

Naive Bayes

Adolfo Marín Arriaga

Juan Carlos López López

Luis Rodrigo Rojo Morales

13 de noviembre de 2016

- Evaluación de patrones

Los atributos que hacen que una persona tenga más probabilidad de caer en la clase >50k son si tiene un capital gain mayor que cero, un grado de educación Doctorate, Masters o Prof-School, una clase de trabajo de self-emp-inc, otras que también aumentan la probabilidad de que caiga en >50k pero menos que las anteriores son: tener entre 41 y 60 años, tener un grado de educación Bachelors, trabajar entre 41 y 80 horas a la semana, tener un estado civil Married-civ-spouse o Married-AF-spouse, tener ocupación de Exec-managerial, o Prof-speciality, tener en Relationship Wife o Husband, tener una clase de trabajo Federal-gov. Si no tiene ninguna de estas o muy pocas del segundo grupo y ninguna del primero es muy probable que caiga en la clase <=50k.

- Clasificadores para la comparación

Para la comparación, los clasificadores Naive Bayes que se usaron son los del archivo NaiveBayes.java que es nuestra implementación y el de la biblioteca e1071 en r.

- Primer clasificador

Usando el de NaiveBayes.java para clasificar los elementos que vienen en el archivo AdultDataSetTest.csv nos da el resultado:

```
Clasifico 3033 mal
Clasifico 13248 bien
Hay 3846 >50K
Hay 12435 <=50K
Hay 5027 predicciones >50K
Hay 11254 predicciones <=50K
```

- Segundo clasificador

Usando la biblioteca e1071 de r para clasificar los elementos que vienen en el archivo AdultDataSetTest.csv nos da el resultado:

```

> table(ok)
ok
FALSE  TRUE
 2851 13430
> table(datest$V15)

<=50K  >50K
12435  3846
> table(pr)
pr
<=50K  >50K
13552  2729

```

- Comparación

Lo que nos dicen estos resultados es que hay 12435 personas de la clase $\leq 50k$ y hay 3846 de la clase $> 50k$.

El primer clasificador clasificó 11254 personas en la clase $\leq 50k$ y 5027 en la clase $> 50k$.

El segundo clasificador clasificó 13552 personas en la clase $\leq 50k$ y 2729 en la clase $> 50k$.

El primer clasificador clasificó bien a 13248 personas y mal a 3033 personas.

El segundo clasificador clasificó bien a 13430 personas y mal a 2851 personas.

El primer clasificador acertó en el 81.37% de los casos.

El segundo clasificador acertó en el 82.49% de los casos.