reconstrucción

Algoritmo de detección de anomalías basado en reconstrucciones

Dado un objeto original x:

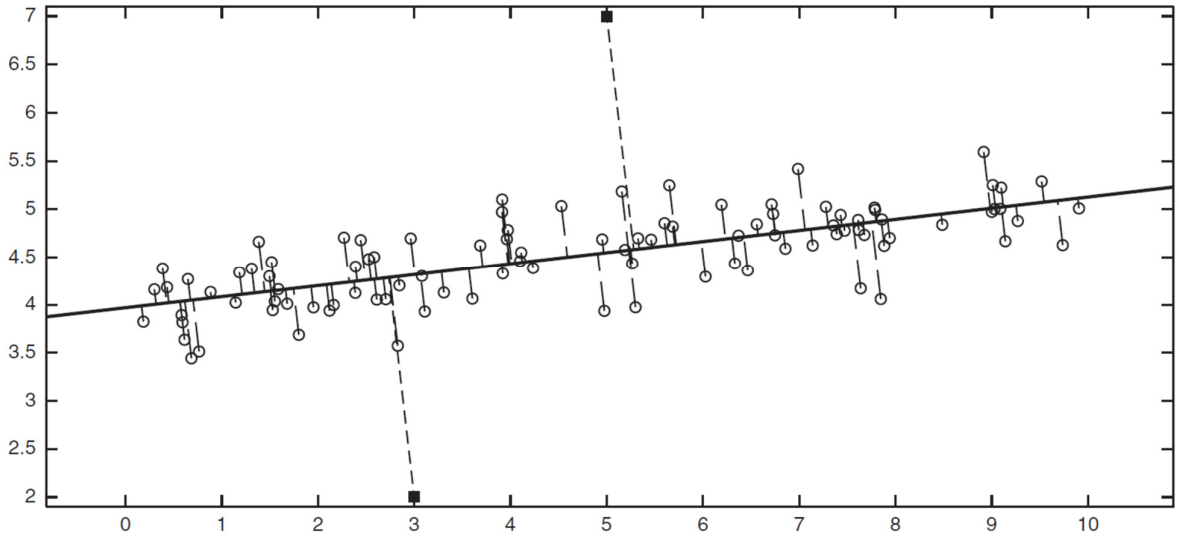
- Obtener la representación de x en menos dimensiones.
- Proyectar esa representación en el espacio original.
- Medir el error de reconstrucción:

$$error_{reconstruction}(\mathbf{x}) = \|\mathbf{x} - \hat{\mathbf{x}}\|$$

Las anomalías son los objetos con mayor error de reconstrucción

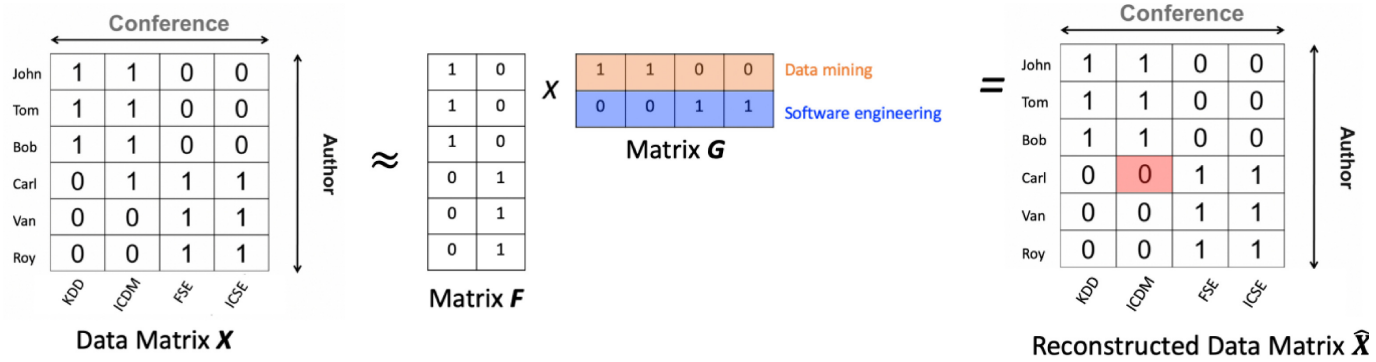
Detección de anomalías
Reconstrucción

EJEMPLO: Reconstrucción 2D

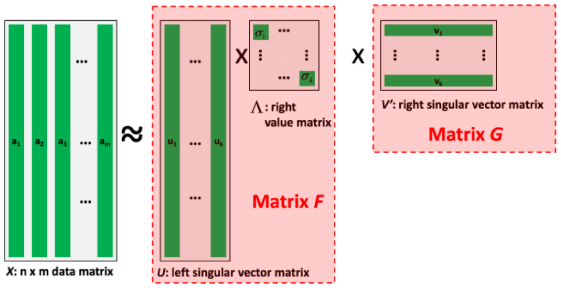


Detección de anomalías
Reconstrucción

Refactorización de matrices



EJEMPLO: SVD [Singular Value Decomposition]



Detección de anomalías
Reconstrucción

Creación de un código

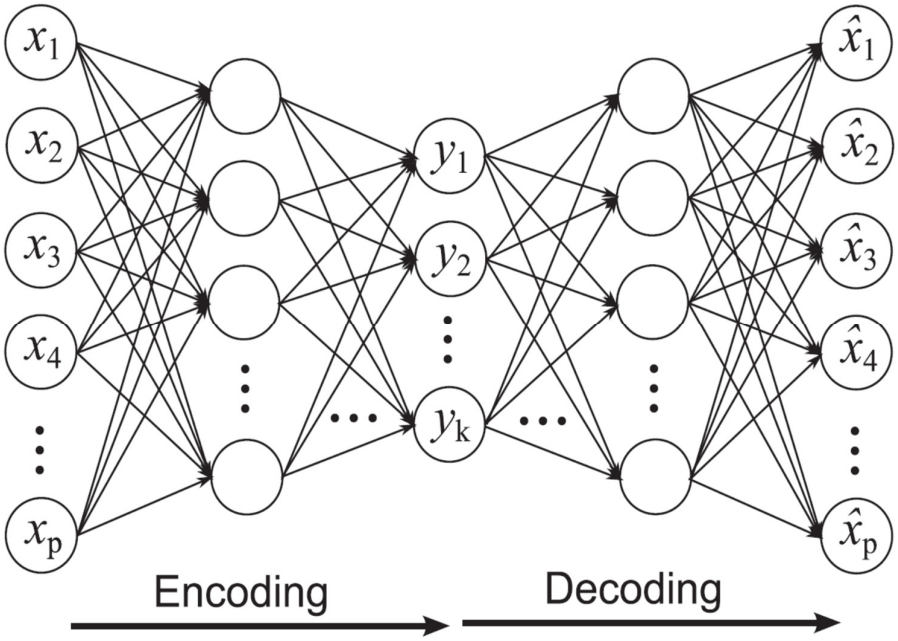
Code word	Code	Usage	Code Length
[I/H, C/H, P/H]	01	2	2
[I/L, C/L, P/L]	10	2	2
[I/H, C/L]	11	2	2
[P/M]	001	1	3
[P/L]	010	1	3

La longitud del código para representar un objeto nos indica su grado de anomalía [outlierness].

	Income	Credit	Purchase
John	High	High	High
Amy	High	High	High
Carl	Low	Low	Low
Mary	Low	Low	Low
Tom	High	Low	Medium
Jim	High	Low	Low

Detección de anomalías
Reconstrucción

Reconstrucción con autoencoders

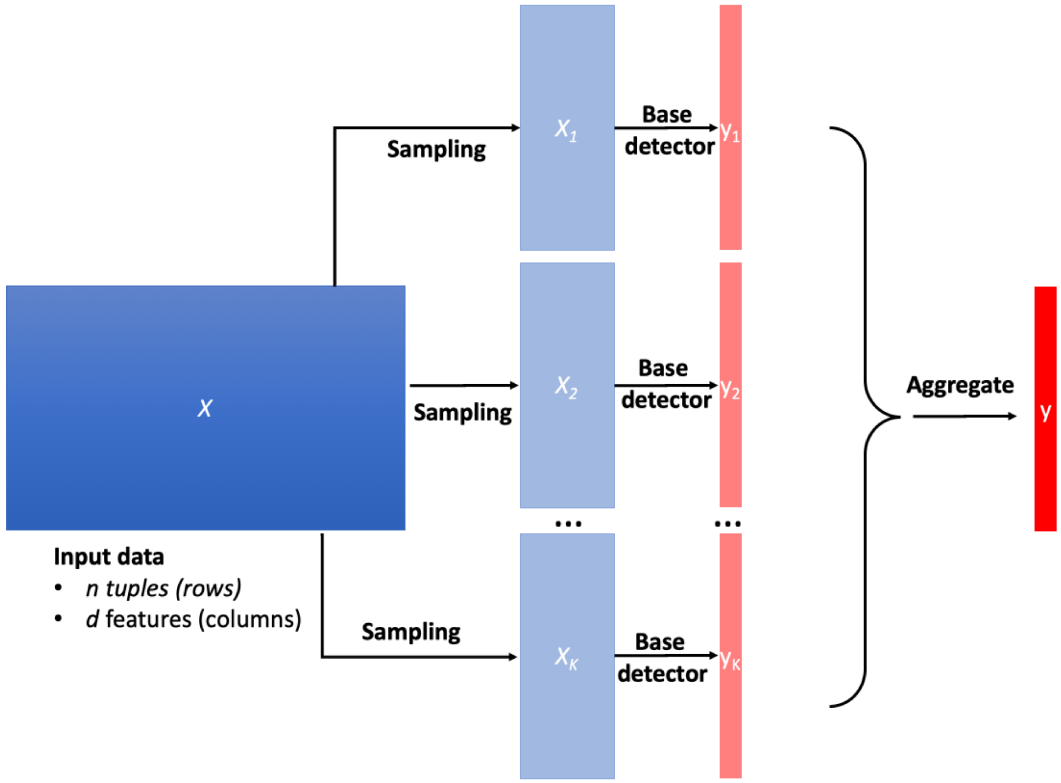


Detección de anomalías
Reconstrucción

Fortalezas y debilidades

- No requiere suposiciones sobre la distribución de la clase normal.
- Múltiples técnicas de reducción de la dimensionalidad.
- El error de reconstrucción se mide en el espacio original (puede ser un problema si la dimensionalidad de ese espacio es elevada).

Detección de anomalías
Ensembles



Detección de anomalías
Ensembles

En espacios de muchas dimensiones,
las medidas de proximidad/distancia se deterioran,
por lo que se construyen subespacios
en los que detectar anomalías más fácilmente:

- Se buscan anomalías en múltiples subespacios.
- Si se determina que un objeto es una anomalía en un subespacio, ese subespacio da información crítica para interpretar por qué y hasta qué punto el objeto es una anomalía.

Detección de anomalías
Ensembles

Subespacios aleatorios

- **Feature bagging**:
  Se seleccionan aleatoriamente unas cuantas características/variables de los datos originales.
- **Rotated bagging**:
  Se genera un subespacio aleatorio de dimensionalidad d (d≪D) y se proyectan los datos en ese subespacio.

Ensembles

Agregación de resultados de los detectores base

- **Media [mean]**:
  Promedio de los K detectores base.

- **Máximo [max]** :
  Mayor grado de anomalía detectado por los K detectores base (que pueden ser más selectivos).

Es importante normalizar (min-max, z-score) los grados de los detectores base antes de la agregación.

Evaluación de resultados

Detección supervisada de anomalías

Como en otros problemas de clasificación, se usan métricas estándar que resulten adecuadas para clases poco frecuentes:

- Precisión [precision, no accuracy]

- Sensibilidad [recall] = TPR [True Positive Rate]

- FPR [False Positive Rate]

Evaluación de resultados

Detección no supervisada de anomalías

Se suelen usar las medidas propias del método de detección de anomalías:

p.ej.

- Error de reconstrucción
- Ganancia de información

Evaluación de resultados

Es aconsejable consultar el histograma de grados de anomalía [anomaly scores], que debería tener una cola:

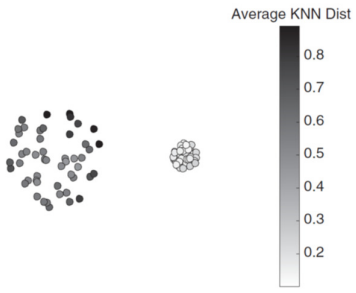


Figure 10.17. Anomaly score based on average distance to fifth nearest neighbor.

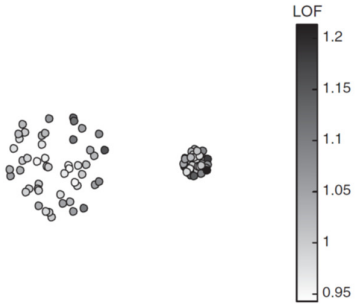
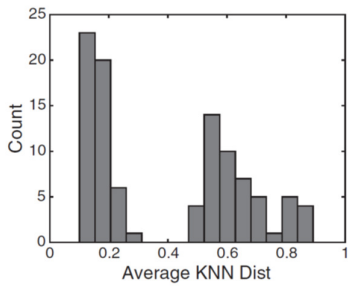
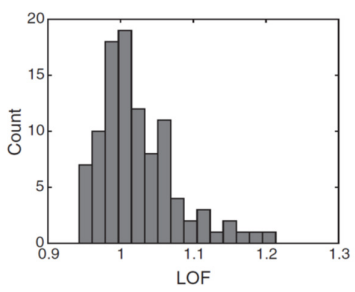


Figure 10.18. Anomaly score based on LOF using five nearest neighbors.





Evaluación de resultados

Dependiendo del problema, puede que nos interesen distintos tipos de anomalías:

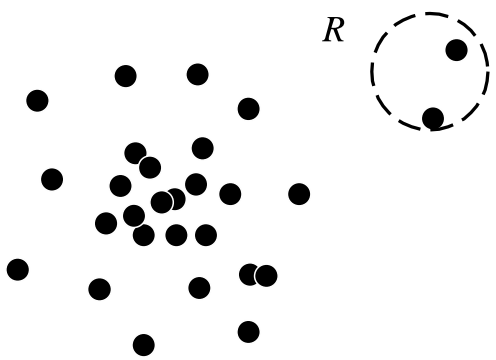
- Anomalías globales/puntuales
- Anomalías contextuales/condicionadas
- Anomalías colectivas

Evaluación de resultados

Anomalías globales

Un objeto es una anomalía global o puntual si se desvía significativamente del resto del conjunto de datos.

Tipo más simple de anomalía que podemos detectar:



Evaluación de resultados

Anomalías contextuales/condicionadas

Dado un conjunto de datos, un objeto es una anomalía contextual/condicionada si se desvía significativamente con respecto al contexto del objeto.

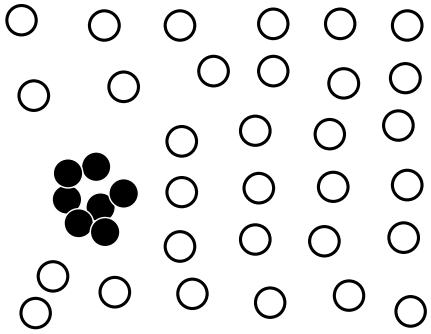
e.g. 40ºC en Granada puede ser una anomalía en noviembre, pero no en agosto.

- Generalización de la idea de anomalía local (propuesta por los métodos basados en densidad).
- Generalización de las anomalías globales (un caso particular de las contextuales, con contexto vacío).

Evaluación de resultados

Anomalías colectivas

Dado un conjunto de datos, un subconjunto de esos datos forma una anomalía colectiva si, en conjunto, se desvía significativamente del resto de los datos, aunque individualmente cada dato no sea anómalo.



Evaluación de resultados

Desafíos de la detección de anomalías

- Modelado de la "normalidad" (la frontera entre objetos normales y anomalías no siempre está clara).
- Las relaciones existentes entre objetos son muy dependiented de la aplicación específica.
- La presencia de ruido puede dificultar la detección de anomalías.
- El usuario puede estar interesado en detectar anomalías, sino también en entender por qué lo son (interpretabilidad).

Evaluación de resultados

Problemas adicionales: masking & swamping

Una anomalía enmascara [masks] a otra, si la segunda puede considerarse una anomalía por sí misma, pero no en la presencia de la primera:

- Tras eliminar la primera anomalía, la segunda emerge.
- Causa: Un grupo de anomalías sesga el modelo de lo que se considera "normal", haciendo que las anomalías parezcan más cercanas a la normalidad.

Evaluación de resultados

Problemas adicionales: masking & swamping

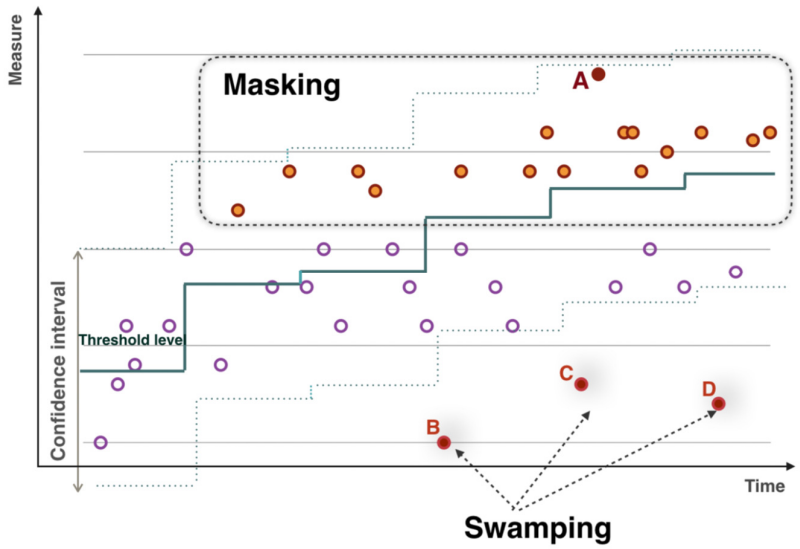
Una anomalía empantana [swamps] a otra observación si la segunda se considera anómala sólo por la presencia de la primera:

- Tras eliminar la primera anomalía, la segunda desaparece.
- Causa: Un grupo de anomalías sesga el modelo de lo que se considera "normal", haciendo que los datos normales parezcan más lejanos a la normalidad.

Evaluación de resultados

Problemas adicionales: masking & swamping

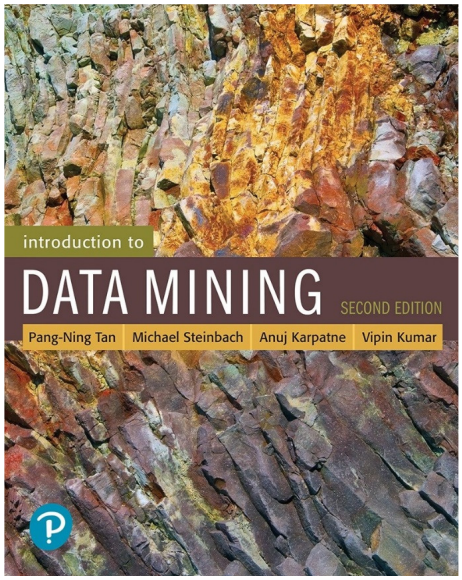
Pueden aparecen siempre que hay grupos de anomalías



Zahra Zohrevand & Uwe Glässer: "Should I Raise The Red Flag? A comprehensive survey of anomaly scoring methods toward mitigating false alarms", arXiv'2020

Bibliografía

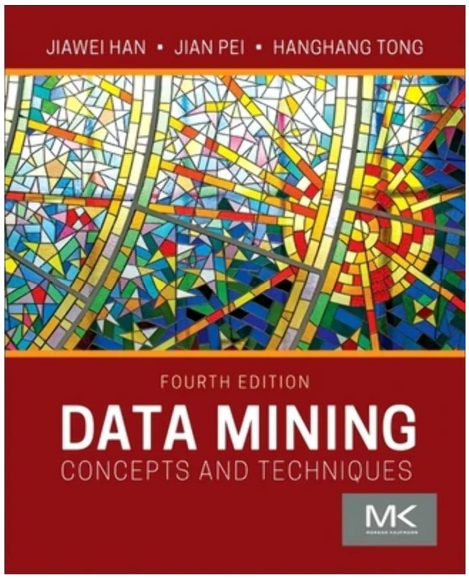
Pang-Ning Tan,
Michael Steinbach,
Vipin Kumar &
Anuj Karpatne:
Introduction to Data Mining,
2nd edition, Addison Wesley, 2018.
ISBN 0133128903



9 Anomaly Detection
10.6 Statistical Testing for Anomaly Detection

Bibliografía

Jiawei Han,
Jian Pei &
Hanghang Tong:
Data Mining:
Concepts and Techniques,
4th edition, Morgan Kaufmann, 2022.
ISBN 0128117605

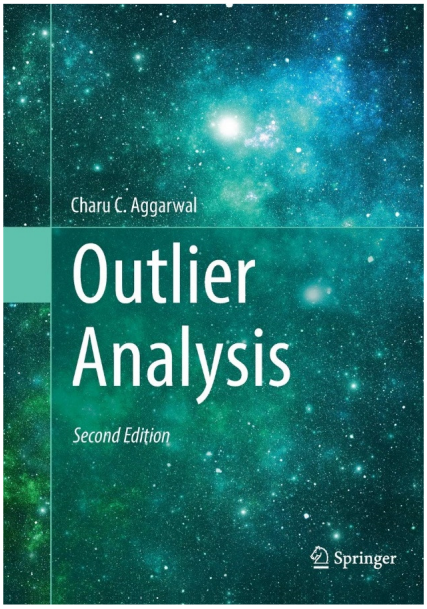


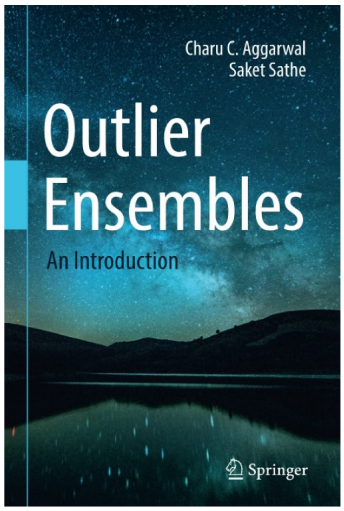
11 Outlier detection

Bibliografía

- Charu C. Aggarwal:
  Outlier Analysis.
  2nd edition, Springer, 2017.
  ISBN 3319475770.

- Charu C. Aggarwal
  & Saket Sathe:
  Outlier Ensembles.
  Springer, 2017.
  ISBN 331954764X.





Bibliografía

Andrew T. Ng:
Machine Learning:
Anomaly Detection
http://ml-class.org/

Originalmente (2011)
http://mlclass.stanford.edu/