

Machine learning applied to

Pablo Arranz ropero y Juan Alberto camino Sáez

heart diseases dataset

El propósito de este proyecto es aplicar el aprendizaje automático a un conjunto de datos de enfermedades cardíacas encontrado en la siguiente URL:

Pablo Arranz ropero y Juan Alberto camino Sáez

<https://archive.ics.uci.edu/ml/datasets/Statlog+%28Heart%29>

El conjunto de datos presenta un total de 270 instancias con 13 atributos cada una.

Un ejemplo de caso de entrenamiento sería el siguiente:

70.0 1.0 4.0 130.0 322.0 0.0 2.0 109.0 0.0 2.4 2.0 3.0 3.0 2

Cada uno de los datos representa lo siguiente:

1. Edad
2. Sexo {0, 1} (no se cual es cual, hay que mirar los papers a ver si pone algo)
3. Tipo de dolor en el pecho {1, 2, 3, 4} (Siendo 1= ninguno, 2 = leve (el dolor no le impide hacer actividades normales), 3 = moderado (el dolor le impide hacer actividades y le despierta) y 4 = intenso (el dolor es el peor que ha tenido en su vida))
4. Presión de la sangre en reposo
5. Colesterol sérico (medido en mg/dl)
6. Azúcar en sangre en ayunas mayor que 120 mg/dl {0, 1}
7. Resultados electrocardiográficos en reposo {0, 1, 2} (no se lo que significa cada uno hulio)
8. Frecuencia cardiaca máxima alcanzada
9. Angina inducida por el ejercicio {0, 1}
10. Depresión ST inducida por el ejercicio en relación con el reposo (oldpeak)
11. Pendiente del segmento ST de ejercicio máximo (número de pendiente, primera segunda…)
12. Número de vasos principales (0-3) coloreados por flouroscopía {0, 1, 2, 3}
13. Thal {3, 6, 7} (siendo 3 = normal, 6 = defecto arreglado, 7 = defecto reversible) \*

\*Thal se refiere a puntos fríos en el scanner, donde no aparece talio, que indican áreas del corazón que no están recibiendo un suministro adecuado de sangre. Los puntos fríos que aparecen tanto en reposo como durante el ejercicio pueden indicar áreas donde el tejido cardíaco ha sido dañado. Sin embargo, los puntos fríos reversibles que aparecen solo durante el ejercicio suelen indicar algún tipo de bloqueo en las arterias coronarias.

Todos estos datos (cada uno en su respectiva columna) generan un resultado que hallamos en la última columna:

Enfermedad cardíaca {1, 2} (Siendo 1 = ausencia y 2 = presencia)