



Dpto. Lenguajes y Ciencias de la Computación
E.T.S.I. en Informática, Universidad de Málaga

Aprendizaje Computacional

24 de Enero de 2018

Apellidos:
DNI:

Nombre:

PARTE 3

1. Un ensemble consiste en entrenar una serie de clasificadores pobres (por ejemplo, árboles de decisión o perceptron multicapa) con un subconjunto escogido aleatoriamente del dataset. Se realiza una predicción con cada uno de los clasificadores y se obtiene como salida del ensemble la clase más votada (es decir, la clase con mayor número de ocurrencias).
 - a) Realiza un programa en R que implemente el algoritmo descrito usando perceptron multicapa (busca el número más adecuado de neuronas en la capa oculta). Usa el dataset de la recidiva.
 - b) Modifica el programa anterior usando árboles de decisión como el clasificador pobre. Usa el mismo dataset que en el apartado anterior.
 - c) Compara ambos algoritmos. ¿Por cuál de ellos te decidirías?

Sugerencia:

Para poder manejar una lista de clasificadores puedes usar el código siguiente:

```
library(rpart)
df<-data.frame(x=c(1,2,3,3,3), y=factor(c("a", "a", "b", "a",
"b")), z=c(5,4,4,2,5))

mytree1<-rpart(y ~ x+z , data = df, minbucket = 1, minsplit=1)
mytree2<-rpart(y ~ x+z , data = df, minbucket = 1, minsplit=2)
mytree3<-rpart(y ~ x+z , data = df, minbucket = 1, minsplit=3)

lista1<- list(mytree1, mytree2)
lista2<- list(mytree3)
w <- c(lista1,lista2)
w[[1]]
plot(w[[3]])
```

Dicho código crea tres clasificadores y dos listas que concatena en una única lista. `w[[1]]` hace referencia a `mytree1` y `plot(w[[3]])` dibuja el árbol `mytree3`.

2. Usa el comando *importance* de **Ramdon Forest**. para determinar la importancia de los atributos en un dataset (puedes usar cualquiera que prefieras o crear uno nuevo), Dibuja también el grafico. Explica que significa que una variable tenga mayor importancia que otra.

3.

- a) Explica el siguiente programa:

```
x1=rnorm(1000)
x2=rnorm(1000)
```



Dpto. Lenguajes y Ciencias de la Computación
E.T.S.I. en Informática, Universidad de Málaga

Aprendizaje Computacional *24 de Enero de 2018*

Apellidos:
DNI:

Nombre:

```
y=2*x1+.7*x2+rnorm(1000)
df=data.frame(y,x1,x2,x3=rnorm(1000),x4=rnorm(1000),x5=rnorm(1000)
)

# run the randomForest implementation
library(randomForest)
rfl <- randomForest(y~., data=df, mtry=2, ntree=50,
importance=TRUE)
importance(rfl,type=1)
```

- b) Modifica el programa de forma que la importancia de las variables sea equiprobable (es decir, que cambie dependiendo de la ejecución).