

# Trabajo de Reglas de Asociación - Aprendizaje no supervisado y detección de anomalías

*M<sup>a</sup> Cristina Heredia Gómez*

*1-3-2018*

Para esta práctica se ha elegido inspeccionar el dataset Zoo del paquete **data**. Puesto que el dataset consta de algunas variables numéricas como “legs” es necesario convertir las columnas a factores.

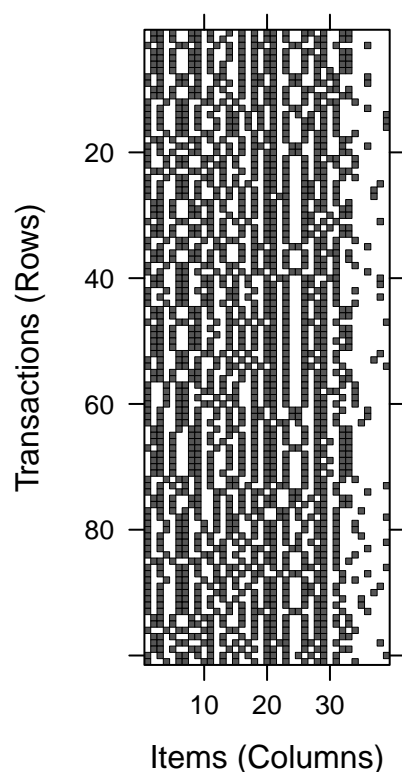
## Preprocesamiento

```
has_legs <- Zoo[["legs"]]>0
Zoo[["legs"]] <- has_legs
#pasamos a factores todas las columnas
Zoo[,1:16]<-lapply(Zoo[, 1:16],as.factor)
```

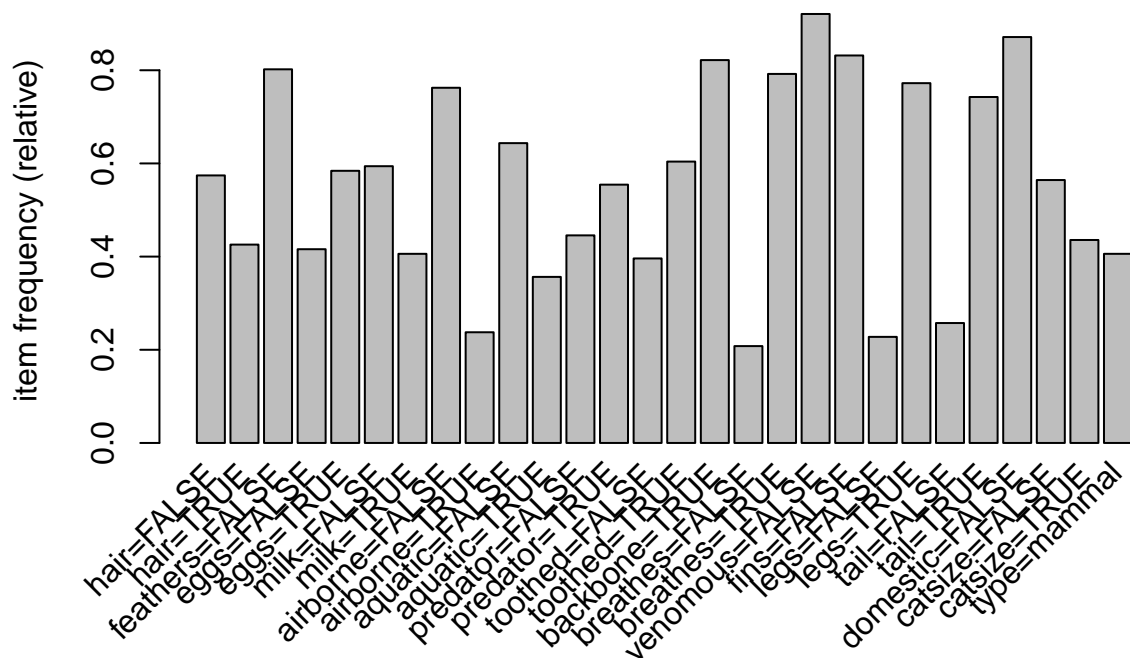
En primera instancia, como la variable “legs” toma valores entre 0 y 8, la transformamos a una variable binaria que toma valores TRUE o FALSE, en función de si el animal no tiene piernas o tiene alguna pierna. Tras esto, se pasan todas las columnas del dataset a tipo factor para poder crear las transicciones que necesita Apriori.

## Búsqueda de reglas

Tras crear las transacciones se visualizan los items y se analiza su frecuencia.



Se observa que hay items como el 20 que figuran en muchas transacciones, mientras que hay otros como el 35 o el 40, que figuran en muy pocas.



En el gráfico de los itemsets frecuentes se observa que hay pocos animales sin piernas, que no respiren, que no vuelen o que se transporten por el aire. Sin embargo, hay items muy frecuentes como no tener plumas, no ser venenoso o no tener aletas. Parece interesante pues centrar la búsqueda inicial en torno a los animales que no tienen piernas.

Tras lanzar Apriori para extraer reglas con minSop=0.2 ,conf=0.85 y minlen=2 inicialmente y filtrar por mayor valor de lift, se filtran aquellas reglas en las que aparezca el tipo “mamífero” dado que aparecía dentro de los items frecuentes en poca medida y queríamos estudiarlo mejor junto con el item “piernas”. Tras la poda de reglas relevantes, obtuve las siguientes 6 reglas de 2 items en el antecedente:

lhs	rhs	support	confidence	lift	count
[1] {milk=TRUE,type=mammal}	=> {breathes=TRUE}	0.4059406	1	1.262500	41
[2] {milk=TRUE,type=mammal}	=> {feathers=FALSE}	0.4059406	1	1.246914	41
[3] {milk=TRUE,type=mammal}	=> {backbone=TRUE}	0.4059406	1	1.216867	41
[4] {milk=TRUE,type=mammal}	=> {venomous=FALSE}	0.4059406	1	1.086022	41
[5] {breathes=TRUE,type=mammal}	=> {feathers=FALSE}	0.4059406	1	1.246914	41
[6] {breathes=TRUE,type=mammal}	=> {backbone=TRUE}	0.4059406	1	1.216867	41

que dicen cosas del tipo si da leche y es mamífero entonces respira, o si da leche y es mamífero no tiene plumas. Además, todas ellas dan una confianza del 100%, por tanto, no nos cuentan nada nuevo. Seguí buscando, manteniendo el mismo soporte y confianza pero esta vez filtrando las reglas que tuvieran que ver algo con mi otro punto de mira: las piernas. Obtuve lo siguiente:

lhs	rhs	support	confidence	lift	count
[1] {hair=FALSE,legs=FALSE}	=> {airborne=FALSE}	0.2178218	1.0000000	1.311688	22
[2] {airborne=FALSE,legs=FALSE}	=> {hair=FALSE}	0.2178218	0.9565217	1.665667	22
[3] {hair=FALSE,legs=FALSE}	=> {feathers=FALSE}	0.2178218	1.0000000	1.246914	22
[4] {feathers=FALSE,legs=FALSE}	=> {hair=FALSE}	0.2178218	0.9565217	1.665667	22
[5] {hair=FALSE,legs=FALSE}	=> {domestic=FALSE}	0.2079208	0.9545455	1.095558	21
[6] {legs=FALSE,domestic=FALSE}	=> {hair=FALSE}	0.2079208	0.9545455	1.662226	21

Ojeando las reglas, vemos que hay algunas, por ejemplo la [4] que empiezan a parecer curiosas. La mayoría de los animales que no tienen plumas ni piernas, no tienen pelo. Lógico, los peces, un claro ejemplo. Pero

la confianza de esa regla es 0.9565. Dicho de otra forma, sí que hay algún animal que, sin tener plumas ni piernas, tiene pelo. Lo buscamos:

```
which(Zoo$feathers==FALSE & Zoo$legs==FALSE & Zoo$hair==TRUE)
```

```
## [1] 75
```

```
Zoo[75,]
```

```
##      hair feathers  eggs milk airborne aquatic predator toothed backbone
## seal  TRUE     FALSE FALSE TRUE     FALSE     TRUE     TRUE     TRUE     TRUE
##      breathes venomous fins  legs  tail domestic catsize  type
## seal      TRUE     FALSE TRUE FALSE FALSE     FALSE     TRUE mammal
```

y encontramos a la foca! Aumentando un poco el tamaño mínimo de las reglas que estamos buscando (esta vez lo fijamos en 3) pero de nuevo, filtrando por los animales que no tienen piernas encontramos otra regla interesante:

```
[1] {hair=FALSE,feathers=FALSE,airborne=FALSE,legs=FALSE} => {domestic=FALSE} 0.2079208 0.9545455 1.09
leverage  phi      gini
[1] 0.01813548 0.1312 0.003860832
```

La mayoría de los animales sin pelo ni plumas, ni piernas que no vuelan no son domésticos con un 95% de confianza, un soporte del 20% y un lift de 1.09. Pero entonces hay algunos animales en los datos sin pelo ni plumas, ni piernas que no vuelan y que sí son domésticos. ¿Cuáles? ¿una iguana? ¿una serpiente? Lo buscamos:

```
which(Zoo$feathers==FALSE & Zoo$legs==FALSE & Zoo$hair==FALSE & Zoo$airborne==FALSE &
      Zoo$domestic==TRUE)
```

```
## [1] 8
```

```
Zoo[8,]
```

```
##      hair feathers eggs  milk airborne aquatic predator toothed backbone
## carp FALSE     FALSE TRUE FALSE     FALSE     TRUE     FALSE     TRUE     TRUE
##      breathes venomous fins  legs tail domestic catsize  type
## carp  FALSE     FALSE TRUE FALSE TRUE     TRUE     FALSE fish
```

y encontramos la carpa. Pues sí, gente que tiene carpas de mascota. Como a medida que se aumenta el tamaño de los itemsets con el filtro de no tener piernas no se generan reglas, pasamos a mirar los animales que sí tienen piernas. Buscamos itemsets de mínima longitud cinco, manteniendo el soporte mínimo en 0.2 y la confianza mínima en 0.9:

```
lhs                                     rhs          support  confidence
[1] {eggs=TRUE,milk=FALSE,airborne=TRUE,breathes=TRUE,legs=TRUE} => {toothed=FALSE} 0.2178218 1
[2] {eggs=TRUE,milk=FALSE,airborne=TRUE,fins=FALSE,legs=TRUE}   => {toothed=FALSE} 0.2178218 1
[3] {eggs=TRUE,airborne=TRUE,breathes=TRUE,fins=FALSE,legs=TRUE} => {toothed=FALSE} 0.2178218 1
[4] {milk=FALSE,airborne=TRUE,breathes=TRUE,fins=FALSE,legs=TRUE} => {toothed=FALSE} 0.2178218 1
hyperConfidence leverage  phi      gini
[1] 0.8382353      0.1315557 0.6516775 0.203162
[2] 0.8382353      0.1315557 0.6516775 0.203162
[3] 0.8382353      0.1315557 0.6516775 0.203162
[4] 0.8382353      0.1315557 0.6516775 0.203162
```

Pero solo encontramos reglas con confianza 100% que no nos dicen mucho nuevo.

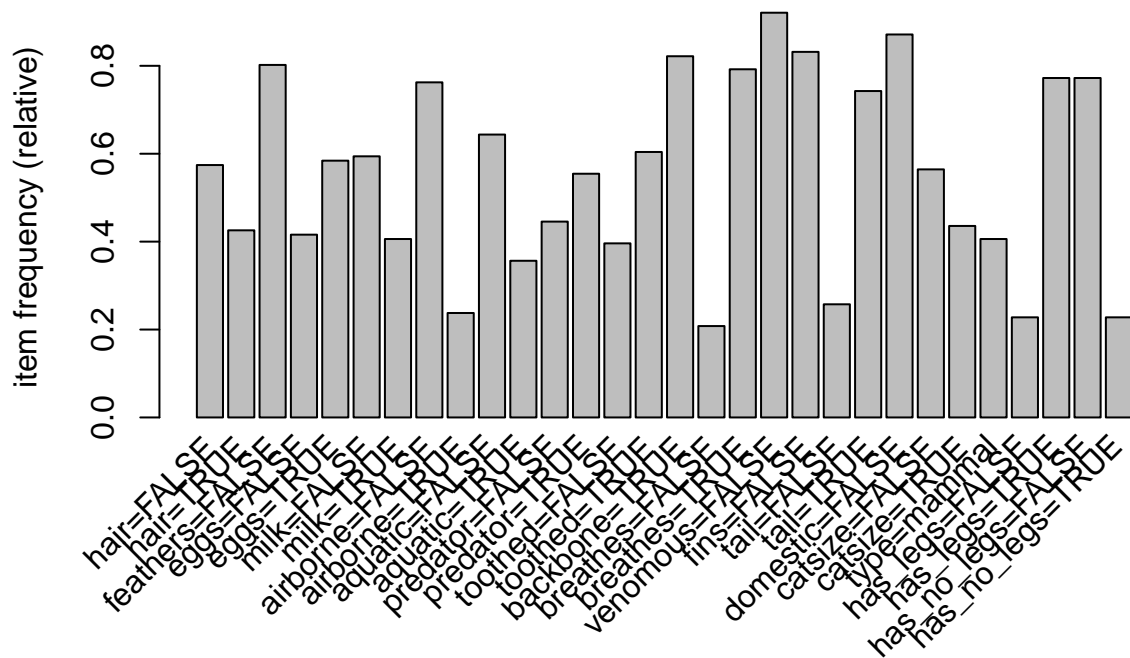
## Items negados

Se incluye experimentación con items negados tal como se pide en la práctica. Se ha escogido la variable “legs” factor creada anteriormente, creando dos variables a partir de ella (has\_legs y has\_no\_legs). Para ello se crea una variable como anteriormente y la otra como negación de la anterior:

```
has_legs <- Zoo[["legs"]]>0
has_no_legs<-!has_legs # creamos item negado
```

y ambas se añaden al dataset, sustituyendo la variable “legs” actual. Luego de nuevo pasamos a factores las variables, sacamos las transacciones y lanzamos Apriori con soporte=2 para buscar los items más frecuentes.

```
itemFrequencyPlot(zooNegs, support=0.2)
```



Cuando tenemos los itemsets frecuentes, los ordenamos por soporte, calculamos sus maximales y cerrados, y entrenamos de nuevo Apriori para obtener las reglas, usando soporte mínimo 0.2, confianza mínima 0.8 y longitud mínima 5. Tras inspeccionar reglas, filtrar y podar, obtuve:

```
#[2] {feathers=FALSE,aquatic=FALSE,predator=FALSE,has_legs=TRUE,has_no_legs=FALSE} => {hair=TRUE}
#[3] {feathers=FALSE,aquatic=FALSE,predator=FALSE,breathes=TRUE,has_legs=TRUE}      => {hair=TRUE}
#[4] {feathers=FALSE,aquatic=FALSE,predator=FALSE,fins=FALSE,has_legs=TRUE}        => {hair=TRUE}
#[5] {feathers=FALSE,aquatic=FALSE,predator=FALSE,breathes=TRUE,has_no_legs=FALSE} => {hair=TRUE}
#[6] {feathers=FALSE,aquatic=FALSE,predator=FALSE,fins=FALSE,has_no_legs=FALSE} => {hair=TRUE}
```

Reglas de hasta 5 items en el antecedente con un soporte del 22%, una confianza de 0.85 que decían que los animales sin plumas, que no son acuáticos ni depredadores y que tienen piernas, entonces tienen pelo. Ante estas reglas tan poco descriptivas, decidí bajar otra vez la longitud mínima a dos y subir el valor de lift a 2.3. Como parece que hay una relación entre tener patas y tener pelo y tener pelo y ser mamífero, esta vez decidí filtrar las reglas generadas por tipo mamífero con pelo y un lift mínimo de 2.3.

lhs	rhs	support	confidence	lift	count
[1] {hair=TRUE,type=mammal}	=> {milk=TRUE}	0.3861386	1.000000	2.463415	39
[2] {hair=TRUE,type=mammal}	=> {eggs=FALSE}	0.3762376	0.974359	2.343101	38
[3] {hair=TRUE,milk=TRUE,type=mammal}	=> {eggs=FALSE}	0.3762376	0.974359	2.343101	38
[4] {hair=TRUE,eggs=FALSE,type=mammal}	=> {milk=TRUE}	0.3762376	1.000000	2.463415	38
[5] {hair=TRUE,catsize=TRUE,type=mammal}	=> {milk=TRUE}	0.2970297	1.000000	2.463415	30

```
[6] {hair=TRUE,toothed=TRUE,type=mammal} => {milk=TRUE} 0.3762376 1.000000 2.463415 38
```

Algunas reglas, como la [4] parecen obvias. Otras, como la [5] son curiosas. Por ejemplo, la [5] indica que un mamífero con pelo de tamaño grande produce leche con un 100% de confianza. Hay varios mamíferos con pelo que no producen leche, así que esta regla puede indicar que el tamaño del mamífero en cuestión influye en que produzca o no leche. Pero entre ellas hay una regla especial. La regla [3] indica que un mamífero con pelo que produce leche no pone huevos **con un 0.97 de confianza**. Pero ¿entonces existe un mamífero con pelo que pone huevos y produce leche? Normalmente cuando un mamífero da leche es porque no pone huevos. Pero hay un animal que es una excepción, lo buscamos:

```
which(ZooNigs$type=="mammal" & ZooNigs$hair==TRUE & ZooNigs$eggs==TRUE)
```

```
## [1] 64
```

```
ZooNigs[64,]
```

```
##          hair feathers eggs milk airborne aquatic predator toothed
## platypus TRUE     FALSE TRUE TRUE  FALSE    TRUE      TRUE   FALSE
##          backbone breathes venomous  fins tail domestic catsize  type
## platypus      TRUE      TRUE   FALSE FALSE TRUE    FALSE    TRUE mammal
##          has_legs has_no_legs
## platypus      TRUE      FALSE
```

Y encontramos al animal, el platypus u ornitorrinco.