



Lección 4. Métodos de detección de anomalías basados en estadística en Python



La detección de anomalías

Métodos basados en estadística en Python

UNIVERSIDAD DE CÓRDOBA

Vamos a trabajar en Python, que se ha utilizado ya en esta asignatura y dispone de una gran cantidad de librerías para poder aplicar métodos de detección de anomalías.

Para poder ver los ejemplos, no necesitas tener nada instalado en tu máquina, ya que lo puedes hacer desde *google collaboration* que te proporciona todo lo que vas a necesitar para ver los ejemplos, ejecutarlos y modificarlos. Para su acceso solamente necesitas un navegador, conexión a Internet y una cuenta de google.

En esta sección, verás los ejemplos que tienes disponibles y los iremos comentando. Todos los métodos serán basados en estadísticas de los vistos en la lección 1.

1. Aplicando métodos basados en estadística

En esta sección vamos a ver cómo funcionan los métodos basados en estadísticas aplicándolos a conjunto de datos para detectar anomalías.

Como material adicional tienes varios notebooks utilizando las técnicas estadísticas que hemos visto para detectar anomalías en la lección 1. Vamos a usar estos ejemplos para encontrar datos anómalos. Esta lección es muy importante para poder hacer la tarea de esta semana, ya que estos notebook serán la base para la tarea que debes entregar al final de la semana. En los dos vídeos que tienes disponibles de esta lección se comenta con detalle los diferentes ejemplos.

Los ejemplos disponibles son:

- **Zscore_ejemplo_v1.ipynb**: este ejemplo muestra el uso de z-score para detección de anomalías. Su uso nos indica a cuántas desviaciones estándar se encuentra un punto de datos de la media. Para su cálculo utilizamos la fórmula vista en la lección 1.
- **Zscore_ejemplo_v2.ipynb**: este ejemplo muestra el uso de z-score para detección de anomalías. Su uso nos indica a cuántas desviaciones estándar se encuentra un punto de datos de la media. Para su cálculo utilizamos la función que calcula los z-score disponible en Python.
- **TestTukey_ejemplo.ipynb**: este ejemplo muestra el Test de Tukey o cálculo del rango intercuartílico (IQR) para la detección de anomalías. El resultado del test nos confirma si existen o no datos anómalos.

- **HBOS_ejemplo.ipynb**: este ejemplo aplica el método HBOS basado en histogramas para detección de anomalías. HBOS asume la independencia entre las características y calcula el grado de anomalías mediante la construcción de histogramas. El resultado nos confirma si existen o no datos anómalos. Se utiliza la librería PyOD de Python.

Te animo a cambiar configuraciones de los métodos y de los datos iniciales y ver cómo afecta a los resultados.