



Fundamentos de la Ciencia de Datos. Prueba de Laboratorio 5 (PL5)

La Prueba de Laboratorio constará de tres partes:

1. La primera parte consiste en la realización de cuatro ejercicios en clase con ayuda del profesor en el que se va a realizar un análisis de detección de datos anómalos con R aplicando todos los conceptos vistos en el tema.
 - 1.1. En el primero se realizará un análisis de detección de datos anómalos utilizando medidas de ordenación sobre la resistencia, método de caja y bigotes sobre la misma muestra¹ que se utilizó para hacer de forma manual el ejercicio de detección de outliers.
 - 1.2. En el segundo se realizará un análisis de detección de datos anómalos utilizando medidas de dispersión sobre la densidad, desviación típica sobre la misma muestra del ejercicio 1.
 - 1.3. En el tercero se realizará un análisis de detección de datos anómalos sobre la regresión de las variables, densidad en función de la resistencia, utilizando el error estándar de los residuos sobre la misma muestra del ejercicio 1.
 - 1.4. En el cuarto se resolverá, utilizando el algoritmo K-vecinos, el mismo problema que el visto en la descripción teórica del tema. Es decir, para la misma muestra² que se utilizó para hacer de forma manual el ejercicio de detección de datos anómalos, se deberán obtener los outliers.
2. La segunda parte consiste en el desarrollo por parte de cada grupo del enunciado y la solución de un ejercicio en el que se realicen análisis con R de detección de datos anómalos utilizando todos los métodos vistos en teoría, con excepción de LOF, e introduciendo modificaciones sobre el ejercicio hecho en clase.
3. La tercera parte se resolverá, utilizando el algoritmo LOF, el mismo problema que el visto en la descripción teórica del tema. Es decir, para la misma muestra² que se utilizó para hacer de forma manual el ejercicio de detección de datos anómalos, se deberán obtener los outliers.

Se debe completar en el documento del grupo utilizando Sweave y Noweb, para realizar Literate-Programming y entregar los archivos .Rnw, .tex y .pdf generados por R con el resultado de las dos partes de la PL5 y los ficheros de datos utilizados. La PL5 se entregará como actividad en Blackboard. El nombre de los ficheros será: Grupo.NúmerodeGrupo-PL5

¹ Está formada por los siguientes 7 valores de resistencia y densidad para diferentes tipos de hormigón {resistencia, densidad}: {3, 2; 3.5, 12; 4.7, 4.1; 5.2, 4.9; 7.1, 6.1; 6.2, 5.2; 14, 5.3}.

² Está formado por las siguientes 5 calificaciones de estudiantes: 1. {4, 4}; 2. {4, 3}; 3. {5, 5}; 4. {1, 1}; 5. {5, 4} donde las características de las calificaciones son: {Teoría, Laboratorio}.