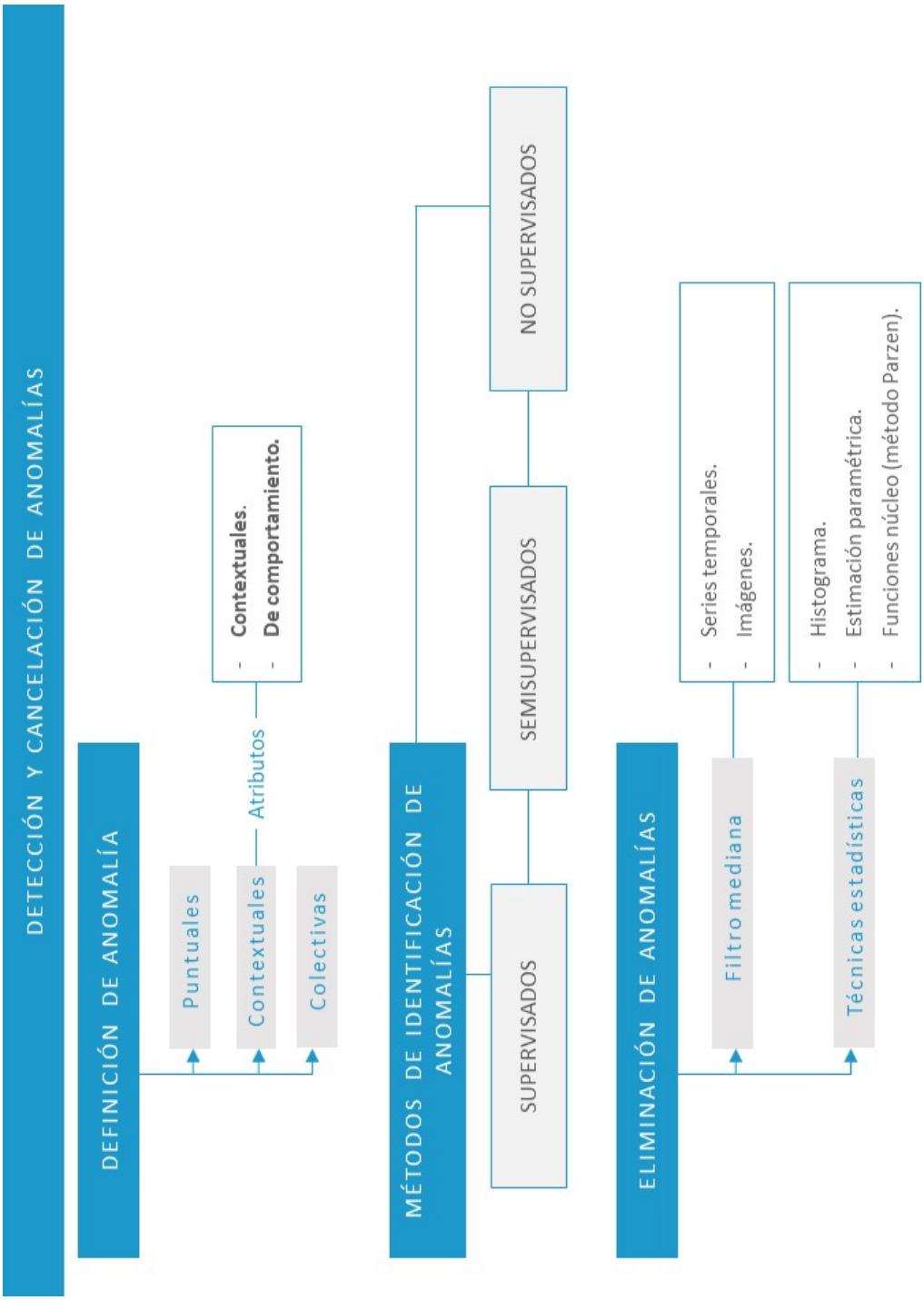
Visión Artificial

# Tema 5. Detección y cancelación de anomalías

# Índice

# Esquema



## 5.1. ¿Cómo estudiar este tema?

Para estudiar este tema deberás leer con atención las ideas clave que se desarrollan a continuación.

En el siguiente tema se aborda, como una etapa más en la preparación de las señales capturadas, la detección y eliminación de muestras de datos atípicas conocidas como anomalías *(outliers).*

El tema se organiza de la siguiente forma:

- Definición de anomalía. En primer lugar, se proporciona una definición de anomalía en el contexto del procesado de datos y señales. Así, se indican los **diferentes tipos de anomalías** que podremos observar en un conjunto de datos y las características de cada uno de ellos.
- Métodos de identificación. Se indica la taxonomía de los métodos que pueden emplearse para la **identificación** de anomalías.
- Eliminación de anomalías. Por último, se profundiza en aquellos métodos que son comúnmente empleados para la **detección y eliminación** de anomalías en señales, dado que estas son la fuente de información con la que se trabaja en la asignatura.

El siguiente esquema permite memorizar de forma sencilla los elementos de este tema.

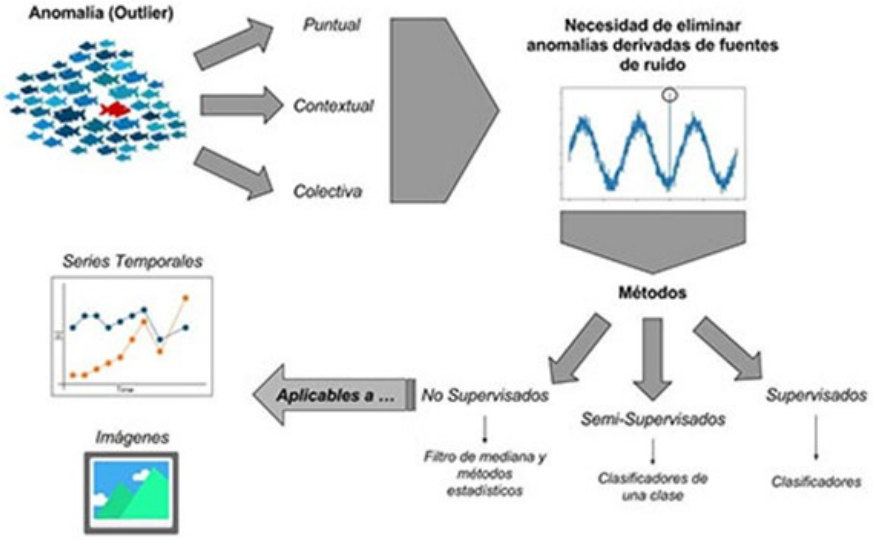



Figura 1. Conceptos relacionados con la presencia e identificación de anomalías en señales como series temporales e imágenes.

## 5.2. Definición de anomalía

La detección de anomalías tiene como objetivo la **identificación de valores atípicos en la fuente de información,** comúnmente conocidos por su vocablo en inglés, *outliers.* Estos se definen como patrones inusuales que no se ajustan al comportamiento esperado. La aparición de *outliers* en una señal o imagen refleja la existencia de ruido, generalmente de tipo impulsivo motivado, por ejemplo: por un valor de pico en un campo eléctrico cercano, o de inestabilidades en el procedimiento de captura, por ejemplo: el movimiento brusco de una cámara.

La detección de anomalías tiene aplicación directa en diferentes escenarios prácticos. A continuación se citan algunas:

- Detección de intrusos en una red. Identificación de patrones atípicos en el tráfico de red que pueden indicar un ataque.
- Diagnóstico médico. Reconocimiento de lesiones con una bajo índice en la población que pueden indicar la existencia de alguna patología.
- Detección de transacciones fraudulentas. La gran mayoría de las transacciones que se realizan son lícitas y solo una pequeña proporción se corresponden con actividades fraudulentas.
- Predicción de fuga de clientes en grandes compañías. En los sectores de banca, seguros y telecomunicaciones, una pequeña parte de los clientes abandona la compañía, por lo que la identificación de estos comportamientos puede realizarse mediante técnicas de detección de anomalías.

Existen diferentes tipos de anomalías. Vamos a definir cada uno de ellos:

## Anomalías puntuales

En el caso de que una muestra individual pueda considerarse notablemente diferente respecto al resto de los datos, esta puede ser tomada como un outlier. Este tipo de anomalía es el más simple y el foco de la mayoría de los trabajos de investigación sobre este tema.

Un claro ejemplo correspondiente a un escenario real sería el fraude cometido con una tarjeta de crédito. Si nos fijamos en una variable como la cuantía de la transacción, aquellas transacciones para las que la cantidad sea muy alta en comparación con el promedio del resto de transacciones previas son susceptibles de ser anomalías puntuales y, por tanto, sospechosas de fraude. Así, una anomalía puntual se expresa mediante la aparición de **valores pico** que se alejan excesivamente del conjunto de valores que encontramos.

En la siguiente figura podemos ver una señal en la que una de las muestras toma un valor que no observamos en ninguna otra. Se trata, claramente, de una muestra candidata a ser una anomalía.

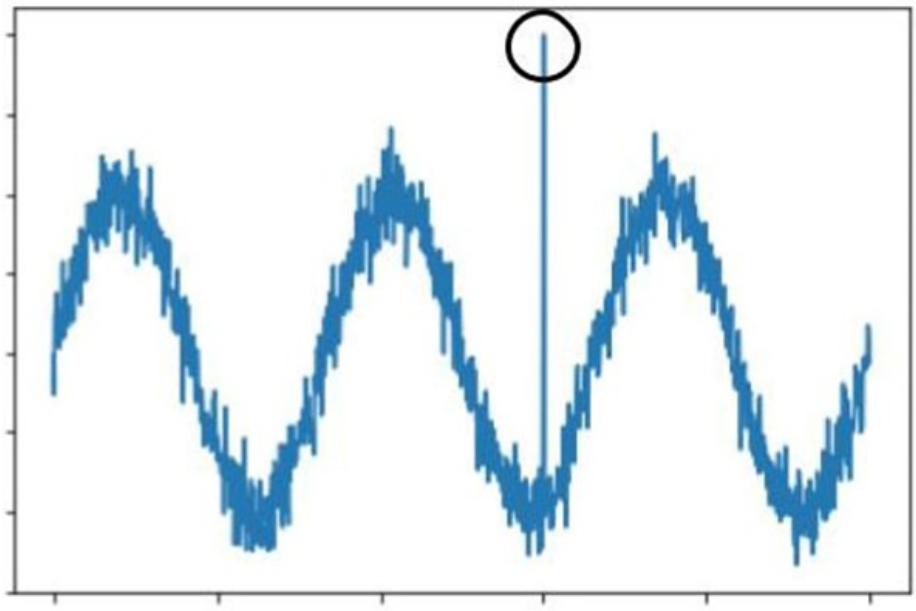


Figura 2. Ejemplo de anomalía puntual en una señal temporal.

## Anomalías contextuales

Si una muestra de datos es anómala en un contexto específico (pero no de otro modo), se denomina anomalía contextual. La noción de contexto viene dada por la naturaleza de los datos. Cada muestra de datos se define teniendo en cuenta los siguientes atributos.

- **Atributos contextuales:** estos se usan para determinar el contexto (o vecindad) para esa muestra. Vienen dados por la **naturaleza de la fuente de datos.** Por ejemplo, en conjuntos de datos espaciales, la longitud y la latitud de una ubicación son los atributos contextuales. En una serie temporal, el tiempo es un atributo contextual que determina la posición de una muestra en toda la secuencia.
- **Atributos de comportamiento:** estos definen el carácter no contextual de una instancia. Es decir, representa el **valor de la muestra.** Siguiendo con el ejemplo de los datos espaciales, si se trata de cuantificar la precipitación promedio en cualquier punto de la superficie mundial, la cantidad de lluvia en cualquier lugar es un atributo de comportamiento.

El comportamiento anómalo se determina usando los valores de los atributos de comportamiento dentro de un contexto específico.

Una instancia de datos podría ser una anomalía contextual en un determinado contexto, pero una instancia de datos idéntica (en términos de atributos de comportamiento, es decir, de su valor) podría ser considerada normal en un contexto diferente. Esta propiedad es clave para identificar atributos contextuales y de comportamiento para una técnica de detección de anomalía contextual.

A diferencia de las anomalías puntuales definidas previamente, en las que únicamente se lleva a cabo una comparación de las muestras de datos disponibles para identificar un valor atípico, en señales temporales (series temporales) e imágenes se tiene en cuenta el contexto para definir un valor anormal.

Por ejemplo, en una imagen, es posible identificar un píxel anómalo si su intensidad es muy diferente a la de los píxeles vecinos. De la misma forma, en una serie temporal también será el entorno de un punto el que nos dé la información contextual necesaria para identificar un valor anómalo, tal y como se ejemplifica en la siguiente imagen.

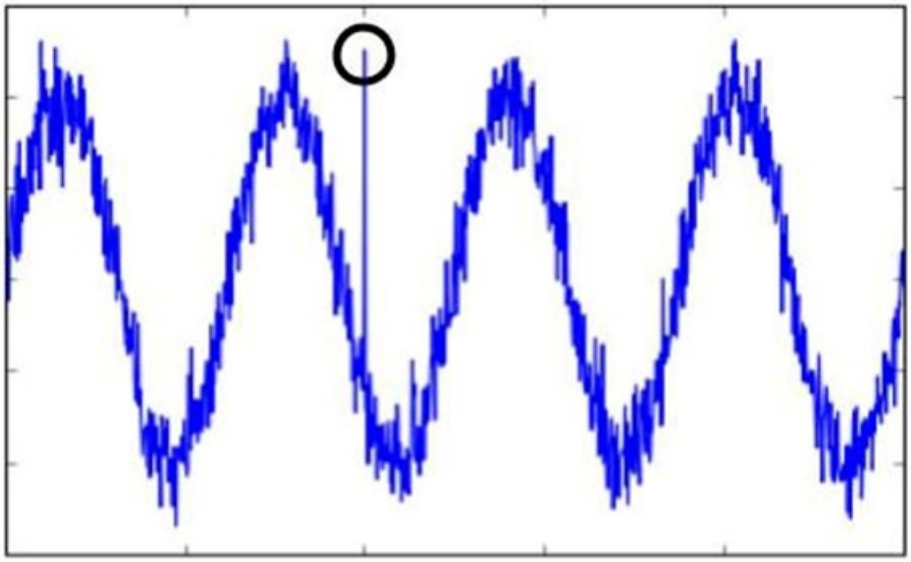


Figura 3. Ejemplo de anomalía contextual identificada en una serie temporal.

En ella vemos que la serie toma en algún momento valores similares al de la anomalía, pero el contexto nos indica que en este caso es una muestra atípica.

## Anomalías colectivas

Si una **colección de instancias de datos** relacionadas es anómala con respecto a todo el conjunto de datos, se denomina anomalía colectiva. Las instancias de datos individuales en una anomalía colectiva pueden no ser anomalías por sí mismas, pero su ocurrencia conjunta como colección es anómala.

El siguiente gráfico ilustra un ejemplo de anomalía colectiva en una señal electrocardiográfica. La región resaltada denota una anomalía porque la señal toma aproximadamente el mismo valor durante un tiempo inusualmente largo. Sin embargo, ese valor no es una anomalía por sí mismo.

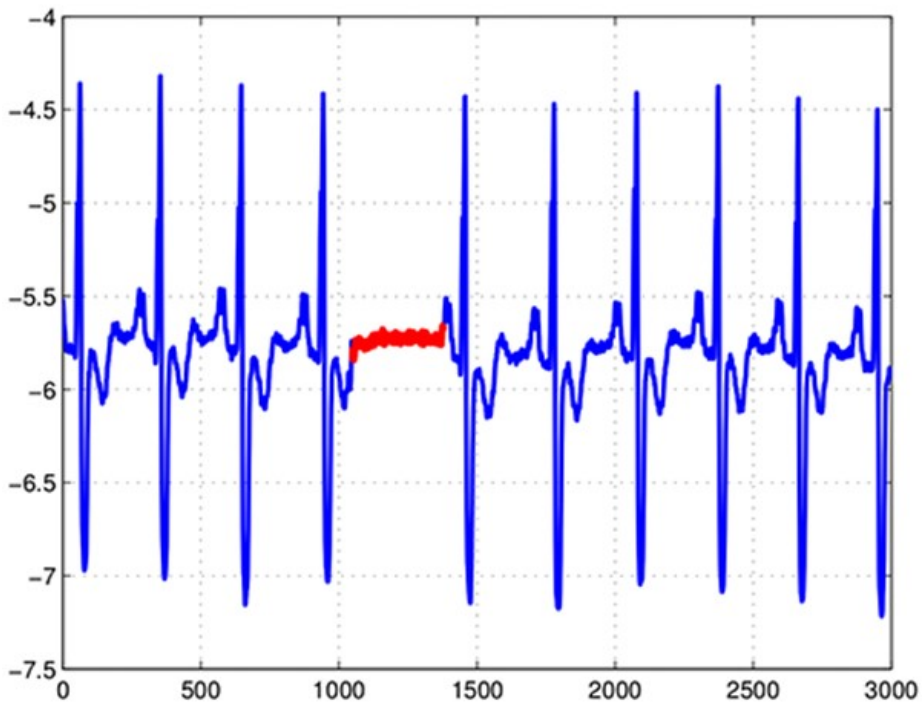



Figura 4. Ejemplo de anomalía colectiva en una señal de ECG. Fuente: https://www.datasciencecentral.com/profiles/blogs/anomaly-detection-for-the-oxford-data-science-for-iot-course

# Ideas clave

En este tema se tratará de identificar la presencia de anomalías en nuestras fuentes de información como una etapa más en la limpieza y preparación de la señal. Nos ceñiremos al escenario en el que la anomalía se corresponde con un artefacto ruidoso que debería ser eliminado a fin de preservar la calidad de la información.

En este sentido, los tres tipos de anomalías descritos previamente pueden tener como origen un elemento perturbador de la señal. Generalmente, las anomalías o artefactos a los que nos enfrentaremos vendrán dados por valores de pico inusuales derivados de la presencia de **fuentes ruidosas de tipo impulsivo.**

## 5.3. Métodos de identificación de anomalías

A diferencia de los problemas convencionales de clasificación, donde se cuenta con un conjunto de datos de entrenamiento, previamente etiquetados con su correspondiente clase, y un conjunto de test que permite la estimación objetiva del rendimiento del modelo, hay múltiples configuraciones posibles cuando se habla de detección de anomalías.

Básicamente, la configuración de detección de anomalías que se utilizará depende de las etiquetas disponibles en el conjunto de datos. Así, podemos distinguir entre tres tipos principales:

### Métodos supervisados

En este caso, se dispone de un conjunto de datos en los que cada muestra está asociada a una etiqueta que indica si es o no una anomalía. A partir de estos datos, se puede **entrenar un clasificador que capture el patrón característico de la anomalía.** Posteriormente, este clasificador se emplearía como identificador automático de estos valores atípicos.

Este escenario es muy similar al reconocimiento tradicional de patrones en el que las dos clases (no anomalías vs. anomalía) están fuertemente desequilibradas. Como consecuencia, debe tenerse en cuenta que no todos los algoritmos de clasificación se adaptan perfectamente a esta tarea, pues depende de la función de error que trate de optimizarse durante su entrenamiento.

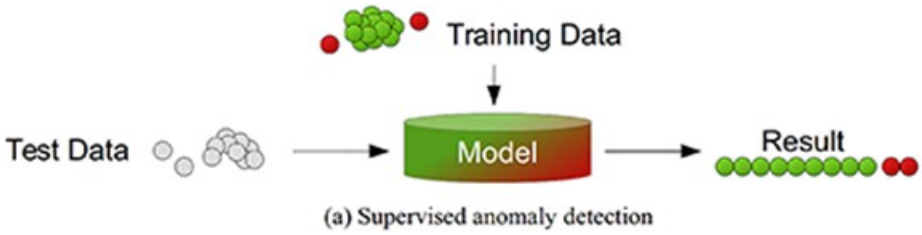



Figura 5. Métodos y escenarios para la implementación de técnicas de detección de anomalías: supervisado. Fuente: Goldstein y Uchida, 2016.

Por ejemplo: los árboles de decisión, que buscan divisiones de los datos en las que se maximice la pureza de las categorías, no responden bien a conjuntos desequilibrados, mientras que las máquinas de vector de soporte (SVM, *Support Vector Machines*) o las redes neuronales artificiales (ANN, *Artificial Neuronal Network)* proporcionarán un resultado más plausible.

Sin embargo, la identificación de anomalías basada en técnicas supervisadas **no está muy extendida,** pues el punto de partida es la suposición de que las anomalías son conocidas y etiquetadas correctamente. Para muchas aplicaciones, las anomalías no se conocen de antemano o pueden ocurrir espontáneamente como novedades durante la fase de test, por lo que estas técnicas no tendrían cabida.

La utilización de modelos derivados del aprendizaje supervisado como herramientas de identificación de anomalías es común en aplicaciones prácticas como la detección de fraude o el diagnóstico médico, donde el número de muestras positivas es notablemente menor en proporción y, por tanto, se consideran una anomalía.

## Métodos semisupervisados

Se trata de un procedimiento similar al anterior, pues también se emplea un conjunto de datos para entrenar un modelo. Sin embargo, en este caso, los datos de entrenamiento **contienen únicamente muestras no anómalas.** La idea fundamental es que el modelo aprenda la clase normal, de forma que se detectaría la anomalía al identificarse una desviación del patrón aprendido.

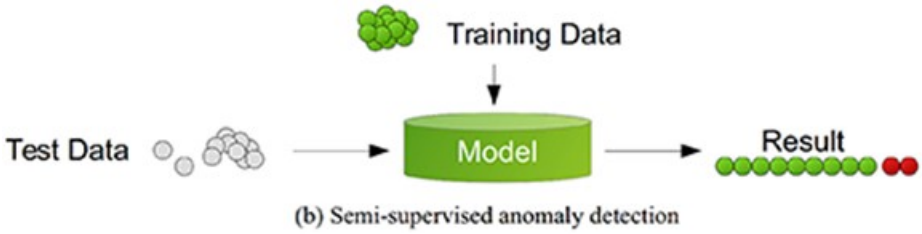



Figura 6. Métodos y escenarios para la implementación de técnicas de detección de anomalías: semisupervisado. Fuente: Goldstein y Uchida, 2016.

Esta aproximación se conoce como clasificadores de una sola clase *(one-class classifiers)*. Usualmente, suelen emplearse para tal fin los algoritmos SVM de una clase y los *autoencoders.* Además, cualquier método de estimación de densidad puede usarse para modelar la función de densidad de probabilidad de las muestras normales. Por ejemplo, los modelos de mezcla gaussiana o la estimación basada en funciones núcleo.

## Métodos no supervisados

Representa el procedimiento **más flexible** para la implementación de un método de identificación de anomalías pues no requiere de ningún tipo de conocimiento previo sobre los datos (etiquetas). La idea es que un algoritmo de detección de anomalías no supervisado califique los datos únicamente en función de las propiedades intrínsecas de estos. Normalmente, las **distancias o densidades** se utilizan para dar una estimación de lo que es normal y lo que es un valor atípico.

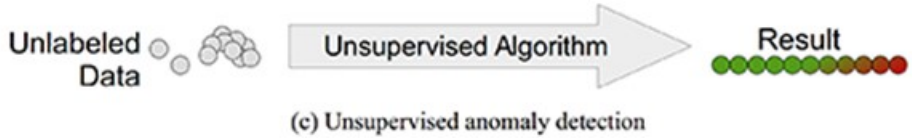



Figura 7. Métodos y escenarios para la implementación de técnicas de detección de anomalías: no supervisado. Fuente: Goldstein y Uchida, 2016.

La salida devuelta por un algoritmo de detección de anomalías puede ser de dos tipos diferentes. Por un lado, puede emplearse **una etiqueta para cada muestra.** Esta indica de forma categórica si una instancia es anómala o no. Este es el caso, por lo general, de los métodos basados en aprendizaje supervisado.

En segundo lugar, puede obtenerse un **valor continuo** que refleje el grado de anormalidad de la muestra. Los algoritmos de detección de anomalías semisupervisados y no supervisados producen una salida de este tipo. Esto se debe a razones prácticas, pues el procedimiento de identificación solo informa de los casos más sospechosos, por lo que es necesario establecer un orden en las instancias a partir del grado de anomalía asignado.

# Ideas clave

El objetivo de la asignatura es el manejo de señales que, generalmente, se corresponden con:

- Series temporales: señales unidimensionales con el tiempo como variable independiente.
- Imágenes: señales bidimensionales con las coordenadas espaciales como variables independientes.

En este contexto, como se indicó previamente, las anomalías son el resultado de una fuente ruidosa que conlleva la aparición de elementos espurios en la señal. Por tanto, el propósito de la etapa de detección de anomalías es eliminarlas para mejorar la calidad de la señal en etapas posteriores. Es decir, no formaría parte de esta etapa la identificación de patrones anómalos de naturaleza no ruidosa en la señal.

Por ejemplo: el resultado de una arritmia reflejada en una señal de ECG o la presencia de determinados objetos en una imagen. De acuerdo a este contexto, el procedimiento habitual para la detección y eliminación de anomalías será la utilización de **métodos no supervisados** dado que, inicialmente, no disponemos de ninguna descripción de las posibles anomalías que pudiera haber en la señal.

## 5.4. Eliminación de anomalías

A continuación se explican los procedimientos no supervisados más habituales para la eliminación de anomalías en señales.

### Filtro de mediana

El filtro de mediana ha sido comúnmente empleado sobre señales 1D y 2D para la eliminación de ruido impulsivo. Estos artefactos se reconocen fácilmente mediante la inspección visual de la señal, pues están asociados a valores de pico que destacan notablemente sobre el resto de la señal (anomalía puntual) o sobre el entorno más cercano (anomalía de tipo contextual, pues el valor atípico iría en discordancia con aquellos de su entorno).

En imágenes, este tipo de anomalías se conoce como ruido «sal y pimienta» *(salt & pepper),* ya que el efecto que genera es el de píxeles colocados de forma aleatoria que toman valores extremos de intensidad (1 o 0). Aquí vemos una imagen afectada por este tipo de ruido.

Figura 8. Imagen con ruido de tipo «sal y pimienta» reflejado en píxeles con una valor extremo de intensidad. Fuente: http://in.mathworks.com/matlabcentral/mlc-downloads/downloads/submissions/16201/versions/3/previews/toolbox_image/html/content.html

El filtro de mediana es una operación que se aplica punto a punto mediante una ventana deslizante. El tamaño de esta viene determinado por el usuario.

En el caso de señales unidimensionales como series temporales, se trata de una ventana de longitud N, mientras que en imágenes la ventana se define en ambas coordenadas y es de tamaño N x N. El valor de N es impar, ya que la ventana se centra en el punto de la señal que se pretende filtrar. Así, el valor resultante en este punto viene dado por la mediana de los puntos considerados por la ventana.

El filtro, tal y como se puede apreciar en su definición, no crea nuevos valores de la señal, sino que **selecciona como salida uno de los valores entrantes.** A continuación veremos el resultado de aplicar el filtro de mediana a una serie temporal y a una imagen.

En la mitad izquierda del siguiente gráfico tenemos una serie temporal donde pueden observarse dos puntos que representan sendas anomalías frente al resto de valores de la señal. Para eliminar estos *outliers,* se aplica un filtro de mediana con una ventana de longitud 3. El resultado del filtro aparece en la parte derecha del gráfico. Tal y como se aprecia, el filtro permite suprimir los valores atípicos, estimando su valor como la mediana de los puntos de su vecindad.

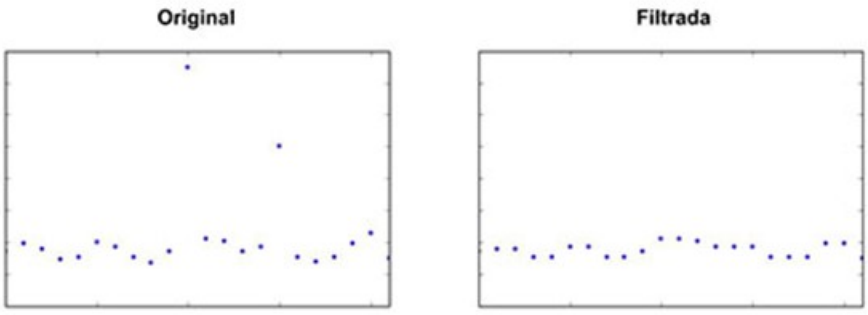



Figura 9. Aplicación de un filtro de mediana sobre una serie temporal afectada por la presencia de outliers.

Y a continuación vemos el resultado de utilizar el filtro de mediana para la eliminación de ruido de tipo «sal y pimienta» en la imagen anterior (figura 8). El tamaño de ventana empleado ha sido 3 x 3. Como puede verse en la imagen filtrada, la aplicación del filtro conlleva cierta distorsión de la imagen, pues se modifica el valor de intensidad de los píxeles que no están afectados por este tipo de ruido.

El filtro de mediana es muy similar a un filtro de promedio que obtendría, para cada ventana, el valor medio de los píxeles o puntos considerados. Esta operación es equivalente a la utilización de un filtro paso-bajo en frecuencia, por lo que las variaciones rápidas de la señal, reflejadas como contrastes significativos en una imagen, quedan suavizados por el filtro.

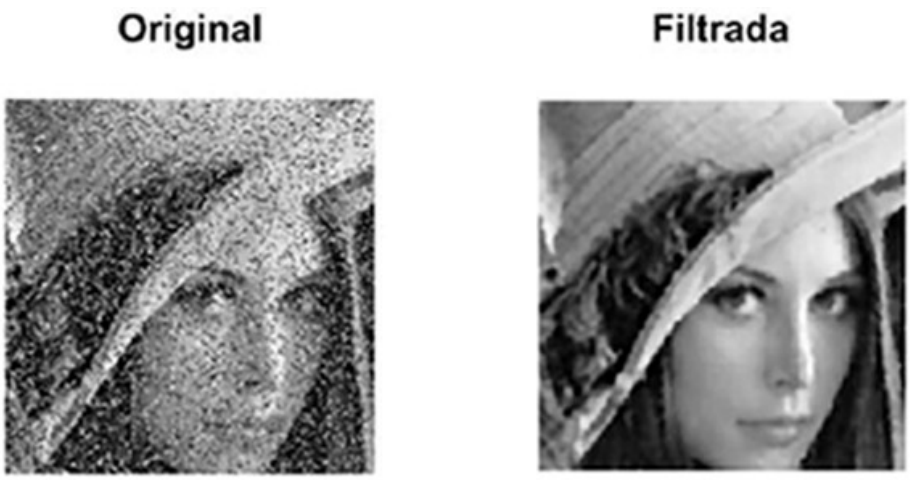



Figura 10. Efecto del filtro de mediana sobre una imagen con ruido impulsivo «sal y pimienta». Fuente: Adaptado de http://in.mathworks.com/matlabcentral/mlc-downloads/downloads/submissions/16201/versions/3/previews/toolbox_image/html/content.html#17

## Técnicas estadísticas

Otra técnica de detección y corrección de anomalías empleada habitualmente se basa en la utilización de la **función de densidad de probabilidad de los datos.** Dada la función de densidad $f(x)$ , donde $x$ es uno de los valores que puede tomar la variable aleatoria correspondiente, puede obtenerse una medida que cuantifica el grado de anomalía para la muestra $x_1$ como la inversa de $f(x_1)$ .

Aquellos valores muy poco probables tenderán a ser identificados como atípicos y, por tanto, debe emplearse una estrategia adecuada para su tratamiento, por ejemplo: eliminarlos o estimar su valor como la media de los puntos vecinos.

Otra estrategia para la identificación de valores atípicos a partir de la función $f(x)$ sería considerar como tales aquellos que estén en los extremos del dominio de $f(x)$ . Por ejemplo, a partir de $f(x)$ puede identificarse los valores $x_a$ y $x_b$ , tales que:

$$P(x \leq x_a) = P(x \geq x_b) = P_{min}$$

Si la muestra $x_1$ es menor que $x_a$ o mayor que $x_b$ , sería considerada una anomalía.

La utilización de estos procedimientos puede llevarse a cabo de forma global o local:

- En el primer caso, la función $f(x)$ corresponde al conjunto total de muestras disponibles. Es decir, todos los puntos de la serie temporal o todos los píxeles de la imagen. En este caso, podríamos detectar **anomalías de tipo puntual.**
- A fin de identificar **anomalías contextuales,** sería necesario aplicar este método en un entorno local del punto a estudiar. Para ello, la función $f(x)$ se correspondería únicamente con la vecindad del punto que se pretende evaluar como anomalía. Tal y como sucede con el filtro de mediana explicado previamente, sería necesario emplear una ventana deslizante centrada en el punto objetivo para definir dicha

vecindad.

Además, ambos métodos basados en $f(x)$ requieren fijar inicialmente un umbral de decisión: para comparar el score obtenido con el primero de los métodos y para definir el valor $P_{min}$ que identifica los valores extremos de una distribución en el caso del segundo procedimiento.

Este umbral determina la definición de anomalía en nuestro conjunto de datos y que ha de ser establecido por el usuario.

En la definición de los métodos basados en el empleo de la función de densidad de probabilidad $f(x)$ se ha asumido el conocimiento de esta. Sin embargo, esta función es desconocida, por lo que será necesario aplicar técnicas de estimación para obtener una aproximación a la misma.

Estas son algunas técnicas que pueden ser utilizadas para la estimación de esta función.

**Histograma**

Representa la técnica más sencilla. A partir de las muestras de una variable, esta se discretiza mediante la división de su dominio en un número limitado de intervalos de igual tamaño e identificados por su punto medio. Estos puntos representan los valores discretos que la variable puede tomar. Así, se obtiene la frecuencia (probabilidad) asociada a cada posible valor a partir del conjunto total de datos inicial contando el número de muestras de la variable que caen en cada intervalo.

L a **elección del número de intervalos** empleado para la discretización de la variable tiene una influencia muy significativa en la aproximación obtenida. Un número demasiado pequeño de intervalos resultará en una aproximación

excesivamente simple que no captura las particularidades de la distribución objetivo. Sin embargo, un número excesivo de intervalos conlleva que la estimación resultante presente discontinuidades (valores nulos) y cambios bruscos en su perfil.

Existen diferentes reglas para obtener el número óptimo de intervalos a considerar. Entre ellas, una de las más comunes es la **regla de Freedman-Diaconis,** que viene dada por la siguiente expresión:

$$T = 2\,IQR(x) \,/\, \sqrt[3]{n}$$

Donde:

- $x$hace referencia a la muestra disponible de la variable.
- $n$ es el número de muestras.
- $IQR$ es el rango intercuartil. Este se obtiene de la diferencia entre los percentiles 75 % y 25 %, y representa el grado de dispersión de la muestra, es decir, si la distribución está extendida en el dominio de  o si, por el contrario, está concentrada en un rango de valores.

**Estimación paramétrica**

Se asume que la función de densidad de probabilidad que caracteriza estadísticamente a la variable es de tipo normal. Por tanto, la media y varianza de esta distribución son los parámetros a obtener. Para ello, se emplean las estimaciones derivadas de la muestra disponible:

$$\mu_x = 1 \,/\, n \sum_{i=1}^{n} x_1$$

$$\sigma_x = 1 \,/\, n \sum (x_i - \mu_x)^2$$

Obviamente, la principal limitación de este método viene dada por la suposición inicial sobre la forma de la distribución. El error en la estimación será más representativo, por tanto, cuanto más difiera la distribución real de la variable del perfil normal inicialmente supuesto.

Funciones núcleo (método de Parzen)

Se trata de un procedimiento híbrido entre la estimación basada en histogramas y la estimación paramétrica. En este caso, la estimación de la función de densidad de probabilidad viene dada por la superposición de funciones núcleo centradas en cada uno de las muestras $x_i$ observadas inicialmente. La expresión de la función estimada se obtiene de la siguiente forma:

$$f\hat{}_x = 1/n \sum_{i=1}^{n} g(x - x_i, \theta),$$

Donde:

- $g(x, \theta)$ es la función núcleo.
- $\theta$ representa el conjunto de parámetros de esta función.

Comúnmente, se emplea una normal Gaussiana como función núcleo, de forma que el conjunto de parámetros θ viene dado únicamente por la varianza de la normal, dado que cada función núcleo se centra en la muestra correspondiente.

Es habitual emplear el mismo valor de varianza para el conjunto de funciones núcleo, por lo que la función de densidad de probabilidad estimada se obtendría como:

$$\widehat{f}_x = 1/n \sum_{i=1}^{n} 1/\sqrt{\left(2\pi\sigma^2\right)} exp\left[-\left(x-x_i\right)^2/2\sigma^2\right)\right]$$

Como se observa, el efecto de la varianza de las funciones núcleo normales es similar al del tamaño del intervalo para el cálculo del histograma. De hecho, el histograma puede verse como un caso particular de estimación basada en funciones núcleo, en el que estas funciones vendrían dadas por pulsos uniformes de altura igual a la unidad y centrados en el punto medio del intervalo.

Una regla extendida para la obtención de un valor adecuado de la varianza de las funciones núcleo es fijar esta al siguiente valor:

$$\sigma = 1.06\sigma_x n^{(}-1/5\Big)$$

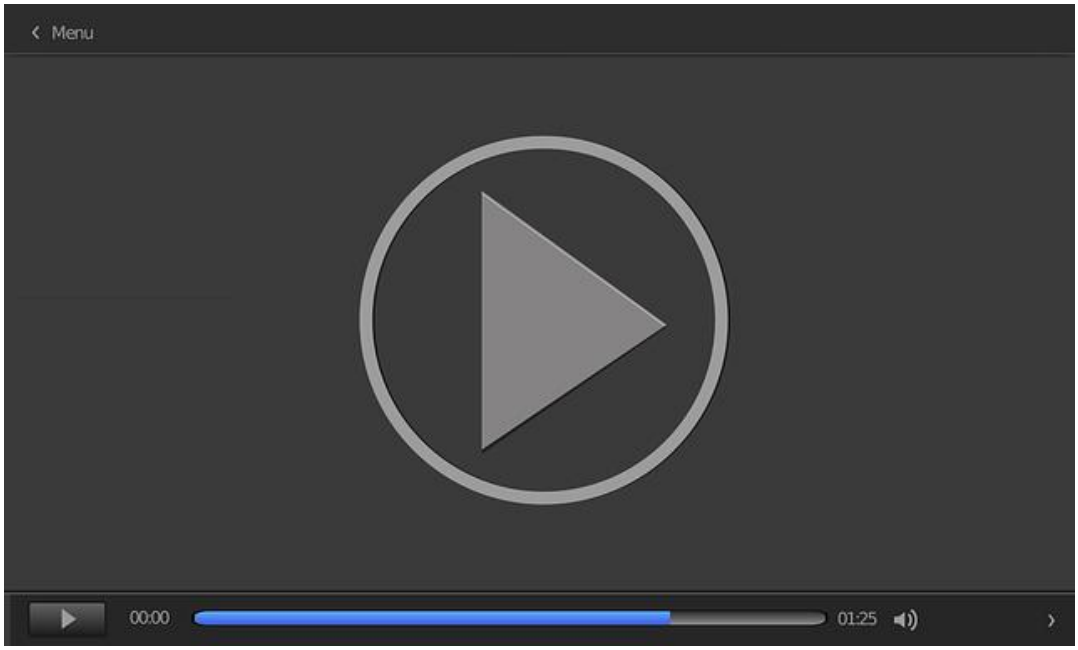
## 5.5. Referencias bibliográficas

Goldstein, M. y Uchida, S. (2016). A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data. *PLOS ONE, 11*(4): e0152173. Recuperado de https://doi.org/10.1371/journal.pone.0152173

# Machine learning

Free Engineeering Lectures. (2014, agosto 22). *Anomaly Detection vs Supervised Learning | Lecture - 73 | Machine Learning* [Vídeo]. YouTube. https://www.youtube.com/watch?v=ROX1VCnVZeQ

Vídeo explicativo de los procesos de detección de anomalías basados en técnicas como el *machine learning.*



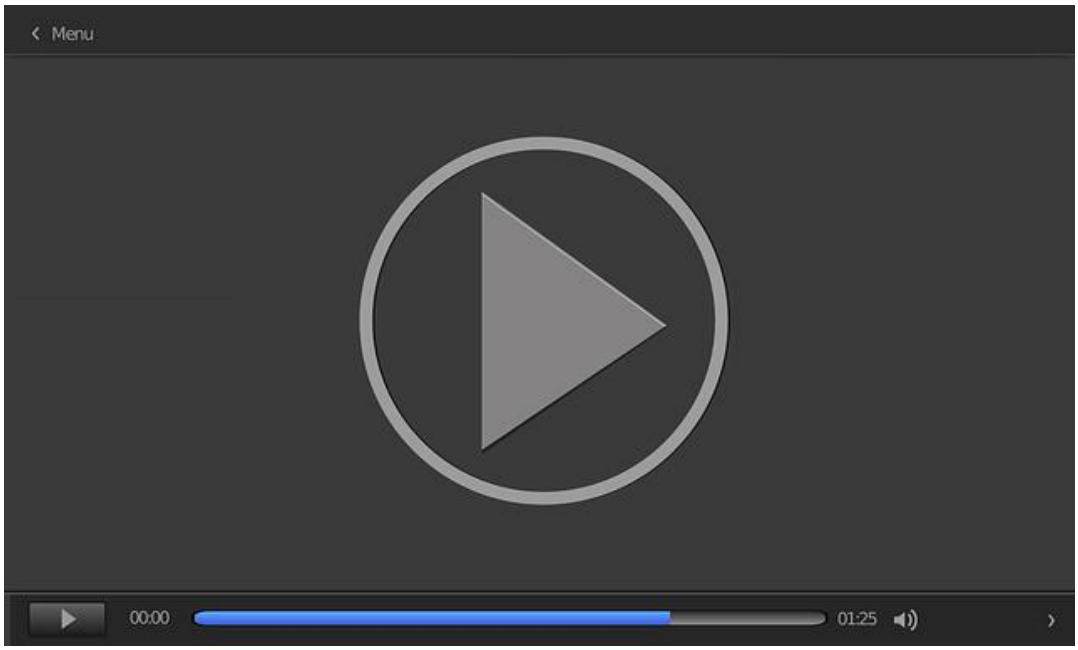

Accede al vídeo:

https://www.youtube.com/embed/ROX1VCnVZeQ

# Machine Learning for Real-Time Anomaly Detection in Network Time-Series Data

Rise SICS (2016, diciembre 1). *Machine Learning for Real-Time Anomaly Detection in Network Time-Series Data - Jaeseong Jeong* [Vídeo]. YouTube. https://www.youtube.com/watch?v=0PqzukqMcdA

Conferencia de Jaeseong Jeong sobre los procesos de detección de anomalías basados en técnicas como el *machine learning* para detección de anomalías a tiempo real.




Accede al vídeo:

https://www.youtube.com/embed/0PqzukqMcdA

## Bibliografía

Akoglu, L., Tong, H. y Koutra, D. (2015). Graph based anomaly detection and description: a survey. *Data Mining and Knowledge Discovery, 29*(3), 626-688. Recuperado de https://scholar.google.com/citations?user=4ITkr_kAAAAJ

Chandola, V., Banerjee, A. y Kumar, V. (2009). Anomaly detection: A survey. *ACM computing surveys (CSUR), 41*(3), 15. Recuperado de https://dl.acm.org/citation.cfm?id=1541882

Chandola, V., Banerjee, A. y Kumar, V. (2012). Anomaly detection for discrete sequences: A survey. *IEEE Transactions on Knowledge and Data Engineering*, *24*(5), 823-839. Recuperado de http://cs.brown.edu/courses/cs227/papers/anomaly-survey-TKDE.pdf

# Test

1. Los términos *outlier* y anomalía:

   A. Se emplean indistintamente.

   B. Son conceptos diferentes.

   C. Solo pueden encontrarse en señales unidimiensionales.

   D. Solo pueden encontrarse en imágenes.

2. ¿Qué es una anomalía?

   A. Un punto de una señal.

   B. Un píxel de una imagen que toma un valor fuera de rango.

   C. Un patrón inusual que no se ajusta al comportamiento esperado.

   D. Un conjunto de puntos que describen un perfil de variación muy significativa.

3. ¿Qué es una anomalía puntual?

   A. Una muestra anómala en un contexto específico (pero no de otro modo).

   B. Una muestra individual notablemente diferente respecto al resto.

   C. Cualquier valor de la señal superior a la varianza de esta.

   D. Colección de instancias relacionadas que reflejan un comportamiento inusual con respecto a todo el conjunto de datos.

4. ¿Qué es una anomalía contextual?

   A. Una muestra anómala en un contexto específico (pero no de otro modo).

   B. Una muestra individual notablemente diferente respecto al resto.

   C. Cualquier valor de la señal superior a la varianza de esta.

   D. Colección de instancias relacionadas que reflejan un comportamiento inusual con respecto a todo el conjunto de datos.

**5.** ¿Qué es una anomalía colectiva?

A. Una muestra anómala en un contexto específico (pero no de otro modo).

B. Una muestra individual notablemente diferente respecto al resto.

C. Cualquier valor de la señal superior a la varianza de esta.

D. Colección de instancias relacionadas que reflejan un comportamiento inusual con respecto a todo el conjunto de datos.

**6.** Los métodos supervisados para la detección de anomalías:

A. Todas las respuestas son correctas.

B. Se derivan de clasificadores de una clase.

C. Son capaces de detectar muestras de datos anómalas sin necesidad de un proceso de entrenamiento previo.

D. Parten de un conjunto de ejemplos etiquetados como anomalía o no para el entrenamiento de un clasificador; este es un procedimiento habitual en identificación de pagos fraudulentos con tarjeta de crédito.

**7.** Los métodos no supervisados para la detección de anomalías:

A. Permiten entrenar clasificadores capaces de detectar eventos anómalos nunca antes identificados.

B. Son los más comunes.

C. Son los más flexibles.

D. Son flexibles y suelen basarse en el cálculo de distancias para detectar muestras anómalas en nuestras señales o conjuntos de datos.

**8.** ¿Qué procedimiento suele emplearse comúnmente, tanto en señales temporales como en imágenes, para la eliminación del ruido impulsivo?

A. Filtro paso-alto.

B. *Clustering* basado en *k-means.*

C. Filtro de mediana.

D. Clasificadores de una clase.

**9.** Si se aplica un filtro de mediana de longitud 5 al punto marcado en negrita de la siguiente serie, ¿cuál será el valor de salida? Serie: [2 1 3 4 5 2 3 **12** 1 4 5 4 5]

A. 3.

B. 4.

C. 12.

D. 2.

**10.** Si se aplica un filtro de mediana de longitud 7 al punto marcado en negrita de la siguiente serie, ¿cuál será el valor de salida? Serie: [2 1 3 4 5 2 3 **12** 1 4 5 4 5]

A. 3.

B. 4.

C. 12.

D. 2.