

Inteligencia artificial: Práctica 8

Álvaro García Tenorio *

Miguel Pascual Domínguez **

1 de mayo de 2018

Índice general

| | |
|--|----------|
| Índice general | 1 |
| 1 Agrupamiento | 2 |
| 1.1. Descripción del conjunto | 2 |
| 1.2. Parametrización del algoritmo de agrupamiento | 3 |
| 1.3. Resultados | 3 |
| 1.4. Conclusiones y respuestas | 3 |
| 2 Clasificación | 4 |
| 2.1. Descripción del conjunto | 4 |
| 2.2. Parametrización del J48 | 4 |
| 2.3. Resultados | 4 |
| 2.4. Conclusiones y respuestas | 4 |
| 3 Regresión | 5 |
| 3.1. Descripción del conjunto | 5 |
| 3.2. Parametrización del perceptrón multicapa y del k-NN | 5 |
| 3.3. Resultados | 5 |
| 3.4. Conclusiones | 5 |
| 4 Ajuste Opcional | 6 |

* alvgar14@ucm.es

** miguepas@ucm.es

Capítulo 1

Agrupamiento

1.1. Descripción del conjunto

Para este experimento hemos seleccionado un conjunto de datos llamado ausencias de trabajo. Los datos pertenecen a un estudio académico realizado por la universidad Nove de Julho de Brasil, por dos estudiantes y un profesor, durante tres años desde 2007 hasta 2010 en una empresa brasileña. Los datos, los hemos obtenido de la siguiente pagina, <http://archive.ics.uci.edu/ml/datasets/Absenteeism+at+work>.

Explicuemos ahora los atributos/variables de este conjunto de datos:

1. ID (Identificador): variable formada por un conjunto de números del 1 al 36 que consisten en 36 identificadores de 36 personas distintas.
2. Motivo de la ausencia: catalogado en 28 números que cada uno corresponde con un motivo de incidencia, los 21 primeros se tratan de las primeras 21 enfermedades del código internacional de enfermedades (ICD) y los 7 ultimos, corresponden con 7 motivos no incluidos en el ICD, que son el acompañamiento de otro paciente (22), una consulta médica (23), donación de sangre (24), análisis médico (25), ausencia injustificada (26), fisioterapia (27) y dentista (28).
3. Mes de ausencia: número real que corresponde con el mes de la ausencia del 1 al 12.
4. Día laboral de la semana: lunes (2), martes (3), miércoles (4), jueves (5) y viernes (6).
5. Estación del año: verano (1), otoño (2), invierno (3) y primavera (4).
6. Gasto de transporte: número real.
7. Distancia desde casa hasta la oficina (km): número real.
8. Tiempo de trabajo: número entero.
9. Edad: número entero
10. Media de carga de trabajo: número real.
11. Porcentaje de cumplimiento de trabajos: número real entre 0.0 y 100.0.
12. Incidencia disciplinaria: Si (1) y No (0).

13. Educación: número real, catalogado de la siguiente manera; bachillerato (1), graduado (2), posgraduado (3) y master y doctorado (4).
14. Número de hijos: número real.
15. Bebedor social: Si (1) y No (0).
16. Fumador social: Si (1) y No (0).
17. Número de mascotas: número real.
18. Peso: número real.
19. Altura: número real.
20. Índice de masa corporal: número real.
21. Tiempo de ausencia en horas: número real.

1.2. Parametrización del algoritmo de agrupamiento

1.3. Resultados

1.4. Conclusiones y respuestas

Capítulo 2

Clasificación

- 2.1. Descripción del conjunto
- 2.2. Parametrización del J48
- 2.3. Resultados
- 2.4. Conclusiones y respuestas

Capítulo 3

Regresión

3.1. Descripción del conjunto

3.2. Parametrización del perceptrón multicapa y del k-NN

3.3. Resultados

3.4. Conclusiones

Capítulo 4

Ajuste Opcional