

Aprendizaje Computacional

19 de Enero de 2022

Apellidos:
DNI:

Nombre:

PARTE 2

1. a) Teniendo en cuenta el dataset que se muestra en la Figura 1, calcula los pesos α según el algoritmo de boosting. Realiza dos separaciones que minimicen el número de puntos mal clasificados.

b) Predice los siguientes puntos marcados con círculos rojos (Figura 2):

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

Aprendizaje Computacional

19 de Enero de 2022

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

Apellidos:
DNI:

Nombre:

3. El conjunto de datos Kyphosis (disponible con el paquete Rpart) nos indica si, tras una operación en la columna vertebral, una muestra de 81 niños presenta o no deformaciones en la columna vertebral (columna **Kyphosis**). El resto de columnas son:

Age : Año en meses.

Number: Numero de la vértebra involucrada.

Start : Numero de la primera vértebra involucrada

Realiza un ensemble con dos árboles Rpart en el primer nivel y votación en el segundo. Usa arboles con un solo nivel. El primer árbol se entrena con los atributos **Age** y **Number** mientras que el otro se entrena con los atributos **Age** y **Start**. Entrena los arboles usando validación cruzada. De la parte reservada para test se obtiene aleatoriamente un 80% para entrenar un árbol rpart y 80% con índices distintos para el otro.

Compara el accuracy del ensemble anterior con un ensemble AdaBoost..