

Extracción Automática del Conocimiento-781006. R-PL3

La prueba se deberá entregar en dos semanas y tiene una puntuación de 1 punto.

La Prueba de Laboratorio constará de dos partes:

1. La primera parte consiste en la realización de dos ejercicios en clase con ayuda del profesor en el que se va a realizar un análisis de clasificación de Datos con R aplicando todos los conceptos vistos en el tema.
 - 1.1. En el primero se resolverá, utilizando el algoritmo construcción de árboles de decisión de Hunt, el mismo problema que el visto en la descripción teórica del tema. Es decir, para la misma muestra¹ que se utilizó para hacer de forma manual el primer ejercicio de clasificación para datos cualitativos, se deberá obtener, utilizando la medida de Ganancia de información, mediante la medida de impureza Gini, la función de clasificación.
 - 1.2. En el segundo se realizará un análisis de regresión lineal para la misma muestra² que se utilizó para hacer de forma manual el primer ejercicio de clasificación para datos cuantitativos.
2. La segunda parte consiste en tres ejercicios:
 - 2.1. Para los datos de un fichero .txt generado a partir de los datos de la muestra³ utilizada para hacer de forma manual el segundo ejercicio de clasificación para datos cualitativos, se debe resolver con R, utilizando el algoritmo de Hunt, el mismo problema que el visto en la descripción teórica del tema, es decir, se deberá obtener utilizando la medida de Ganancia de información, mediante la medida de impureza Gini, la función de clasificación.
 - 2.2. Para los datos de un fichero .txt generado a partir de los datos de la muestra⁴ utilizada para hacer el segundo ejercicio de clasificación para datos cuantitativos, hacer un análisis de regresión lineal.
 - 2.3. 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 introduciendo modificaciones sobre el ejercicio hecho en clase.

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 PL3.

¹ Está formado por las siguientes 9 calificaciones de estudiantes: 1. {A, A, B, Ap}; 2. {A, B, D, Ss}; 3. {D, D, C, Ss}; 4. {D, D, A, Ss}; 5. {B, C, B, Ss}; 6. {C, B, B, Ap}; 7. {B, B, A, Ap}; 8. {C, D, C, Ss}; 9. {B, A, C, Ss}, donde las características de las calificaciones son: {Teoría, Laboratorio, Prácticas, Calificación Global}.

² Está formada por los siguientes 4 radios ecuatoriales y densidades de los planetas interiores : {Mercurio, 2.4, 5.4; Venus, 6.1, 5.2; Tierra, 6.4, 5.5; Marte, 3.4, 3.9}

³ Está formada por las características de 10 vehículos de cuatro tipos diferentes: 1. {B, 4, 5, Coche}; 2. {A, 2, 2, Moto}; 3. {N, 2, 1, Bicicleta}; 4. {B, 6, 4, Camión}; 5. {B, 4, 6, Coche}; 6. {B, 4, 4, Coche}; 7. {N, 2, 2, Bicicleta}; 8. {B, 2, 1, Moto}; 9. {B, 6, 2, Camión}; 10. {N, 2, 1, Bicicleta}, donde los elementos de cada suceso son: {TipoCarnet, NúmeroRuedas, NúmeroPasajeros, TipoVehículo}.

⁴ Está formada por las siguientes 4 muestras de pares de datos: 1. {10, 8.04; 8, 6.95; 13, 7.58; 9, 8.81; 11, 8.33; 14, 9.96; 6, 7.24; 4, 4.26; 12, 10.84; 7, 4.82; 5, 5.68}; 2. {10, 9.14; 8, 8.14; 13, 8.74; 9, 8.77; 11, 9.26; 14, 8.1; 6, 6.13; 4, 3.1; 12, 9.13; 7, 7.26; 5, 4.74}; 3. {10, 7.46; 8, 6.77; 13, 12.74; 9, 7.11; 11, 7.81; 14, 8.84; 6, 6.08; 4, 5.39; 12, 8.15; 7, 6.42; 5, 5.73}; 4. {8, 6.58; 8, 5.76; 8, 7.71; 8, 8.84; 8, 8.47; 8, 7.04; 8, 5.25; 19, 12.5; 8, 5.56; 8, 7.91; 8, 6.89}