Practica 2. CART Spliting.

CART es un algoritmo que genera arboles de decisión binarios a partir de un conjunto de entrenamiento (dataset). Dicho conjunto esta constituido por un conjunto de características (columnas del dataset, también denominadas variables o campos). El algoritmo parte de un nodo raíz que debemos etiquetar con una variable y un valor de la misma tal que el conjunto de datos se divida en dos regiones en las cuales la suma de la varianza de cada región sea mínimo. Mas formalmente, debemos encontrar $\mathbf{X}_i$ y **s** que divida el conjunto de datos en dos regiones R1y R2 tal que:

$$R_1(j,s) = \{X|X_j < s\} \text{ and } R_2(j,s) = \{X|X_j \geq s\},$$

tal que, se minimice la siguiente suma:

$$\sum_{i:\ x_i \in R_1(j,s)} (y_i - \hat{y}_{R_1})^2 + \sum_{i:\ x_i \in R_2(j,s)} (y_i - \hat{y}_{R_2})^2,$$

$\hat{y}_{R_1}$ es la media de las respuestas de la variable de decisión en la Región **R1**.

$\hat{y}_{R_2}$ es la media de las respuestas de la variable de decisión en la Región **R2**.

Por tanto, el algoritmo consiste en iterar por el conjunto de variables y por cada valor de la misma. Evaluar la expresión anterior y determinar la mínima. Con el par variable ($\mathbf{X}_i$) valor (**s**) encontrado dividiremos el nodo en dos regiones: a la izquierda la región **R1** y a la derecha la región **R2** y etiquetaremos el nodo con **Xi<s**.
De esta forma se obtiene el nodo raíz y la primera división del árbol.
a) El primer objetivo de la practica es teniendo en cuenta el algoritmo anterior realizar la primera división del árbol y encontrar la variable y el valor de dicha división. El algoritmo mostrara la variable y valor mínima. Ademas, tendrá un versión *Verbose* que mostrara para cada par variable/valor la cantidad que minimiza.
El algoritmo continua de forma recursiva por cada región encontrada.
b) El segundo objetivo del algoritmo es determinar dos árboles **T1** y **T2** de dos niveles tal que **T1** explote la región derecha (es decir **R2**) con el algoritmo descrito anteriormente, mientra que **T2** explota la izquierda . Por ultimo, predecimos el conjunto completo de resultados usando ambos arboles. Nos quedaremos con el árbol que mejor predicción realice. La salida sera:

Arbol 1.

Nivel 0
- Nodo 1(o raiz): $\mathbf{X}_i$<**s1**

Nivel1
- Derecha
  - Nodo2: $\mathbf{X}_j$<**s2**
  - etiqueta nodo derecha - etiqueta nodo izquierda
- Izquierda
  - etiqueta nodo

Arbol 2.

Nivel 0
- Nodo 1(o raíz): $\mathbf{X}_i$<**s1**

Nivel1
- Derecha
  - etiqueta nodo
- Izquierda
  - Nodo2: $\mathbf{X}_j$<**s2**

etiqueta nodo derecha - etiqueta nodo izquierda

Predicción.

T1: accuracy del árbol **T1**

T2: accuracy del árbol **T2**

c) Explica el criterio que seguirías pasa finalizar el árbol. Pon un ejemplo.

Nota: Para realizar la pruebas podemos usar los siguientes dataset:

- Hitters que se encuentra en el paquete ISLR.
- Recidiva que podemos encontrarlo en el campus virtual.
- Usado en clase:

	X1	X2	Y
1	2.771244718	1.784783929	0
2	1.728571309	1.169761413	0
3	3.678319846	2.81281357	0
4	3.961043357	2.61995032	0
5	2.999208922	2.209014212	0
6	7.497545867	3.162953546	1
7	9.00220326	3.339047188	1
8	7.444542326	0.476683375	1
9	10.12493903	3.234550982	1
10	6.642287351	3.319983761	1