

Prueba y Comparación de Naive Bayes y C4.5

Juan Carlos López López

Adolfo Marín Arriaga

Luis Rodrigo Rojo Morales

12 de noviembre de 2016

1. Introducción

Para hacer las pruebas de estos algoritmos usamos el conjunto de datos que esta destinado a probar los clasificadores, este se puede obtener en: <https://archive.ics.uci.edu/ml/datasets/Adult> y en nuestro repositorio se encuentra en: /Entregable3/NaiveBayes/AdultDataSetTest.csv. Dicho conjunto tiene 16,281 registros, los cuales cada uno tiene los mismos atributos que el conjunto de datos original.

2. Naive Bayes

El script para probar este clasificador se encuentra en /Entregable3/NaiveBayes/TestAdultDataSet.java el objetivo de este script es cargar el conjunto de datos prueba, aplicar el clasificador y dar los datos de cuantos clasifico bien y cuantos mal, para así poder hacer la matriz de confusión, la cual queda de la siguiente manera:

	> 50K	≤ 50K
> 50K	2920	926
≤ 50K	2107	10328

Con esta matriz podemos observar que el clasificador predijo 13248 bien y 3033 mal, a 926 personas le dijimos que iban a ganar $\leq 50K$ pero en realidad ganaron $> 50K$, mientras que a 2107 personas les dijimos que iban a ganar $> 50K$ pero ganaron $\leq 50K$.

El clasificador en general tiene una exactitud del 81.37 %, mientras que individualmente predice a los que ganan más de 50K con una exactitud de 75.92 % y a los que ganan menos o 50K con una exactitud de 83.05 %

3. C4.5

4. Comparación