

Ejercicio Laboratorio: Descubrir región de dataset

Antonio Manuel Milán Jiménez

Una vez realizado el tratado de los datos, haber aplicado apriori y haber eliminado las reglas redundantes, obtenemos este tipo de reglas:

```
> inspect(rulesPruned)
```

	lhs	rhs	support	confidence	lift	count
[1]	{toothed=FALSE}	=> {fins=FALSE}	0.3960396	1.0000000	1.2023810	40
[2]	{milk=TRUE}	=> {type=mammal}	0.4059406	1.0000000	2.4634146	41
[3]	{milk=TRUE}	=> {breathes=TRUE}	0.4059406	1.0000000	1.2625000	41
[4]	{milk=TRUE}	=> {feathers=FALSE}	0.4059406	1.0000000	1.2469136	41
[5]	{milk=TRUE}	=> {backbone=TRUE}	0.4059406	1.0000000	1.2168675	41
[6]	{milk=TRUE}	=> {venomous=FALSE}	0.4059406	1.0000000	1.0860215	41
[7]	{type=mammal}	=> {breathes=TRUE}	0.4059406	1.0000000	1.2625000	41
[8]	{type=mammal}	=> {feathers=FALSE}	0.4059406	1.0000000	1.2469136	41
[9]	{type=mammal}	=> {backbone=TRUE}	0.4059406	1.0000000	1.2168675	41
[10]	{type=mammal}	=> {venomous=FALSE}	0.4059406	1.0000000	1.0860215	41
[11]	{eggs=FALSE}	=> {feathers=FALSE}	0.4158416	1.0000000	1.2469136	42

Sin embargo, en este problema buscamos algún caso especial que nos ayude a determinar la región de este dataset. Para encontrar esos casos especiales buscamos aquellas reglas que tengan una confianza muy cercana a 1 sin ser 1, de forma que exista alguna excepción que no haya cumplido la regla.

Eliminando así aquellas reglas que tengan una confianza de 1 obtenemos reglas de la forma:

```
> inspect(casosEspeciales)
```

	lhs	rhs	support	confidence	lift	count
[1]	{legs=TRUE}	=> {fins=FALSE}	0.7623762	0.9871795	1.1869658	77
[2]	{tail=TRUE}	=> {backbone=TRUE}	0.7326733	0.9866667	1.2006426	74
[3]	{backbone=TRUE,fins=FALSE}	=> {breathes=TRUE}	0.6435644	0.9848485	1.2433712	65
[4]	{aquatic=FALSE}	=> {breathes=TRUE}	0.6336634	0.9846154	1.2430769	64
[5]	{backbone=TRUE,legs=TRUE}	=> {venomous=FALSE}	0.6237624	0.9843750	1.0690524	63
[6]	{eggs=TRUE}	=> {milk=FALSE}	0.5742574	0.9830508	1.6548023	58
[7]	{fins=FALSE,tail=TRUE}	=> {breathes=TRUE}	0.5742574	0.9830508	1.2411017	58
[8]	{legs=TRUE,tail=TRUE}	=> {venomous=FALSE}	0.5544554	0.9824561	1.0669685	56
[9]	{aquatic=FALSE,backbone=TRUE}	=> {venomous=FALSE}	0.5148515	0.9811321	1.0655305	52
[10]	{feathers=FALSE,predator=TRUE}	=> {airborne=FALSE}	0.4554455	0.9787234	1.2837800	46
[11]	{catsize=TRUE}	=> {backbone=TRUE}	0.4257426	0.9772727	1.1892114	43
[12]	{catsize=TRUE}	=> {venomous=FALSE}	0.4257426	0.9772727	1.0613392	43
[13]	{toothed=TRUE,fins=FALSE}	=> {breathes=TRUE}	0.4257426	0.9772727	1.2338068	43
[14]	{hair=TRUE}	=> {legs=TRUE}	0.4158416	0.9767442	1.2647585	42
[15]	{eggs=FALSE}	=> {toothed=TRUE}	0.4059406	0.9761905	1.6163154	41

Estas reglas ya sí nos resultan más interesantes para nuestro objetivo. Aquí concretamente se han estudiado las reglas 1 y 8.

Al no tener una confianza de 1 podemos saber que hay algún animal en los datos que tenga patas y aletas, y algún otro animal que tenga patas, cola y sea venenoso. Buscando ya en Internet sobre estos casos descubrimos que el pinguino tiene patas y aletas, y el ornitorrinco cumple que tenga patas, cola y sea venenoso.

Si nos informamos sobre las regiones en las que viven estos dos animales encontramos una región común, el sur de Australia, por lo que esta será posiblemente la región que se ha estudiado para realizar este dataset.