Machine Learning: El nuevo aliado de la seguridad de la información

Índice

Presentación

- **Santiago Hernández Ramos**
  - CyberSoc Deloitte
  - Doctorando Seguridad de la información
  - ToorCon SanDiego, BlackHat Europa, Navaja Negra, CyberCamp, Noconname, CCN-CERT…

Estado actual

Estado actual Ciberseguridad

- **Crecimiento exponencial del campo de la ciberseguridad, pero las amenazas se mantienen inmutables [1]**

- **La economía detrás de los ciberataques**

"There is strong evidence that the best researchers are now motivated more by monetary gain than prestige" [2]

[1] https://www.enisa.europa.eu/topics/threat-risk-management/threats-and-trends/enisa-threat-landscape/etl2015/enisa-threat-taxonomy-a-tool-for-structuring-threat-information
[2] Miller, C. (2007). The legitimate vulnerability market: Inside the secretive world of 0-day exploit sales. In In Sixth Workshop on the Economics of Information Security.

Estado actual Inteligencia Artificial

- ***"AI is quickly transforming American life and American business, improving how we diagnose and treat illnesses, grow our food, manufacture and deliver new products, manage our finances, power our homes, and travel from point A to point B"*** *[1]*

- ***"Experts forecast that rapid progress in the field of specialized artificial intelligence will continue. Although it is very unlikely that machines will exhibit broadly-applicable intelligence comparable to or exceeding that of humans in the next 20 years, it is to be expected that machines will reach and exceed human performance on more and more tasks"*** *[2]*

[1] https://www.whitehouse.gov/wp-content/uploads/2018/05/Summary-Report-of-White-House-AI-Summit.pdf
[2] https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

Inteligencia Artificial y Ciberseguridad

"Today's AI has important applications in cybersecurity, and is expected to play an increasing role for both defensive and offensive cyber measures." [1]

[1] https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

Inteligencia Artificial y *Machine Learning*

- **En los últimos años, la mayoría de las aplicaciones exitosas de la IA al mundo real provienen de una de sus disciplinas conocía como *Machine Learning***

- **No todos los problemas en IA pueden solucionarse con *Machine Learning*, solo aquellos que pueden solucionarse con datos**

¿Qué es el Machine Learning?

¿Qué es el Machine Learning?

- ***"El aprendizaje automático es un subdominio de la IA que proporciona a los sistemas la capacidad de aprender y mejorar automáticamente a partir de la experiencia sin ser explícitamente programados para ello. Se basa en la hipótesis subyacente de crear el modelo y trata de mejorarlo ajustando más datos en el modelo a lo largo del tiempo." [1]***

[1] La definición de "sin estar explícitamente programado" a menudo se atribuye a Arthur Samuel, quien acuñó el término "aprendizaje automático" en 1959. Pero la frase no se encuentra literalmente en esta publicación, y puede ser una paráfrasis que apareció más adelante. Conferir "Parafraseando a Arthur Samuel (1959), la pregunta es: ¿cómo pueden las computadoras aprender a resolver problemas sin estar explícitamente programadas?" en Koza, John R.; Bennett, Forrest H .; Andre, David; Keane, Martin A. (1996). Diseño automatizado de la topología y el dimensionamiento de los circuitos eléctricos analógicos mediante programación genética. Inteligencia Artificial en Diseño '96. Springer, Dordrecht. pp. 151-170. doi: 10.1007 / 978-94-009-0279-4_9.

Aprendizaje Supervisado

- ***"El aprendizaje supervisado es la tarea de aprendizaje automático que consiste en aprender una función que mapea una entrada a una salida basada en pares de entrada-salida de ejemplo. Se deduce una función de datos de entrenamiento etiquetados"*** [1]

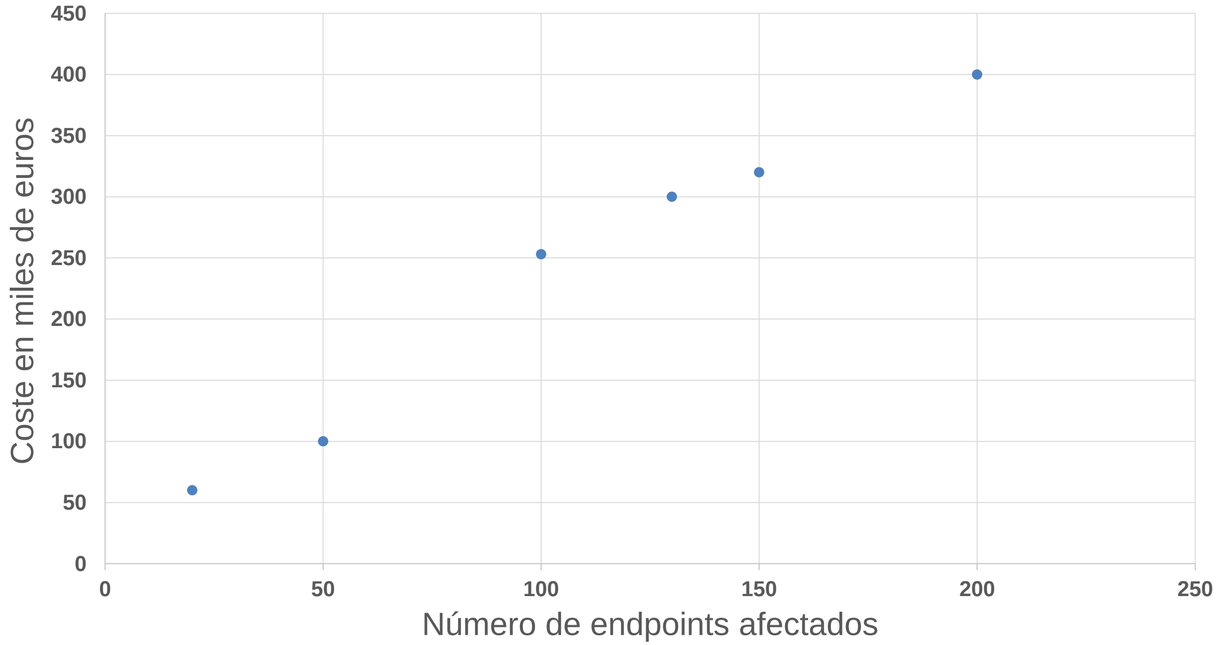
- **Clasificación: Intenta predecir valores discretos**

- **Regresión: Intenta predecir valores continuos**

[1] Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar (2012) Foundations of Machine Learning, The MIT Press ISBN 9780262018258.

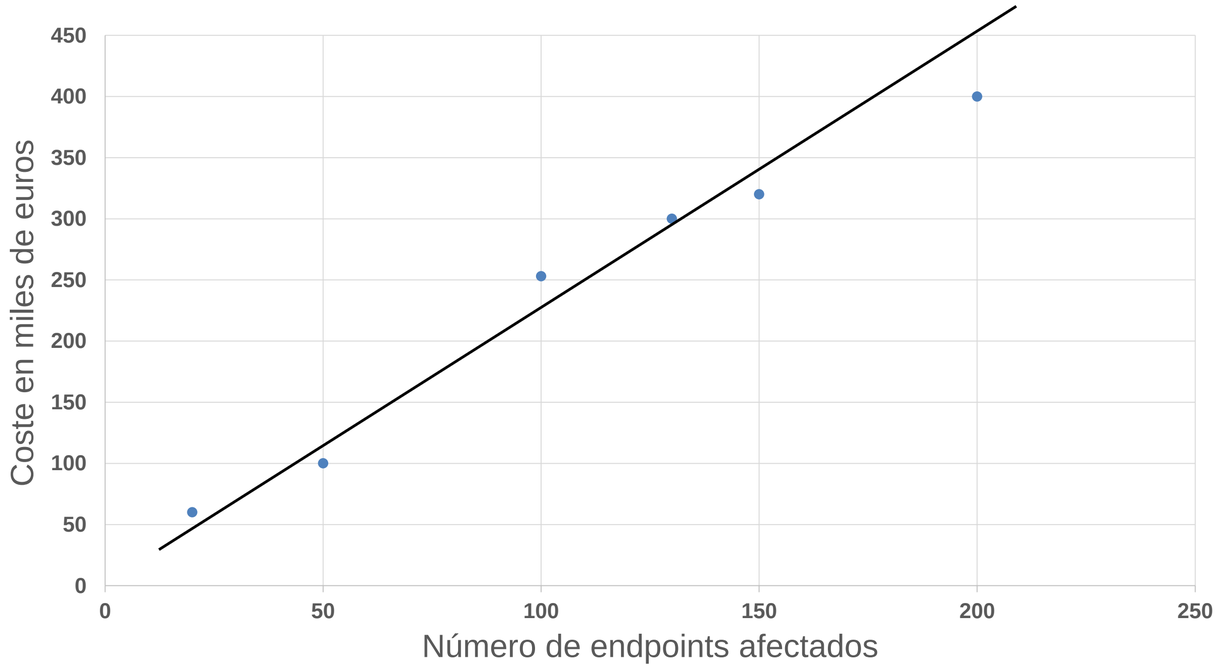
Aprendizaje Supervisado: Regresión

- **Predecir el coste en euros de gestionar un incidente de seguridad**



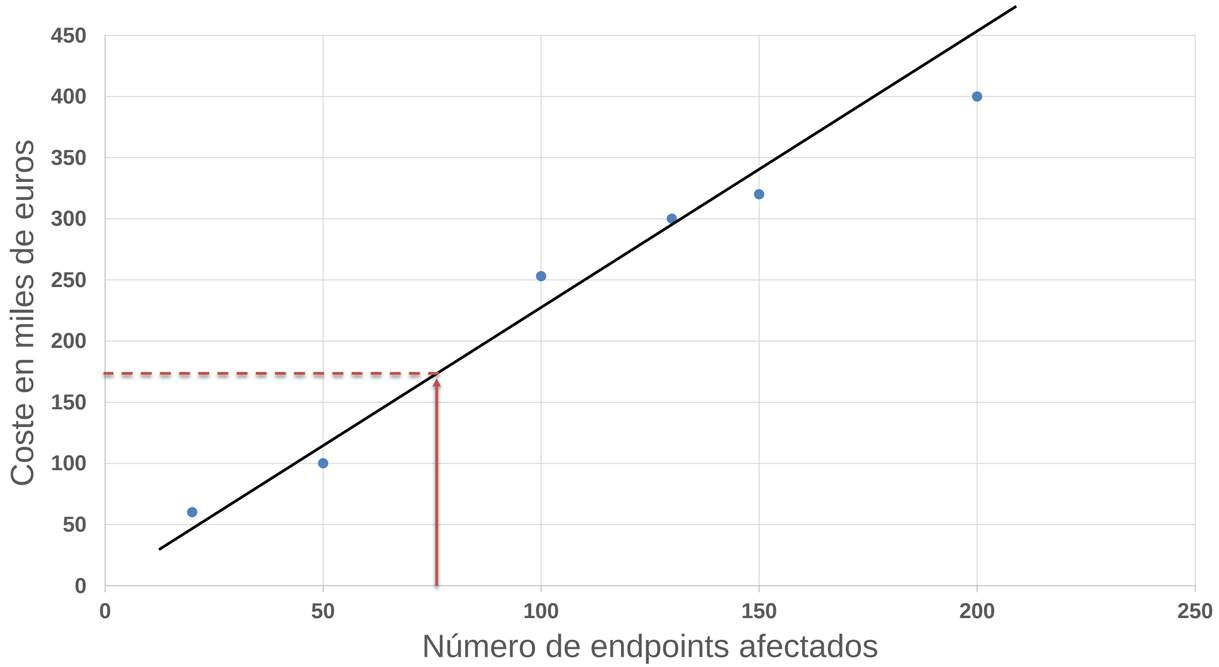
Aprendizaje Supervisado: Regresión

- **Predecir el coste en euros de gestionar un incidente de seguridad**



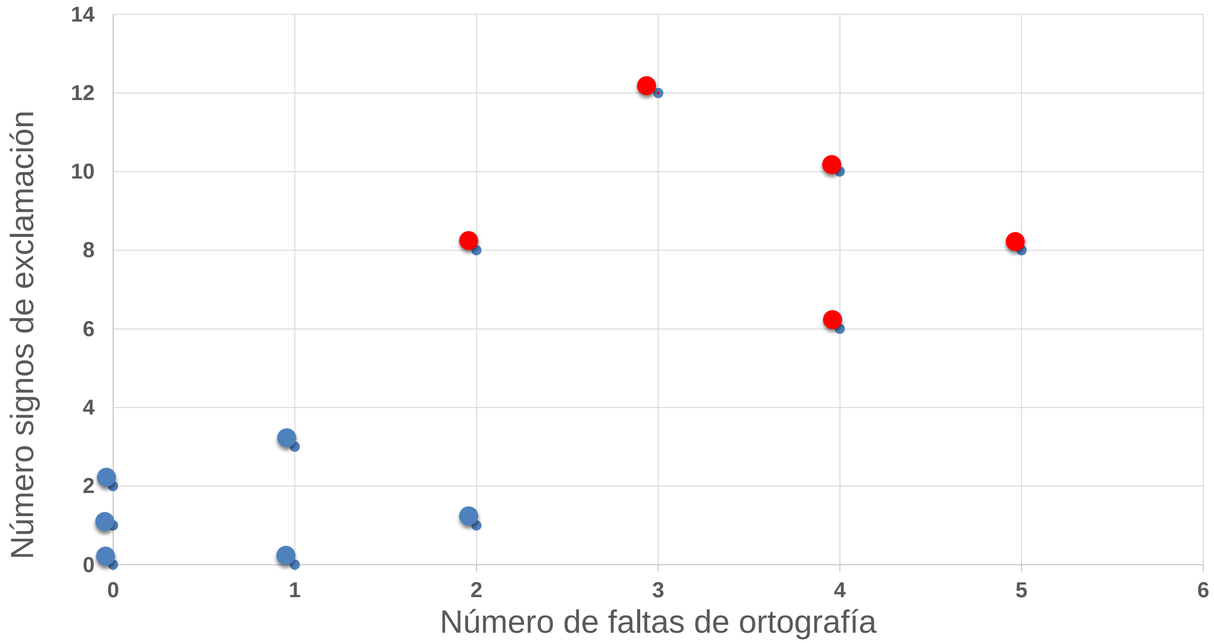
Aprendizaje Supervisado: Regresión

- **Predecir el coste en euros de gestionar un incidente de seguridad**



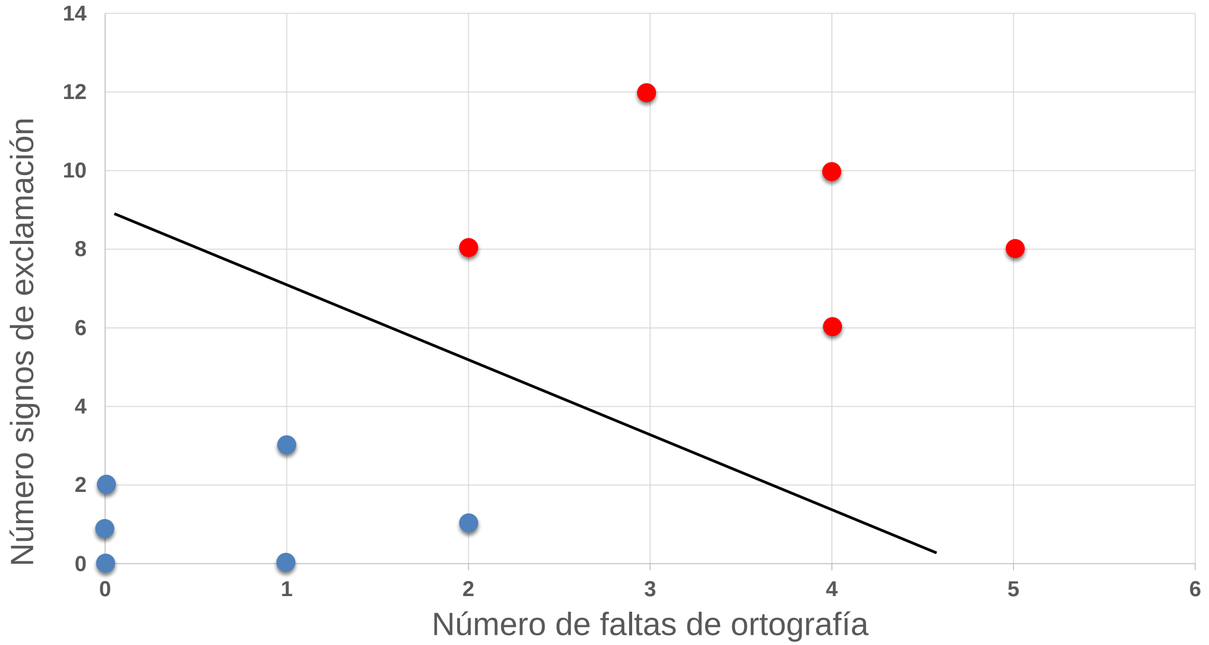
Aprendizaje Supervisado: Clasificación

- **Predecir si un correo electrónico es SPAM**



Aprendizaje Supervisado: Clasificación

- **Predecir si un correo electrónico es SPAM**

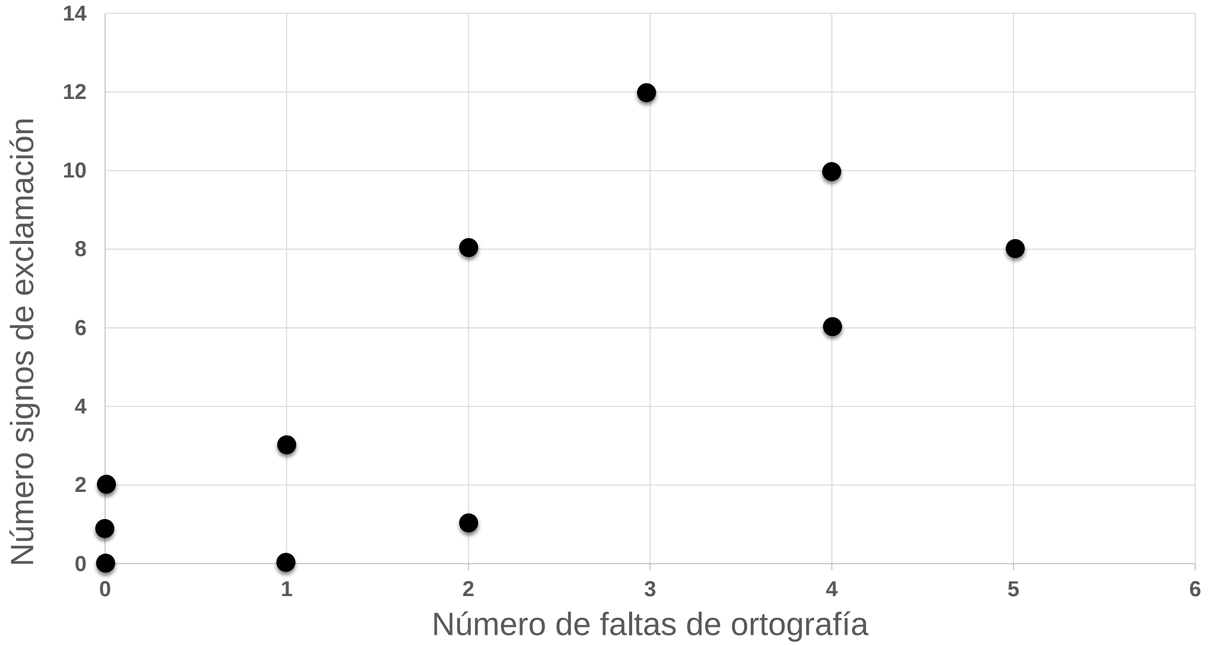


Aprendizaje No Supervisado

- ***"El aprendizaje automático no supervisado es la tarea de aprendizaje automático que consiste en inferir una función que describe la estructura de de un conjunto de datos sin etiquetar (es decir, datos que no se han clasificado ni categorizado)."***

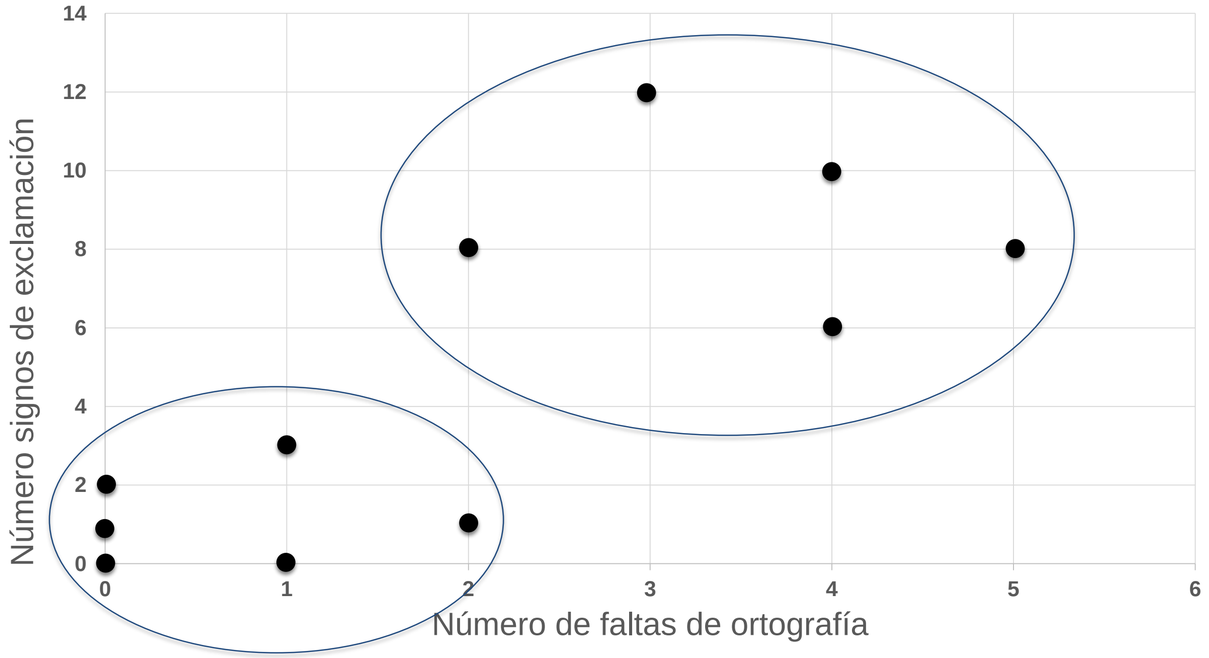
Aprendizaje Supervisado: Clasificación

- **Predecir si un correo electrónico es SPAM**



Aprendizaje Supervisado: Clasificación

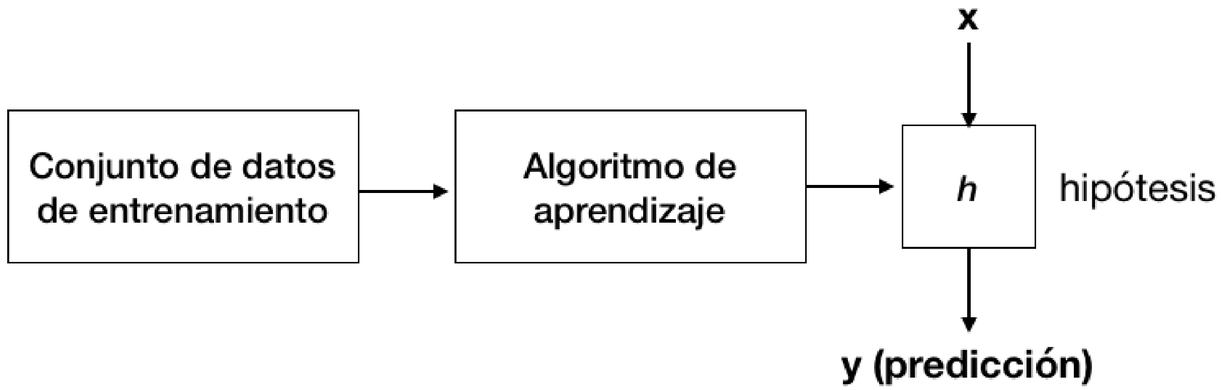
- **Predecir si un correo electrónico es SPAM**



Regresión Lineal
¿Cómo aprende un algoritmo?

¿Cómo aprende un algoritmo?

- ***"El aprendizaje supervisado es la tarea de aprendizaje automático que consiste en aprender una función que mapea una entrada a una salida basada en pares de entrada-salida de ejemplo"*** [1]

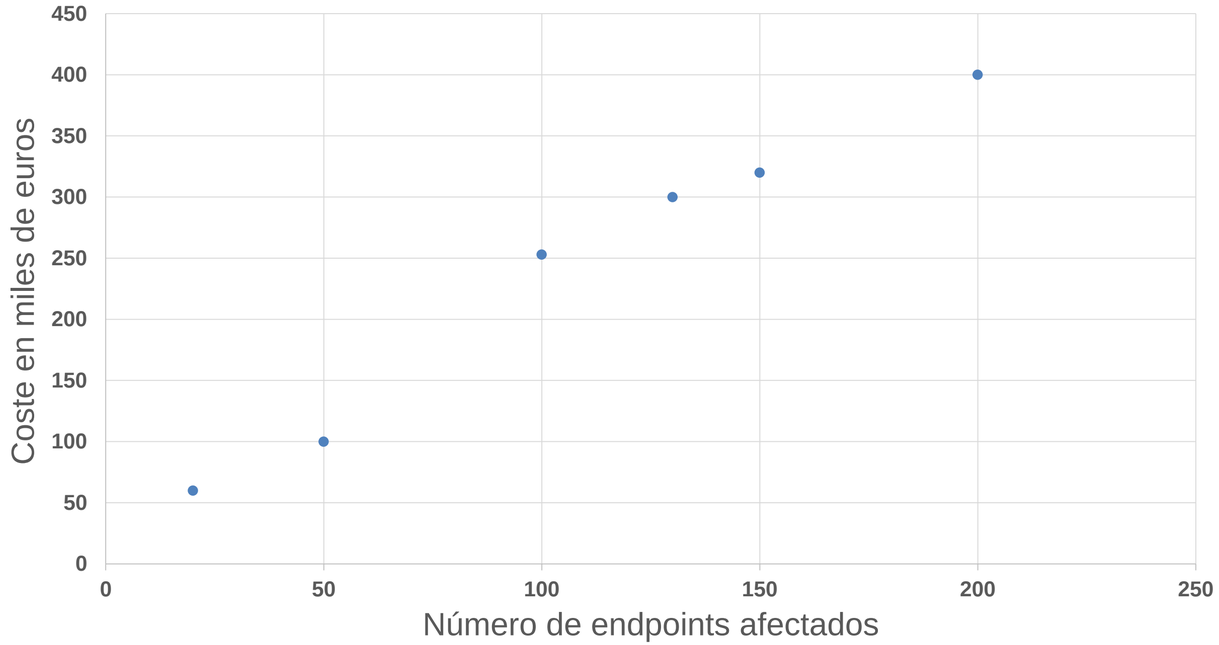


[1] Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar (2012) Foundations of Machine Learning, The MIT Press ISBN 9780262018258.

Aprendizaje Supervisado: Regresión

- **Predecir el coste en euros de gestionar un incidente de seguridad**

Endpoints afectados	Coste en miles de euros
50	100
100	253
130	300
20	60
200	400
150	320



Aprendizaje Supervisado: Regresión

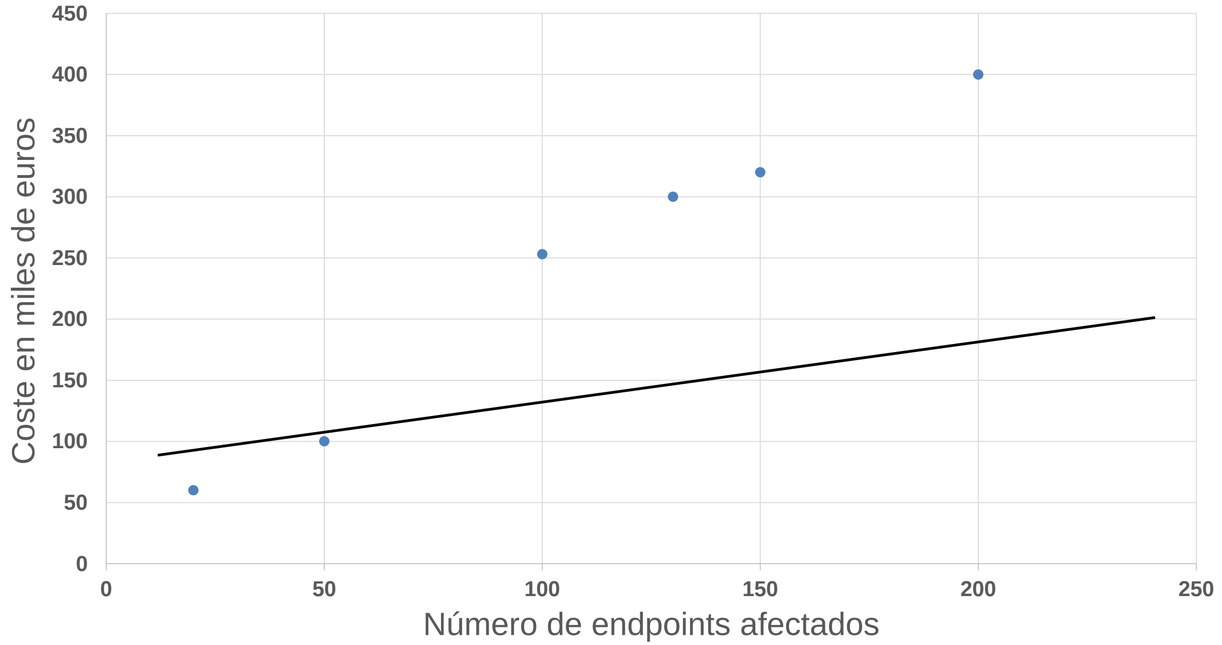
- **Predecir el coste en euros de gestionar un incidente de seguridad**

$$y = m\,x + b$$

Aprendizaje Supervisado: Regresión

- **Predecir el coste en euros de gestionar un incidente de seguridad**

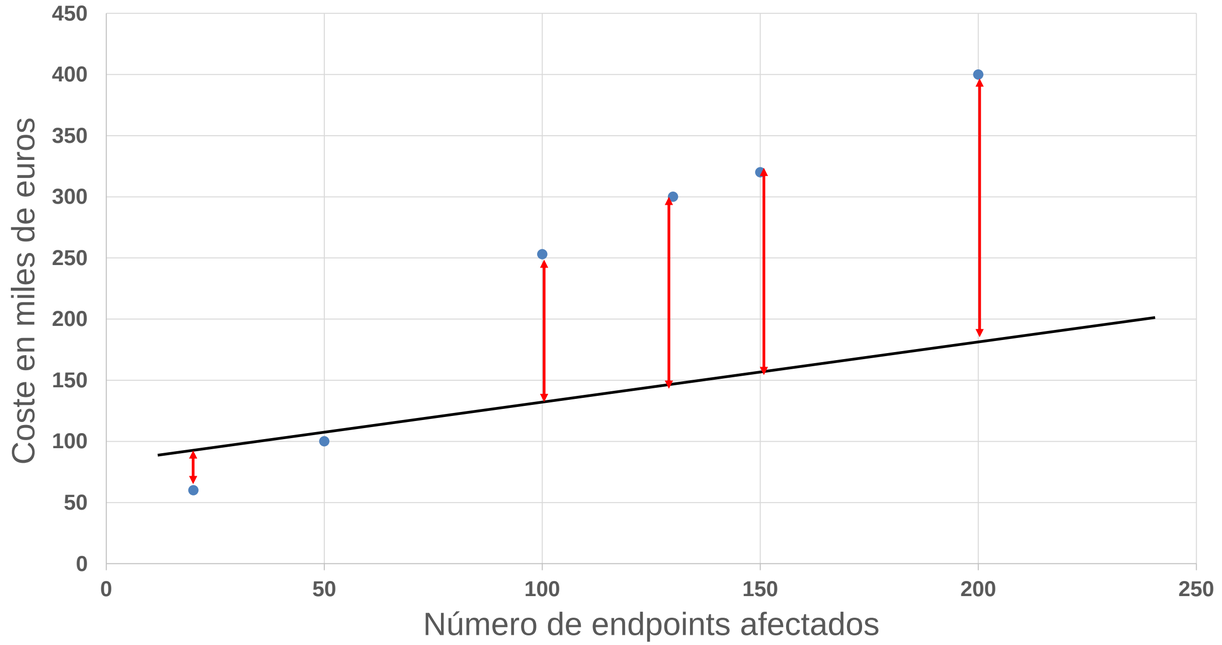
$$y = 0.5x + 85$$



Aprendizaje Supervisado: Regresión

- **Predecir el coste en euros de gestionar un incidente de seguridad**

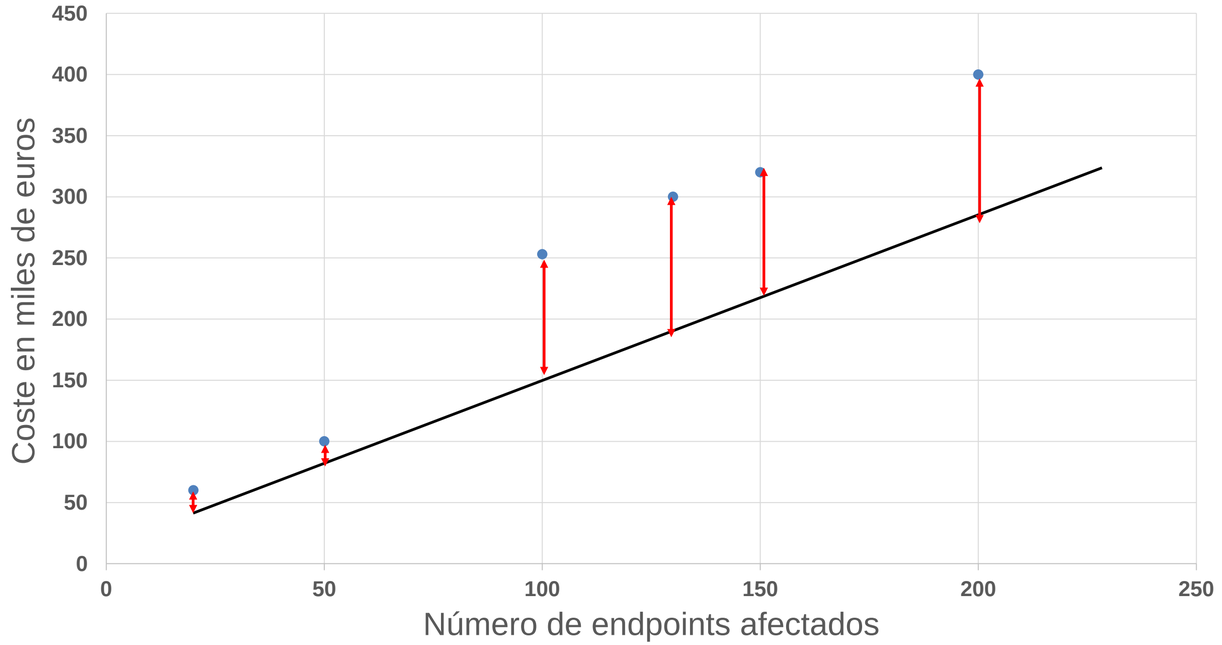
$$y = 0.5x + 85$$



Aprendizaje Supervisado: Regresión

- **Predecir el coste en euros de gestionar un incidente de seguridad**

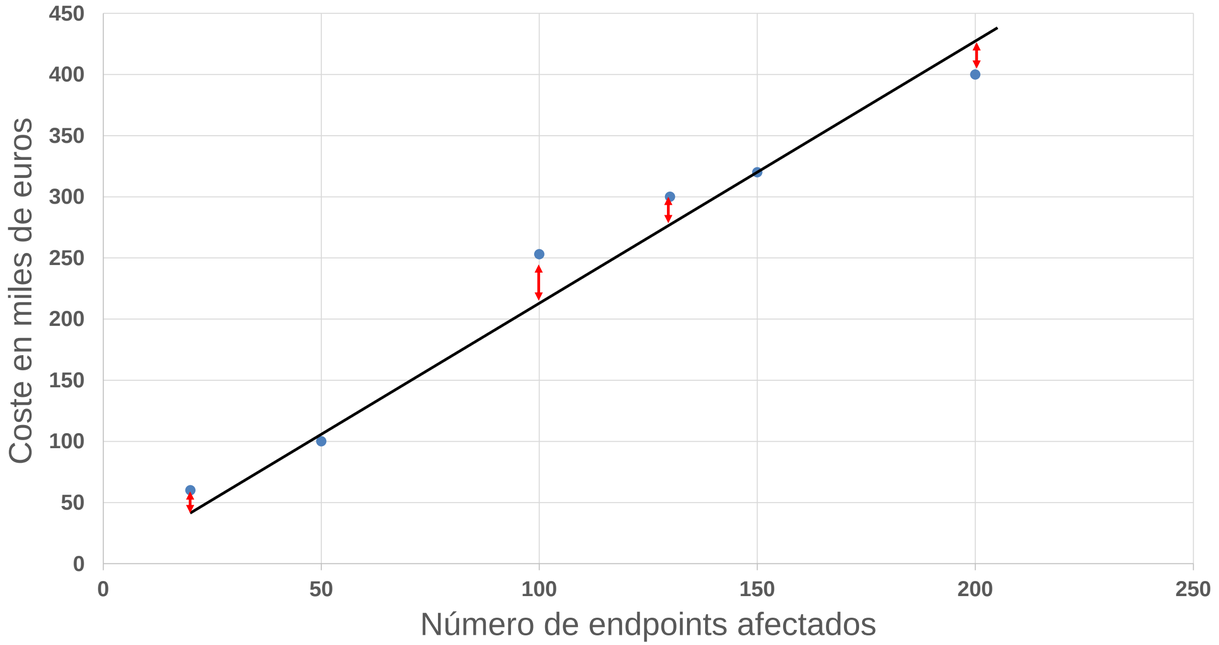
$$y = 1.4x + 10$$



Aprendizaje Supervisado: Regresión

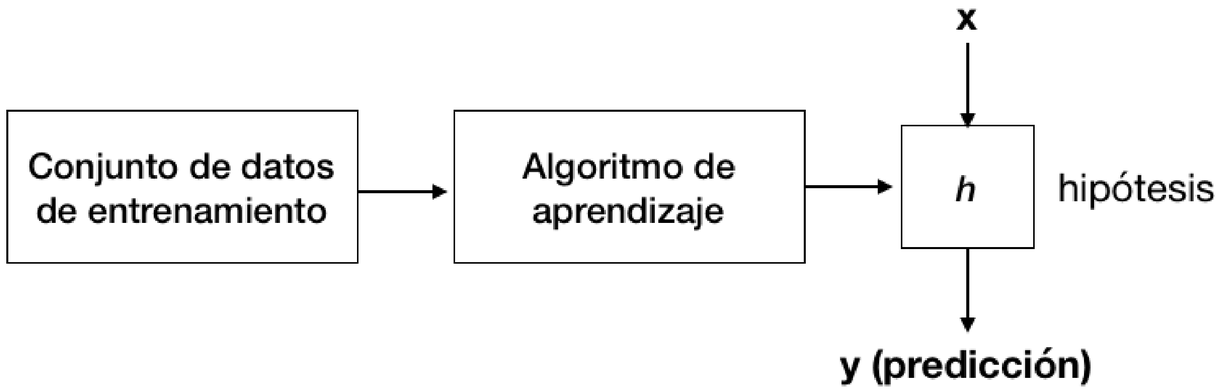
- **Predecir el coste en euros de gestionar un incidente de seguridad**

$$y = 2.1x + 0$$



¿Cómo aprende un algoritmo?

- ***"El aprendizaje supervisado es la tarea de aprendizaje automático que consiste en aprender una función que mapea una entrada a una salida basada en pares de entrada-salida de ejemplo"*** [1]

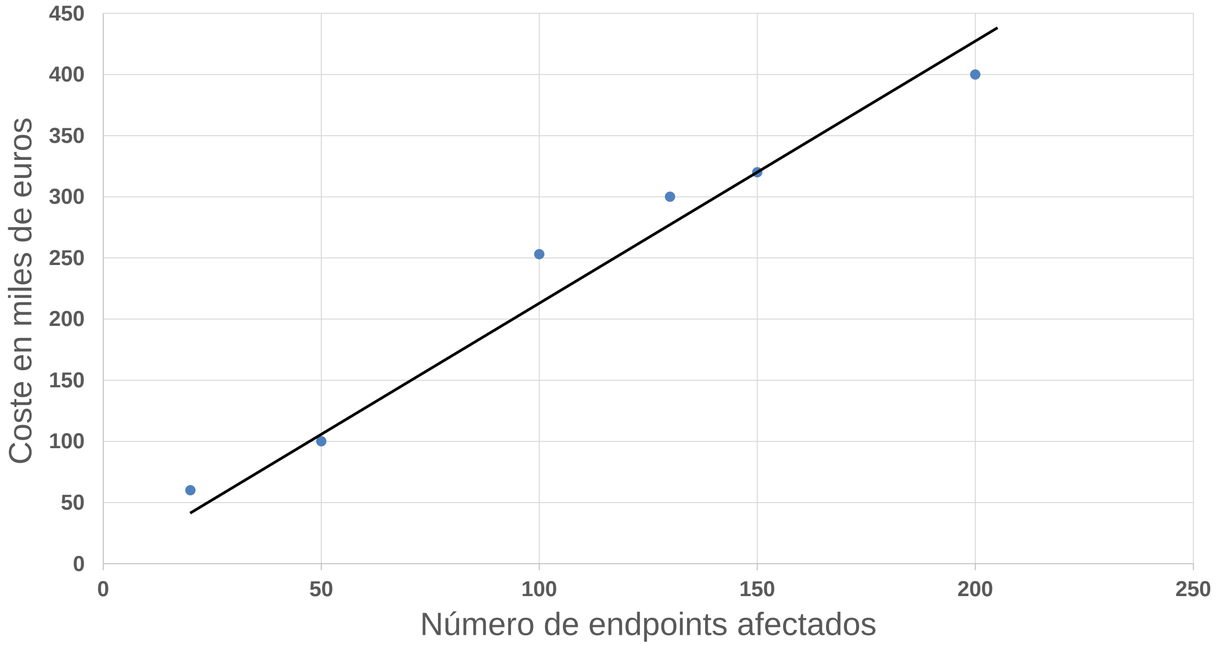


[1] Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar (2012) Foundations of Machine Learning, The MIT Press ISBN 9780262018258.

Aprendizaje Supervisado: Regresión

- **Predecir el coste en euros de gestionar un incidente de seguridad**

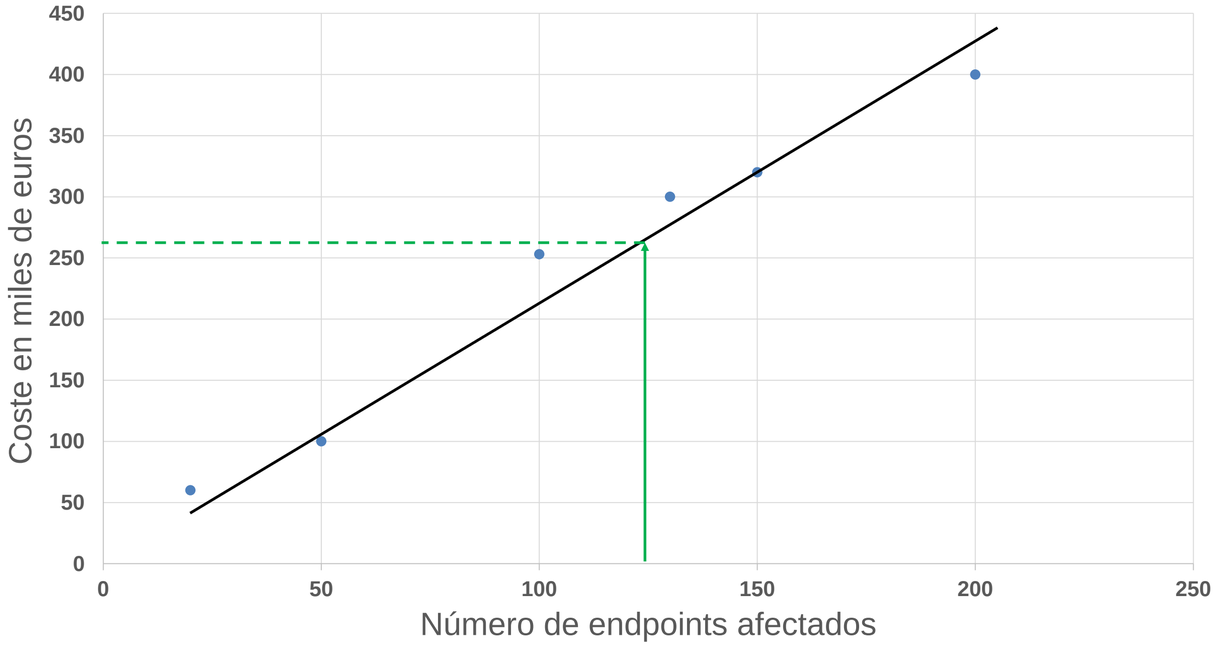
$$h(x) = 2.1x + 0$$



Aprendizaje Supervisado: Regresión

- **Predecir el coste en euros de gestionar un incidente de seguridad**

$$h(x) = 2.1x + 0$$



Clasificación

Clasificación y Ciberseguridad

- **Un profesional de la ciberseguridad se encuentra constantemente realizando procesos de clasificación:**
  - Para cada correo electrónico, ¿el correo es phishing?
  - Para un fichero enviado a través de la red, ¿el fichero contiene malware?
  - Para cada petición hacia el exterior, ¿la petición es una llamada a un CC?
  - Para cada transacción, ¿se trata de una transacción fraudulenta?
  - Para cada petición entrante a la red, ¿la petición forma parte de un DOS?
- **El conjunto de sucesos anteriores solo es una pequeña parte de todos los que se deben revisar de manera constante para mantener un nivel de seguridad adecuado**

Clasificación y Ciberseguridad

- **Las organizaciones incorporan varios procesos de recolección de datos, como logs o eventos**

- **El procesamiento de estos datos para la búsqueda de amenazas de seguridad puede realizarse de diferentes maneras que no necesariamente son Machine Learning**

Clasificación y Ciberseguridad

- **<u>Ejemplo</u>: Se desean identificar los ataques de fuerza bruta sobre los paneles de autenticación de un conjunto de aplicaciones web de una organización**
  - Se comprueban los logs del servidor de aplicación con ataques de fuerza bruta que ocurrieron en el pasado
  - El analista descubre un patrón en los datos, en todos los ataques pasados se han realizado más de 10 peticiones desde la misma IP en menos de 1 minuto
  - Se programa un algoritmo que implemente la siguiente heurística, si se realizan más de 10 peticiones desde la misma IP en menos de un minuto, se bloquea la dirección
  - Este algoritmo obtenido a través de la interpretación de los datos sería capaz de bloquear algunos de los ataques de fuerza bruta, pero, **¿por qué en la heurística se ha elegido 10 peticiones y 1 minuto como límite?, ¿por qué no elegir 11 peticiones y 2 minutos?¿Cada cuanto tiempo se modifican estos parámetros?**

Clasificación y Ciberseguridad

- **En el ejemplo anterior, probablemente el analista haya utilizado algún tipo de principio para determinar que “10” era el número de peticiones óptimas y “1” el tiempo óptimo**

- ***Machine Learning* consiste en utilizar algoritmos que procesan el histórico de datos e infieren las reglas de clasificación que se consideran óptimas de acuerdo con un conjunto de principios matemáticos**

Caso práctico: Clasificación de SPAM

Demostración

Clustering

Clústering y ciberseguridad

- **<u>Principio</u>: La actividad maliciosa en muchas ocasiones ocurre en conjuntos**
  - Si un atacante trata de entrar en tu red, probablemente tenga que realizar varios escáneres que generen tráfico con características muy similares
  - Si un atacante trata de explotar un SQL injection, probablemente tenga que probar varias secuencias similares hasta encontrar la adecuada
  - Si un atacante realiza un ataque de fuerza bruta sobre un panel de autenticación, probablemente probará con un conjunto de usuarios y contraseñas
  - Si un atacante realiza un ataque de phishing a una organización, probablemente enviará correos de phishing muy similares a varios empleados

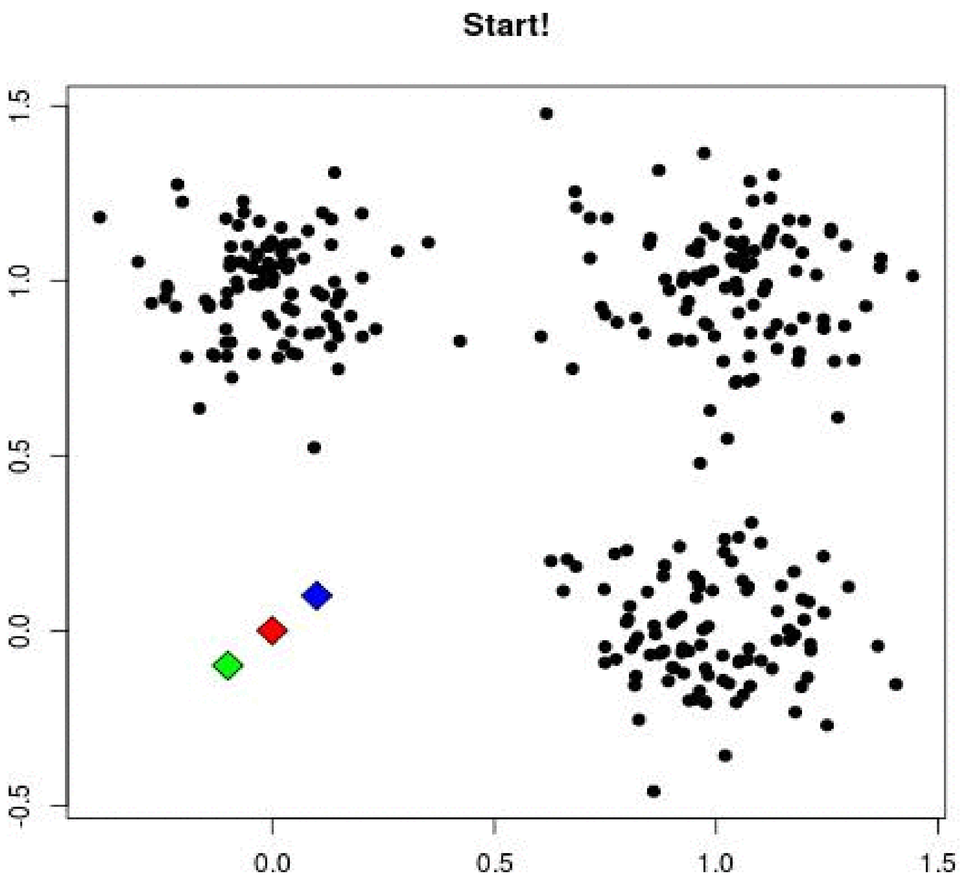
Clústering

- **Aprendizaje no supervisado**
- **El objetivo es agrupar de manera coherente un conjunto de datos sin etiquetar en subconjuntos o clústers**
- **Agrupación de los datos mediante el concepto de proximidad entre ellos**
- **Métrica: método concreto con el que se evalúa la cercanía entre los puntos**
- **Ejemplo rudimentario de clústering:** Escoger una o varias dimensiones y definir cada cluster como el conjunto de elementos que comparten valores en esas dimensiones. Ejemplo: Si eliges la dirección IP, se define un cluster por cada dirección IP (GROUP BY de SQL)

Clústering basado en distancias: KMEANS

- **Algoritmo de clustering más utilizado y popular**
- **La métrica que utiliza para medir la distancia es la distancia Euclídea**
- **Escala muy bien a grandes conjuntos de datos**
- **Consideraciones:**
  - El analista debe intuir por adelantado el número de clusters. Estrategia común: uno y tres veces el número de etiquetas existentes
  - Hay que utilizar normalización
  - No utilizar KMeans con datos categóricos a los que se le aplica one-hot encoding. Codificarse como multiple binary.
  - KMeans pierde eficiencia en conjuntos de datos con muchas dimensiones. Utilizar PCA o SVD
  - KMeans funciona mejor si los centroides iniciales se eligen aleatoriamente.
  - KMeans asume que los clusters son esféricos. **No funciona correctamente en distribuciones de datos no esféricas.**

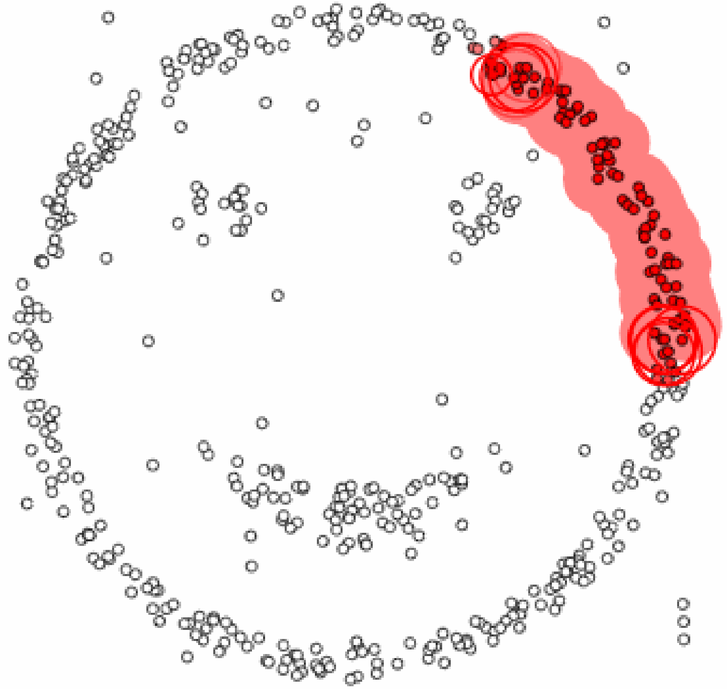
KMEANS



Clústering basado en densidades

- **DBSCAN es un algoritmo de clustering basado en densidad, no en distancia**
- **No requiere elegir el número de clusters, los infiere del conjunto de datos**
- **Funciona correctamente con distribuciones no esféricas**
- **Consideraciones:**
  - No funciona correctamente cuando los clusters del conjunto de datos tienen diferentes densidades
  - La selección de los parámetros epsilon y minPoints determina el correcto funcionamiento del algoritmo
  - No se comporta adecuadamente en conjuntos de datos con muchas dimensiones
  - **¡Hay que tener cuidado con no alterar las densidades del conjunto de datos si se divide en subconjuntos!**

DBSCAN



epsilon = 1.00
minPoints = 4

Restart

Pause

Caso Práctico: Threat Hunting

- **What are you looking for? (Hypothesis)**
  - Se ha producido una campaña de phising en mi empresa. Los usuarios pueden estar accediendo a distintas urls y descargando malware.
- **Investigation (Data)**
  - Urls accedidas por todos los usuarios durante las últimas semanas
- **Uncover Patterns and IOCs (Techniques)**
- **Inform and Enrich Analytics (Takeaways)**

https://sqrrl.com/media/Your-Practical-Guide-to-Threat-Hunting.pdf

Detección de Anomalías

Detección de Intrusiones

- **La detección de intrusiones se cataloga principalmente en dos categorías [1]:**

  - **Basado en reglas y heurísticas**: Genera un numero reducido de falsos positivos. Detecta ataques conocidos. No funciona correctamente para la detección de nuevos ataques

  - **Basado en anomalías**: Perfila el comportamiento normal del sistema. Es capaz de detectar ataques nuevos. Puede generar un numero mayor de falsos positivos.

[1] Lee, W., and Stolfo, S. J. Data mining approaches for intru- sion detection. In Proc. of the 7th USENIX Security Symposium (USA, 1998), vol. 7, USENIX Association, pp. 79–94.

Detección de Anomalías

- **Una anomalía es un evento que se desvía del comportamiento normal o esperado y es sospechoso, en este caso, desde una perspectiva de seguridad.**
- **La detección de anomalías es la identificación de elementos raros, eventos u observaciones que levantan sospechas al diferir significativamente de la mayoría de los datos [1]**
- **La detección de anomalías fue propuesta para sistemas de detección de intrusos (IDS) por Dorothy Denning en 1986 [2]**
- **Las anomalías pueden producirse debido a dos factores principales [3]:**
  - Relacionadas con el rendimiento
  - Relacionadas con la seguridad

[1] https://en.wikipedia.org/wiki/Anomaly_detection
[2] D. E. Denning, P. G. Neumann, Requirements and model for IDES—A real-time intrusion detection system, 1985.
[3] Thottan, M., and Ji, C. Anomaly detection in IP networks. IEEE Transactions on Signal Processing 51, 8 (August 2003), 2191–2204.

Machine Learning y Detección de Anomalías

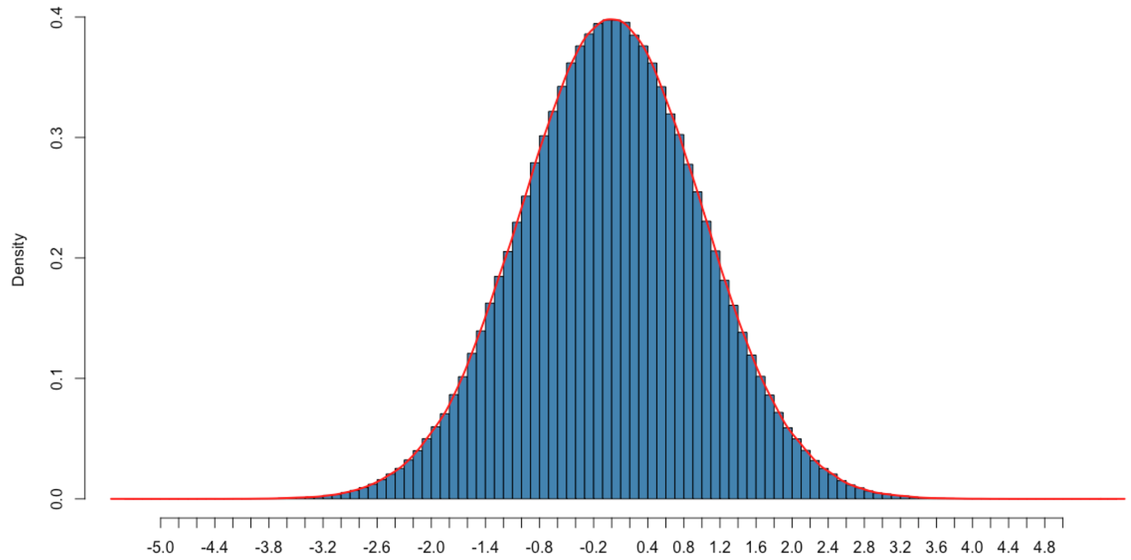
- **“A machine learning algorithm attempts to recognize complex patterns in existing datasets to help make intelligent decisions or predictions when it encounters new or meaningful patterns from a dataset, usually large, in a domain by using a nontrivial learning mechanism”** [1]

- **Aprendizaje semi-supervisado:** pueden usar ejemplos etiquetados de un número limitado de categorías. Por ejemplo, si hay dos etiquetas posibles, sí y no, podemos tener ejemplos etiquetados de la clase sí solamente.

[1] Bhattacharyya, D. K., & Kalita, J. K. (2013). Network anomaly detection: A machine learning perspective. Chapman and Hall/CRC.

Distribución Gaussiana

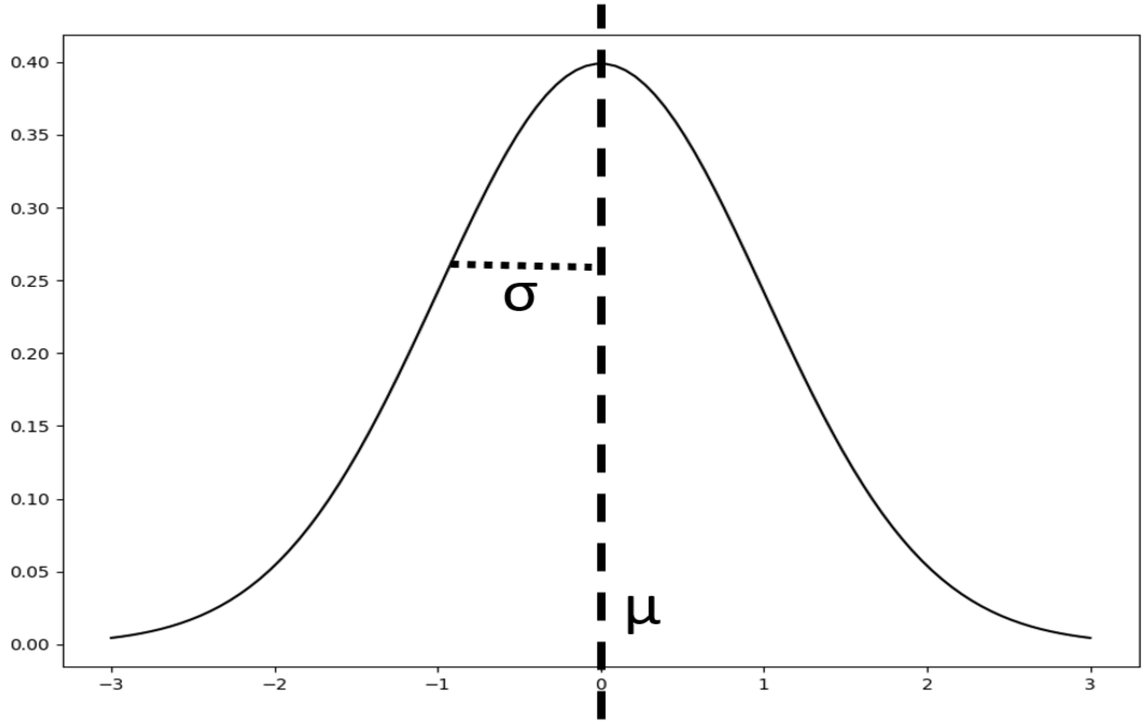
- **Es una de las distribuciones de probabilidad de variable continua que con más frecuencia aparece en estadística y en la teoría de probabilidades.**



Distribución Gaussiana

- **Parámetros del modelo**

$$\phi_{\mu,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$



Distribución Gaussiana: Algoritmo de detección

1. **Seleccionar características que pueden determinar que un ejemplo sea anómalo.**

2. **Ajustamos los parámetros del modelo.**

3. **Dado un nuevo ejemplo, computamos la probabilidad p(x)**

4. **Si p(x) < epsilon*, lo consideramos una anomalía**

*** Límite (probabilidad) a partir del cuál consideramos un ejemplo anómalo**

Caso Práctico: Detección de fraude bancario

Cliente paga con tarjeta

Solicitud de autorización de pago

Autorización de pago a la red de tarjetas

Petición de autorización al emisor de la tarjeta

El emisor de la tarjeta aprueba el pago

Reenvía la autorización

Reenvía la autorización

Completa la transacción

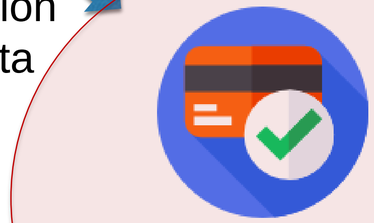
Caso Práctico: Detección de fraude bancario

Cliente paga con tarjeta

Solicitud de autorización de pago

Autorización de pago a la red de tarjetas

Petición de autorización al emisor de la tarjeta



El emisor de la tarjeta aprueba el pago

Reenvía la autorización

Reenvía la autorización

Completa la transacción

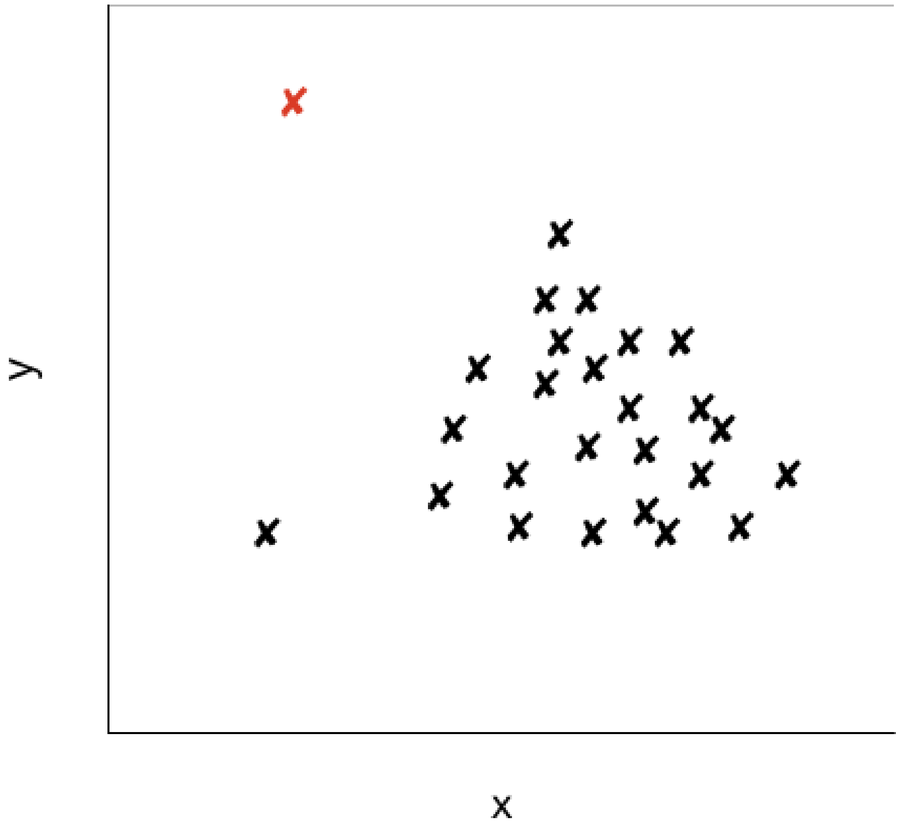
Isolation Forest

- **Algoritmo de detección de anomalías basado en aprendizaje no supervisado**
- **No perfila los datos normales**
- **No se basa en distancias para obtener los resultados**
- **IsolationForest forma un conjunto de árboles a partir del conjunto de datos y las anomalías son aquellos puntos con la longitud de camino promedio más corta**

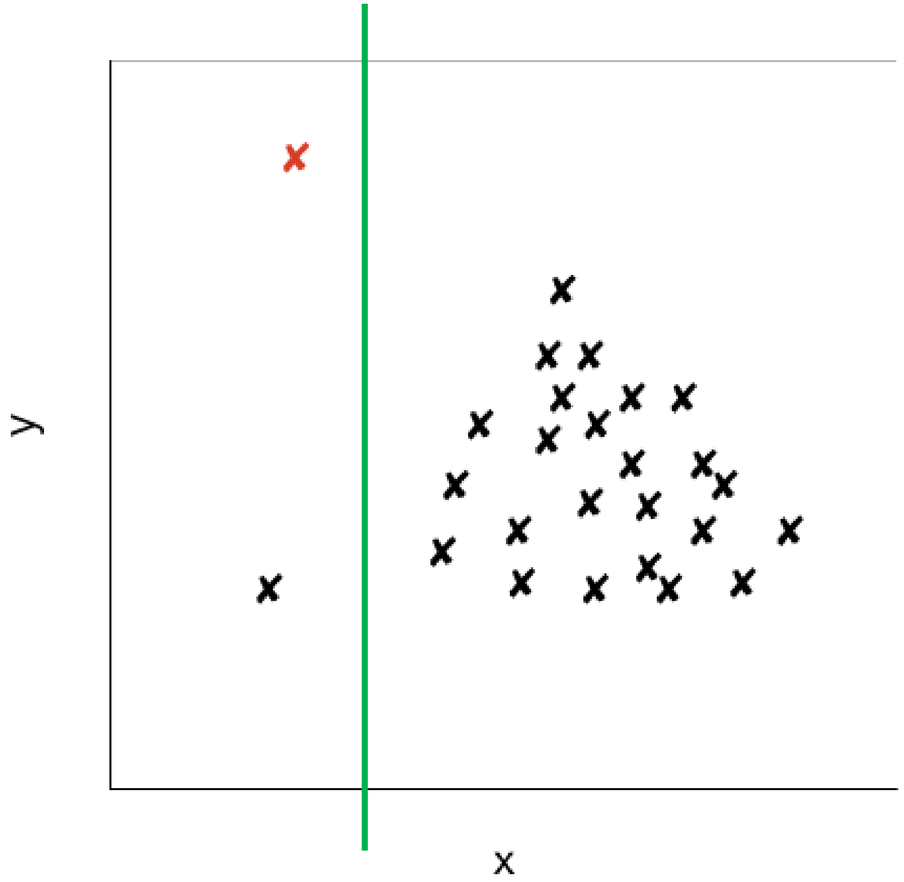
Isolation Forest: Aislando una anomalía

- **Isolation Forest se basa en las siguientes premisas:**
  - El subconjunto de datos anómalos es minoritario en relación con el subconjunto de datos normales
  - Los datos anómalos poseen características que los hacen muy diferentes con relación a los datos normales.

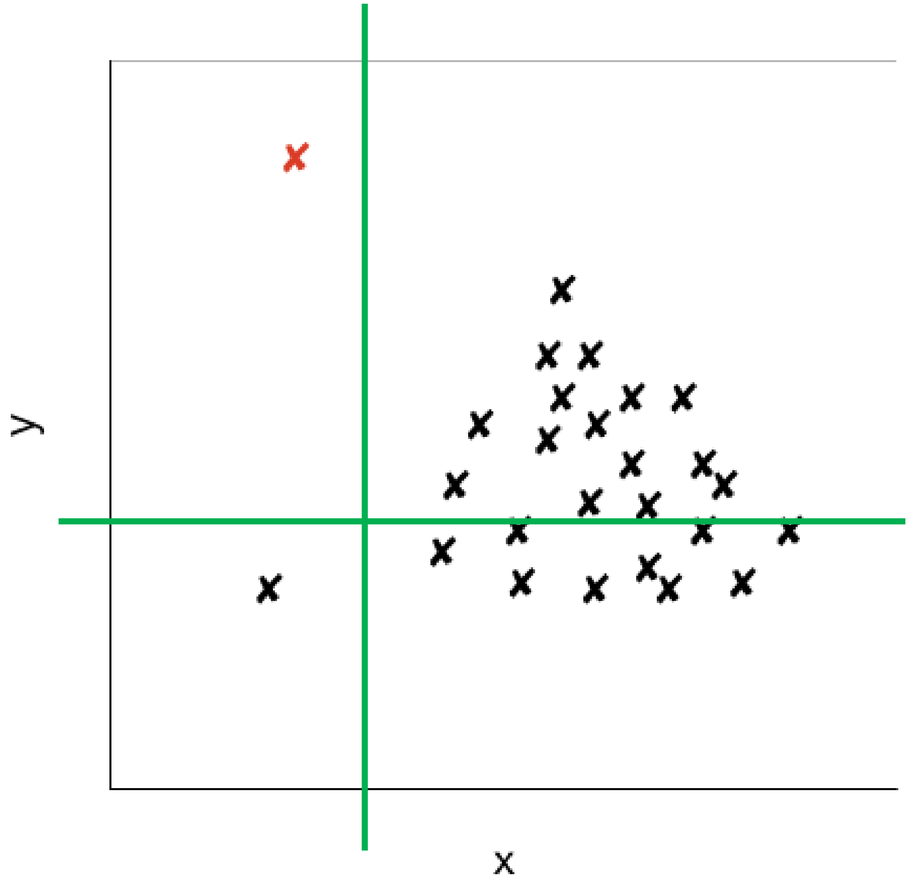
Isolation Forest: Aislando una anomalía



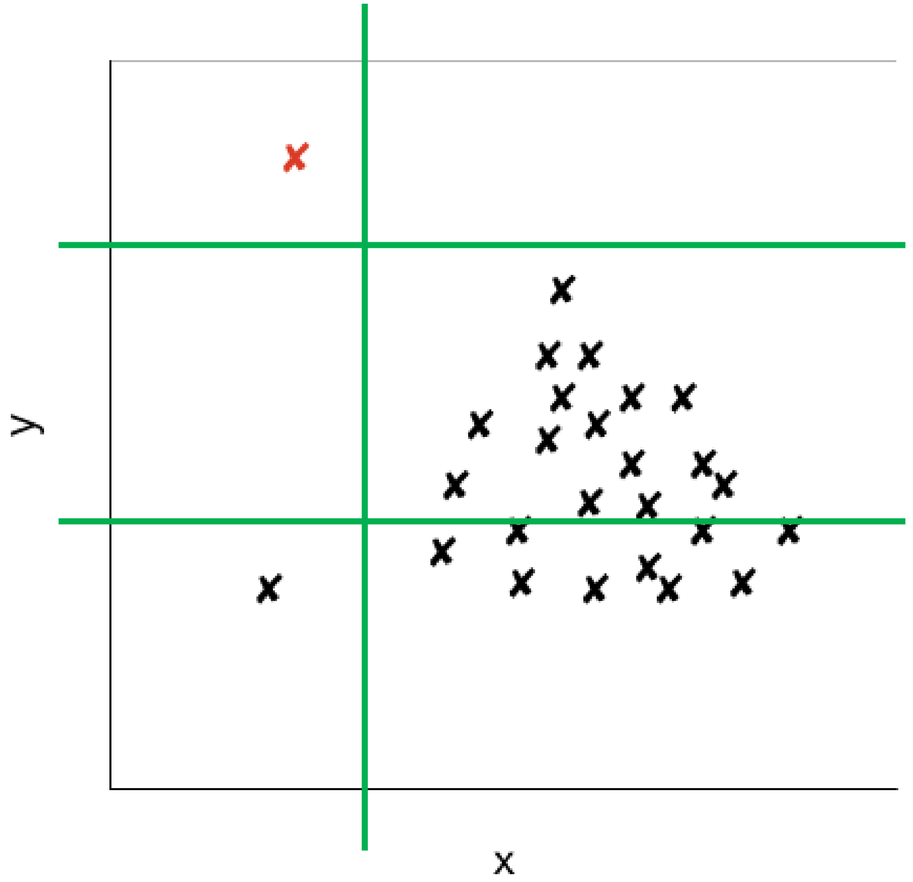
Isolation Forest: Aislando una anomalía



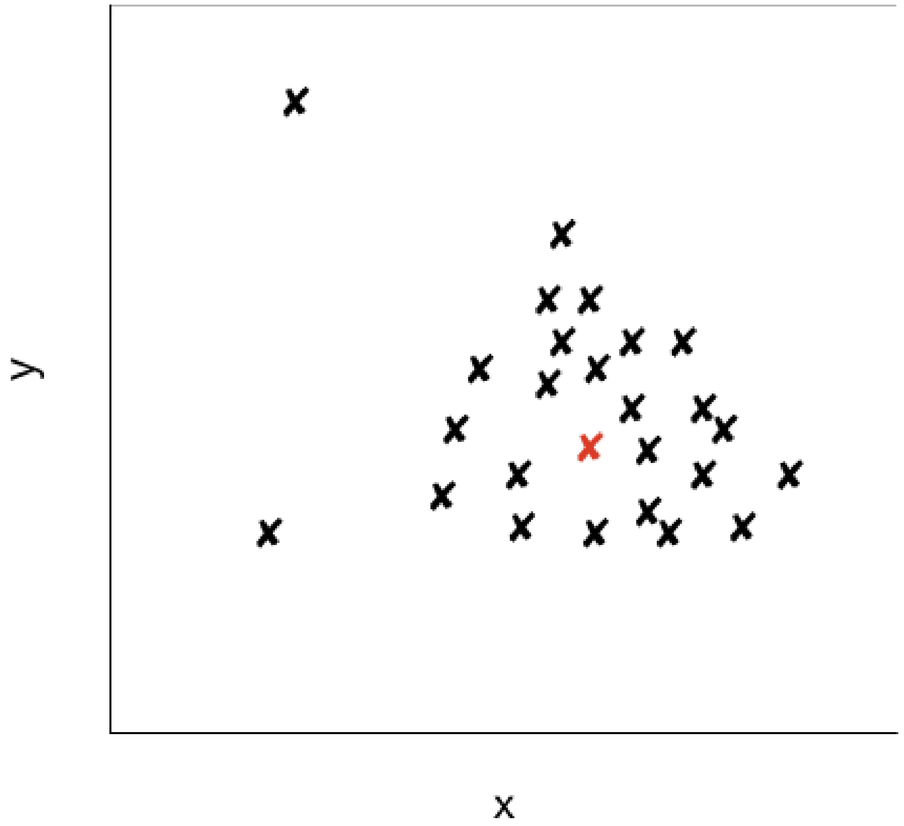
Isolation Forest: Aislando una anomalía



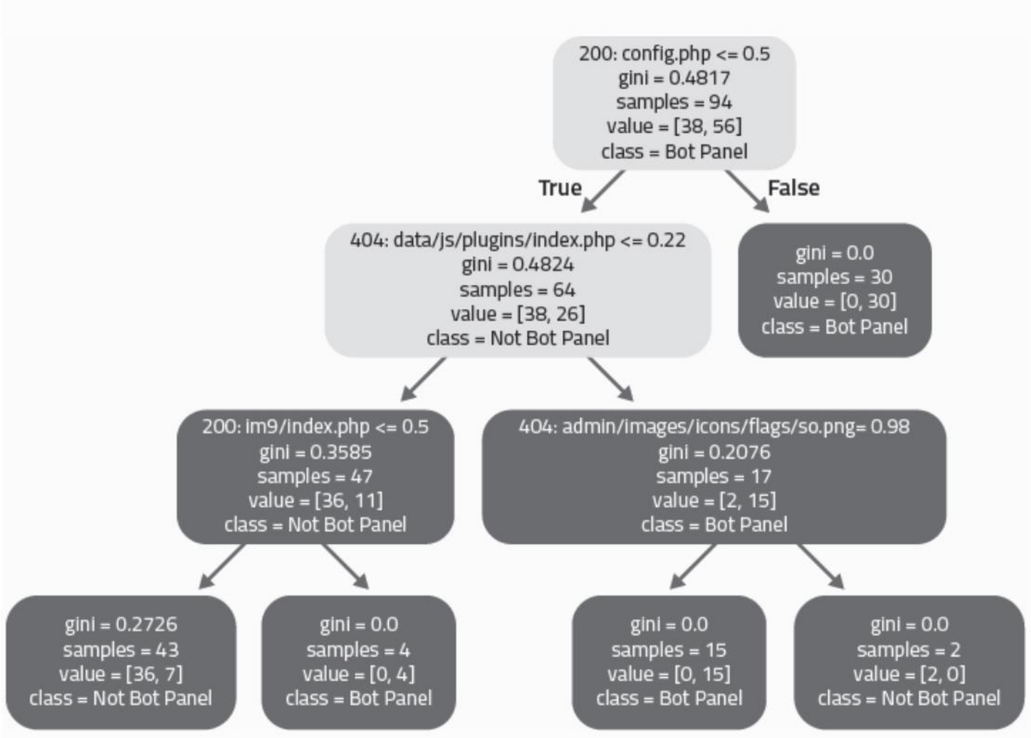
Isolation Forest: Aislando una anomalía



Isolation Forest: **Aislando un ejemplo normal**

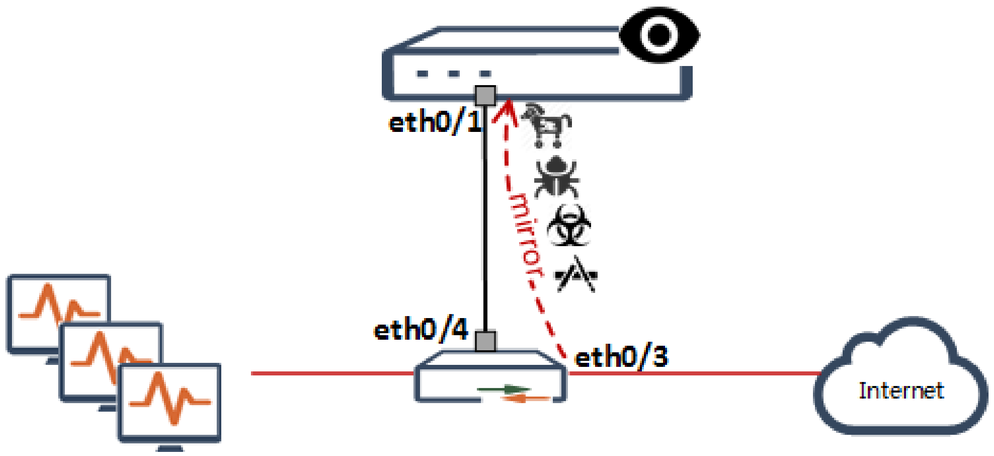


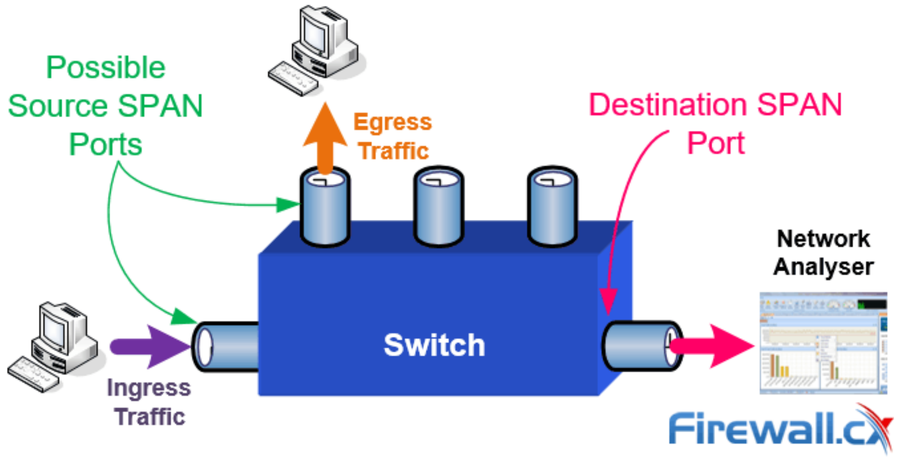
Isolation Forest: árboles de decisión



Cylance. "Introduction to AI for security professionals". 2017

Caso Práctico: Detección de anomalías en tráfico de red





TAP

PORT MIRRORING

GRACIAS

@santiagohramos
shramos@protonmail.com