



Universidad  
de Alcalá

#### **Fundamentos de la Ciencia de Datos. Prueba de Laboratorio 4 (PL4)**

La Prueba de Laboratorio constará de tres partes:

1. La primera parte consiste en la realización de un ejercicio en clase con ayuda del profesor en el que se va a realizar un análisis de clasificación no supervisada de Datos con R aplicando todos los conceptos vistos en el tema. En dicho ejercicio se resolverá, utilizando el algoritmo K-means, el mismo problema que el visto en la descripción teórica del tema. Es decir, para la misma muestra<sup>1</sup> que se utilizó para hacer de forma manual el primer ejercicio de clasificación no supervisada, se deberán obtener los centroides de los clusters obtenidos.
2. La segunda parte consiste en el Desarrollo por parte de cada alumno del enunciado y la solución de un ejercicio en el que se realice un análisis con R de clasificación no supervisada introduciendo modificaciones sobre el ejercicio hecho en clase.
3. Aplicar y analizar el paquete de R: LearnClust: Learning Hierarchical Clustering Algorithms

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 PL4 y los ficheros de datos utilizados. La PL4 se entregará como actividad en Blackboard. El nombre de los ficheros será: Grupo.NúmerodeGrupo-PL4

---

<sup>1</sup> Está formado por las siguientes 8 calificaciones de estudiantes: 1. {4, 4}; 2. {3, 5}; 3. {1, 2}; 4. {5, 5}; 5. {0, 1}; 6. {2, 2}; 7. {4, 5}; 8. {2, 1}, donde las características de las calificaciones son: {Teoría, Laboratorio}.