

Árboles de decisión

N. Corral,
C. Carleos

18 de marzo de 2024

1 Ejemplo

2 Ejercicios

3 Bibliografía

Ejercicio 1: mtcars

- Construir un modelo de regresión lineal para estimar mpg a partir del resto de variables.
- Construir un árbol de regresión con el mismo objetivo.
- Comparar ambos modelos:
 - Con la información que producen las funciones de R.
 - Con el método de retención (entrenamiento y validación).
 - Con validación cruzada.

Ejercicio 2: iris

- Construye un árbol de clasificación para los datos `iris` a partir de todas sus variables (en la presentación usamos sólo las de los pétalos).
- ¿Se gana algo respecto a usar sólo los pétalos?
- Compáralo con el análisis dicriminante.

Iris: a toda mecha

Ejemplo

Ejercicios

Bibliografía

```
library(rpart)
rpart(Species~., iris)

## n= 150
##
## node), split, n, loss, yval, (yprob)
##      * denotes terminal node
##
## 1) root 150 100 setosa (0,3333 0,3333 0,3333)
##    2) Petal.Length< 2.45 50    0 setosa (1,0000 0,0000 0,0000)
##    3) Petal.Length>=2.45 100  50 versicolor (0,0000 0,0000 0,0000)
##      6) Petal.Width< 1.75 54    5 versicolor (0,0000 0,0000 0,0000)
##      7) Petal.Width>=1.75 46    1 virginica (0,0000 0,0000 0,0000)
```

Protocolo para árboles

- 1 Determinar el número adecuado de permutaciones para validación cruzada. Sean
 - n el tamaño muestral.
 - p el número de predictoresentonces podría bastar
 - $n < 1000 \cap p < 100 \Rightarrow xval = n$
 - $n > 10^4 \Rightarrow xval = 10$
- 2 Crear un árbol completo
 - $cp = 0$
 - $minsplit = 2$ ó $minbucket = 1$
 - $xval = 10$ ó n
- 3 Fijarse en si $xstd$ (errores típicos obtenidos por cruzvalidación) son reducidos. Si no, aumentar $xval$
- 4 Buscar el parámetro de complejidad (cp) adecuado
`plotcp (arbol)`
- 5 Poder el árbol
`arbol1 <- prune (arbol)`

Iris: protocolo para árboles

```
a <- rpart(Species~., iris,  
           control=rpart.control(xval=150,  
                                  cp=0,  
                                  minsplit=2))
```

printcp(a)

plotcp(a)

```
a1 <- prune(a, 0.01)
```

Más detalles

- https://es.wikipedia.org/wiki/Aprendizaje_basado_en_%C3%A1rboles_de_decisi%C3%B3n
- <https://cran.r-project.org/web/packages/rpart/vignettes/longintro.pdf>
- https://www.researchgate.net/publication/263671703_Fifty_Years_of_Classification_and_Regression_Trees