

# Importancia de características

## Aprendizaje Automático

Juan David Martínez

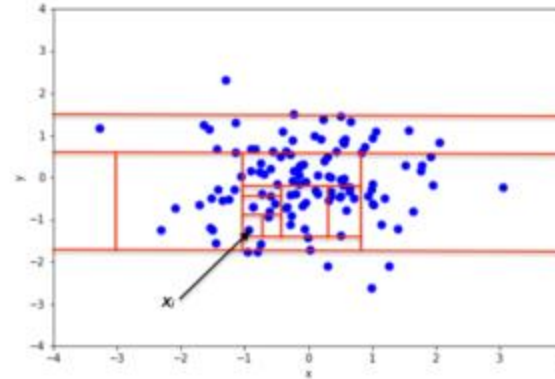
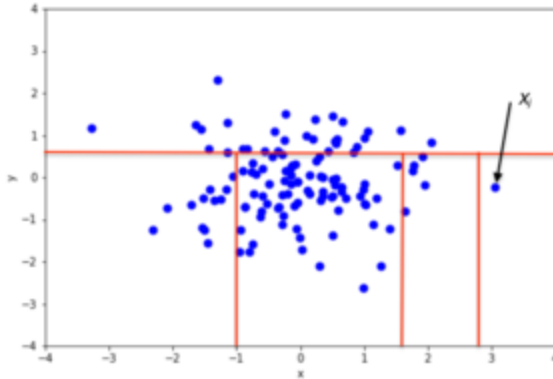
[jdmartinev@eafit.edu.co](mailto:jdmartinev@eafit.edu.co)

# Agenda

- Isolation trees
- Isolation forests

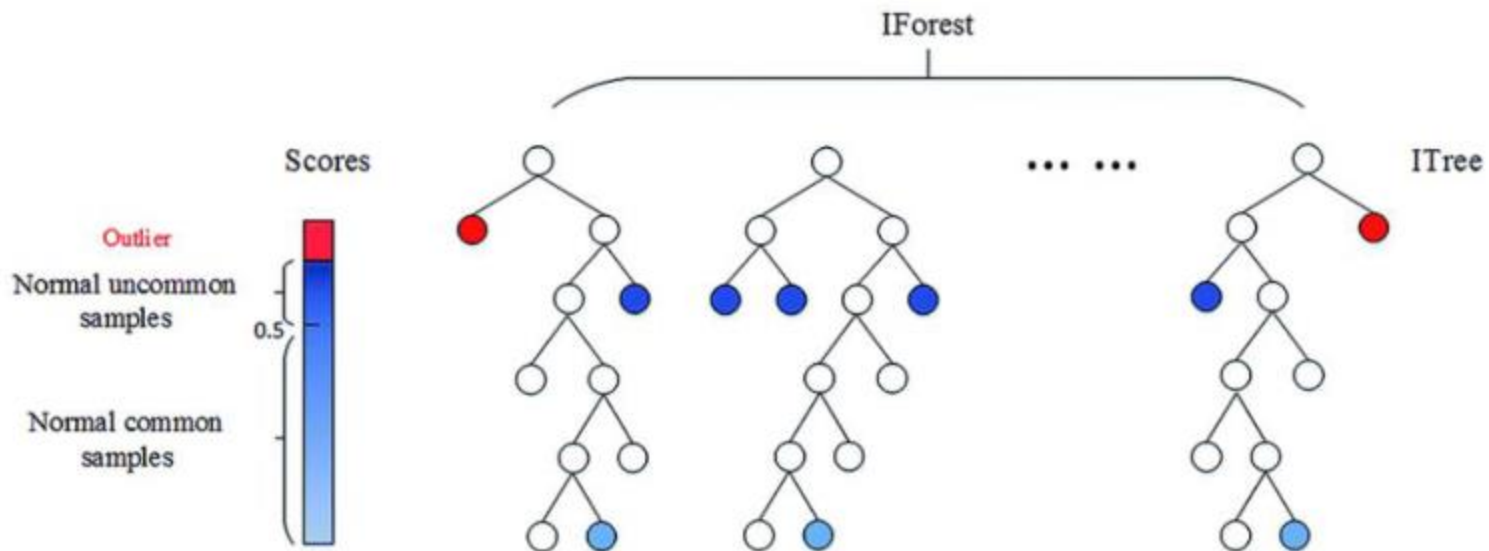
# Isolation trees

- Motivación:
  - Las anomalías generalmente son pocas muestras con características diferentes de las muestras “normales”
  - Podemos construir un árbol para aislar cada muestra
  - Las anomalías se agrupan cerca del nodo raíz
  - Las muestras normales se agrupan al final del árbol



# Isolation forests

- La distancia promedio al nodo terminal se puede ver como un score de novedad de una muestra



# Isolation forest

- **Isolation tree:** Dado un conjunto de muestras  $X$  de  $n$  muestras, este conjunto de datos se divide recursivamente seleccionando aleatoriamente una característica  $q$  con un valor de partición  $p$ , hasta que:
  - El árbol alcance una profundidad establecida
  - $|X| = 1$
  - Todas las instancias de  $X$  tienen el mismo valor
- **Definición:** Longitud de trayectoria
  - La longitud de trayectoria  $h(x^{(i)})$  de una muestra  $x^{(i)}$  se mide por el número de nodos que se atraviesan en un árbol desde el nodo raíz al nodo terminal en el que se encuentra la muestra  $i$
  - $h(x^{(i)})$  se normaliza con la longitud de trayectoria promedio de  $h(x^{(i)})$  dado  $n$ 
$$c(n) = 2H(n - 1) - (2(n - 1))/n$$
$$H(i) = \ln(i) + 0,57721 \text{ (constante de Euler)}$$

# Isolation forest

- **Definición:** score de novedad

- El score de novedad  $s$  de una muestra  $\mathbf{x}^{(i)}$  se define como:

$$s(\mathbf{x}^{(i)}, n) = 2^{-\frac{E(h(\mathbf{x}^{(i)}))}{c(n)}}$$

- $E(h(\mathbf{x}^{(i)})) \rightarrow c(n), s \rightarrow 0,5$
- $E(h(\mathbf{x}^{(i)})) \rightarrow 0, s \rightarrow 1$
- $E(h(\mathbf{x}^{(i)})) \rightarrow n - 1, s \rightarrow 0$

# Isolation forest

[https://scikit-learn.org/stable/auto\\_examples/ensemble/plot\\_isolation\\_forest.html](https://scikit-learn.org/stable/auto_examples/ensemble/plot_isolation_forest.html)