

L U M B R E



Detección de anomalías

Grado: 5

Materia: Aprendizaje No Supervisado · Machine Learning

11/10/2026

Antes de empezar

Un fraude, un fallo inminente, una intrusión: rarezas que conviene cazar antes que clasificar. La detección de anomalías aprende lo normal y alza la mano ante lo que desentona, sin necesidad —casi nunca— de ejemplos previos de lo raro.

QUÉ APRENDERÁS

- Aplicar métodos de aislamiento y de densidad para señalar valores inusuales en problemas donde las anomalías son raras y heterogéneas.

Detección de anomalías

Se aprende lo normal, se caza la desviación

Las anomalías no comparten un patrón: cada rareza es rara a su manera. Por eso no se aprende «qué es una anomalía», sino «qué es normal», y se marca todo lo que se sale. Es un aprendizaje por exclusión, con las etiquetas siempre torcidas hacia el lado mayoritario.

VIDEO

Detección de datos anómalos con bosques de aislamiento

Vídeo en español que muestra cómo Isolation Forest señala los valores atípicos sin necesidad de ejemplos previos de anomalía.



Escanea para verlo en Lumbre

<https://www.youtube.com/watch?v=n2PpD7xnPoM>

Qué cuenta como anomalía

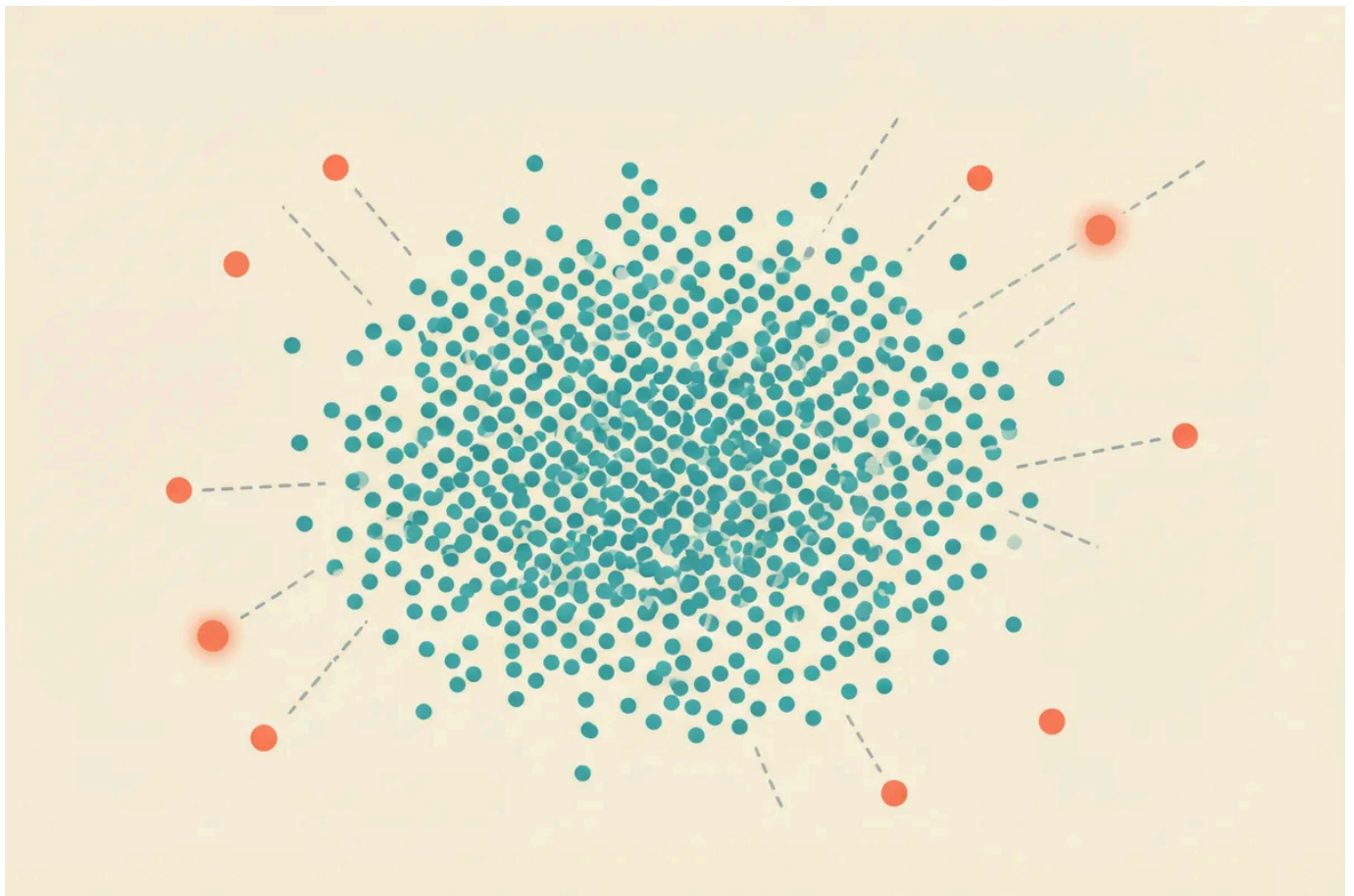
Un punto anómalo es aquel que se desvía tanto de la masa de datos que parece generado por otro proceso. Hay puntuales —un valor disparatado—, contextuales —normales en otra situación, sospechosos aquí— y colectivas —una secuencia rara de eventos por lo demás corrientes.

El reto del dato escaso

A diferencia de la clasificación, aquí casi no hay ejemplos positivos de anomalía, y los que hay son variadísimos. No se puede entrenar a reconocerlas a todas. La estrategia honesta es modelar la normalidad y señalar la baja densidad o lo difícil de aislar.

Bosque de aislamiento

Isolation Forest parte de una intuición brillante: una anomalía es fácil de separar del resto. Construye muchos árboles que parten el espacio aleatoriamente y mide cuántos cortes hacen falta para aislar cada punto. Los raros quedan aislados en pocos niveles: su camino es corto.



La lógica del aislamiento: lo normal forma un corazón denso que exige muchos cortes, mientras los puntos raros del borde quedan solos en pocos trazos y se marcan como anomalías.

Por qué el aislamiento funciona

Los puntos normales viven en el corazón denso y exigen muchas particiones para quedar solos; los atípicos, al borde, se separan en pocos pasos. Esa asimetría, sin suponer ninguna distribución, hace del bosque de aislamiento un método rápido y notablemente robusto en alta dimensión.

Densidad local: LOF

El factor de outliers de localización compara la densidad de un punto con la de sus vecinos. Si un punto es mucho menos denso que sus alrededores, puntúa como anomalía. A diferencia del aislamiento, capta rarezas relativas al contexto: normal en una zona, anómalo en otra.

Una clase SVM

Otra familia dibuja una frontera que envuelva la mayoría de los datos, tolerando un porcentaje esperado de outliers. Todo lo que cae fuera de esa envoltura se marca. Útil cuando la normalidad forma una región compacta identificable, aunque exige cuidar el kernel y sus parámetros.

Puntuación y umbral

Estos métodos no entregan un sí o no, sino una puntuación de anomalía. La decisión llega al fijar un umbral, y ese umbral es una elección de negocio: cuánto ruido admitimos y cuántas rarezas reales podemos permitirnos perder. No hay umbral inocente.

Evaluar sin etiquetas

Sin verdad conocida, validar es delicado. Se recurre a anomalías sintéticas inyectadas, a conjuntos de referencia etiquetados o a la revisión experta de las alertas. Reportar solo la tasa de alertas sin medir su calidad es la trampa favorita del optimismo.

Errores frecuentes

- Tratar la puntuación de anomalía como certeza absoluta en lugar de una ordenación de sospecha.
- Validar con datos donde la anomalía dominante se confunde con rareza inocente del fin de semana.
- Ignorar el coste de falsas alarmas y anegar al analista en alertas.

Ejemplo resuelto: cazar transacciones fraudulentas

Un banco quiere vigilar en línea sus millones de movimientos sin tener una lista completa de fraudes. El montaje:

1. Entrenar con lo normal: ajustar el bosque sobre transacciones corrientes, casi sin ejemplos de fraude.
2. Puntuar: asignar a cada movimiento una puntuación de anomalía según lo breve de su camino de aislamiento.

3. Fijar umbral: decidir cuánta alerta admite el equipo de riesgo sin saturarse.
4. Revisar: contrastar las banderas altas con confirmaciones reales para calibrar.
5. Vigilar: revisar el equilibrio entre fraudes cazados y falsas alarmas conforme cambia el patrón.

¿Para qué sirve en la realidad?

La detección de anomalías protege tarjetas, anticipa fallos en maquinaria, filtra intrusiones de red y destapa defectos de fabricación: donde las rarezas son pocas y muy variadas, aprender la normalidad y señalar la desviación es la única vía practicable.

1. ¿En qué intuición se basa el bosque de aislamiento para detectar anomalías?

- ☐ Que las anomalías son más fáciles de separar del resto mediante cortes aleatorios
- ☐ Que siguen una distribución gaussiana de cola pesada
- ☐ Que siempre son puntos lejanos al origen
- ☐ Que abundan más que los datos normales

2. La detección de anomalías necesita típicamente un gran conjunto de ejemplos etiquetados de anomalía.

- ☐ Verdadero
- ☐ Falso

Anomalías contextuales

Cada tres grados de madrugada es normal; en pleno verano, una ola de calor. El mismo valor es o no anómalo según el contexto, y cazarlo exige comparar no contra la distribución global sino contra la del escenario apropiado. LOF y modelos estacionales abordan esto mejor que un umbral global.

ACTIVIDAD INTERACTIVA DISPONIBLE EN LUMBRE

3. Anomalías



4. Ordena el mecanismo del bosque de aislamiento:

Elementos:

- dividir el espacio con cortes aleatorios repetidamente
- continuar hasta aislar cada punto en una hoja
- medir la profundidad media de aislamiento en muchos arboles
- convertir la brevedad del camino en puntuación de anomalía

Partir	. . .
Aislar	. . .
Promediar	. . .
Puntuar	. . .

5. Une cada método o concepto con su idea.

5.	aisla con cortes aleatorios y mide la profundidad	-----
5.	mide rareza relativa a la densidad de los vecinos	-----
5.	dibuja una frontera que envuelve la mayoría	-----
5.	salida ordenada que exige elegir un corte	-----

Opciones: isolation forest · lof · one-class svm · puntuación de anomalía

Cazar lo raro sin ejemplos

Detección de anomalías

Enfoque

aprender lo normal

cazar desviación

Isolation

cortes aleatorios

camino corto

Densidad

lof

rarezas contextuales

Decisión

puntuación

umbral de negocio

El mapa mental interactivo es interactivo en la versión online.



Guía del bosque de aislamiento <https://www.datacamp.com/es/tutorial/isolation-forest>
Explicación en español del algoritmo y su uso para detectar valores atípicos.

6. Tu detector marca el diez por ciento de las transacciones como fraudulentas, pero los analistas casi nunca confirman ninguna. ¿Qué parámetro y qué supuesto revisarías primero?

Referencia curricular

Grado:	5
Materia:	Aprendizaje No Supervisado · Machine Learning

Versión interactiva completa (con actividades, videos y autoevaluación):



<https://univ.lumbre.lat/object/cdia-nosu-006-deteccion-de-anomalias>

Material educativo generado desde Lumbre · lumbre.lat