

FACULTAD DE CIENCIAS EXACTAS, INGENIERÍA Y AGRIMENSURA

INTRODUCCIÓN A LA INTELIGENCIA ARTIFICIAL

---

# TRABAJO PRÁCTICO 1: ÁRBOLES DE DECISIÓN

---

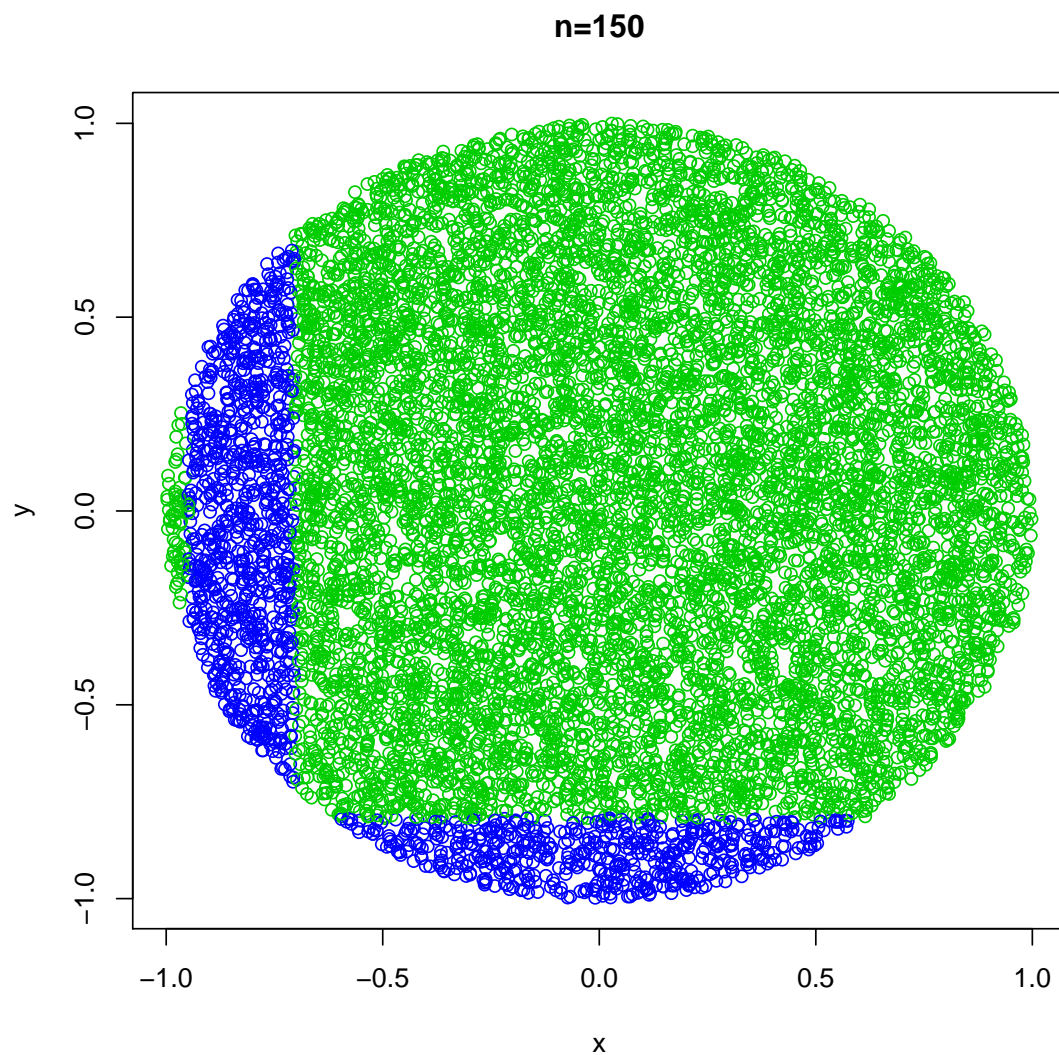
*Alumno: Rodríguez Jeremías*

7 de mayo de 2017

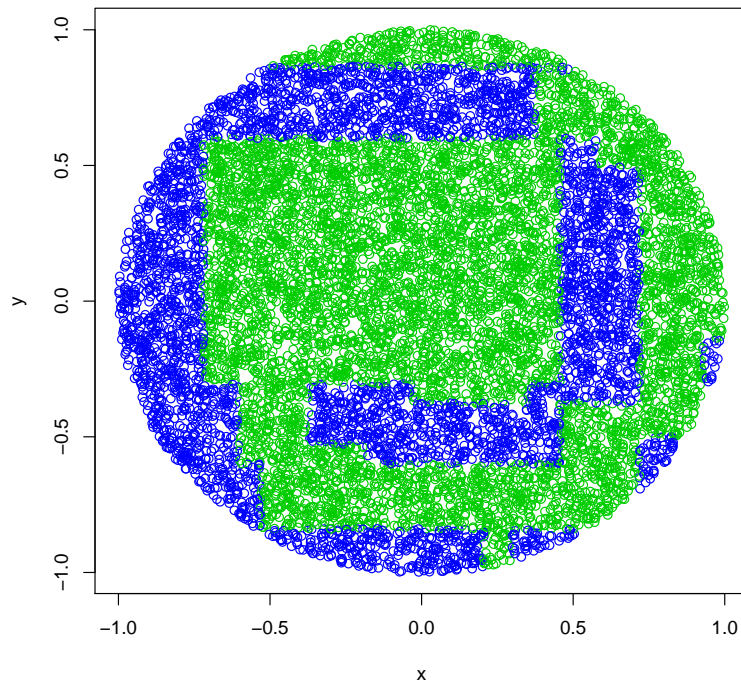
## 1. Ejercicio 4: Espirales

En las siguientes tres imágenes vemos la predicción de c4.5 sobre las espirales de la práctica 0 para distintos tamaños de conjunto de entrenamiento:  $n=150$ , 600 y 3000. Se clasifica un conjunto de test de  $n=10000$ .

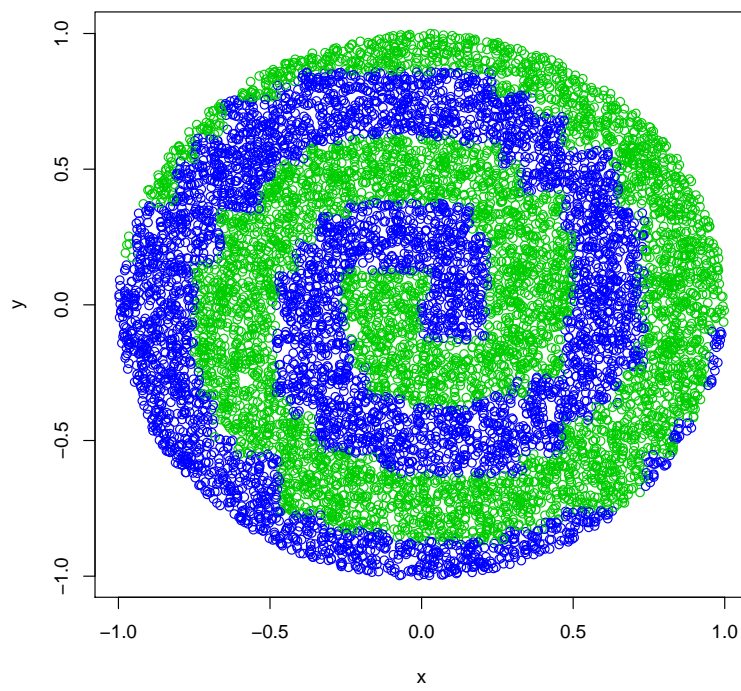
Se aprecia que, si el conjunto de entrenamiento no es suficientemente grande, el resultado es muy inexacto y el árbol es bastante básico; pero que si ampliamos un poco más el tamaño del conjunto de entrenamiento las predicciones son cada vez más exactas. Para  $n=3000$  la predicción es muy buena.



**n=600**



**n=3000**

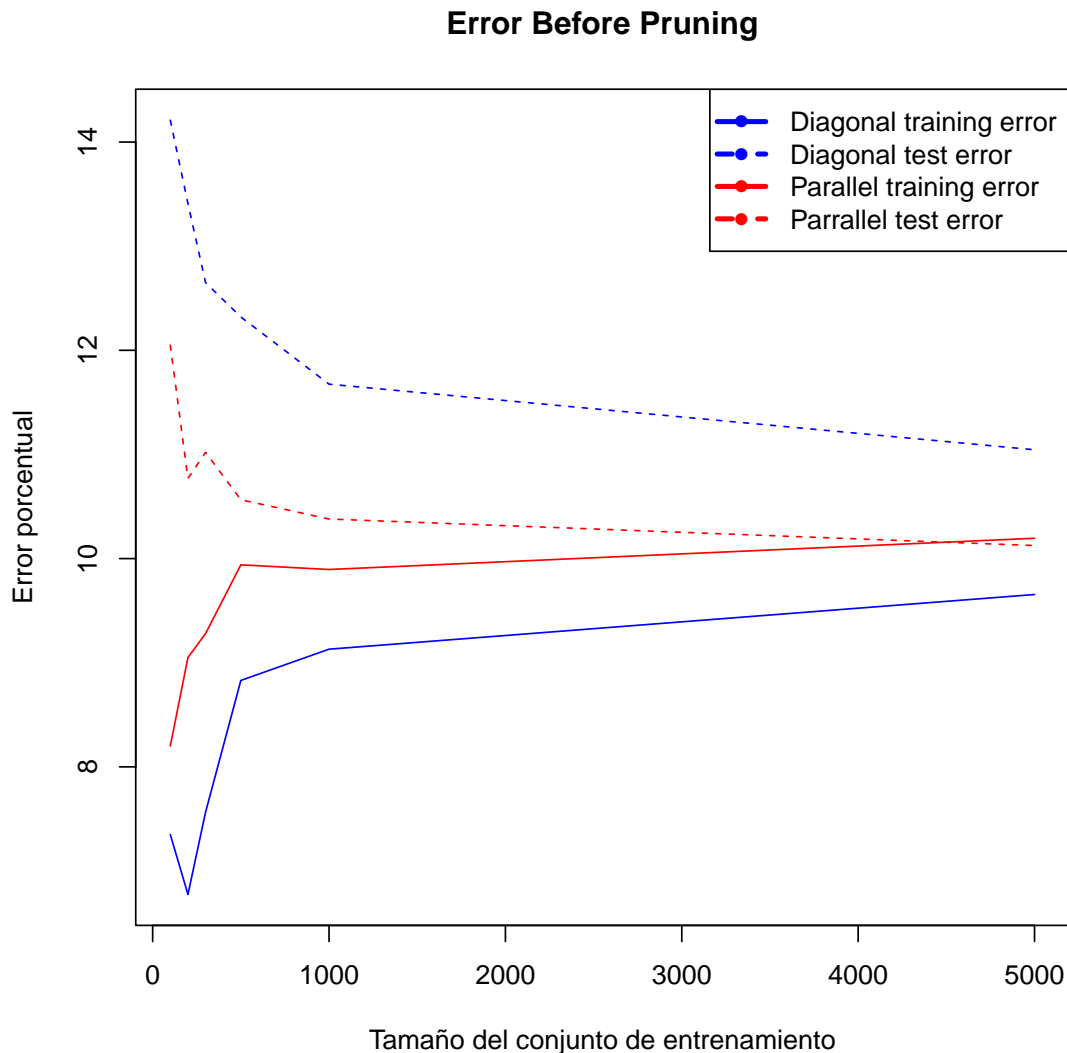


## 2. Ejercicio 5: Dependencia con la longitud del conjunto de entrenamiento - Sobreajuste y pruning

### 2.1. Errores

En los próximos ejercicios se usa el generador de datos diagonal y el paralelo del trabajo anterior. Recordemos que estos generadores, creaban datasets de tamaño  $n$  con una dispersión de datos  $c$  y de  $d$  dimensiones.

En este ejercicio analizaremos cómo se comporta c4.5 cuando utilizamos conjuntos de entrenamiento con  $n$  cada vez mayor. La siguiente gráfica muestra como varía el error porcentual (tanto en el caso diagonal como paralelo) de la predicción de c4.5 sobre un conjunto de test, en el árbol before pruning:



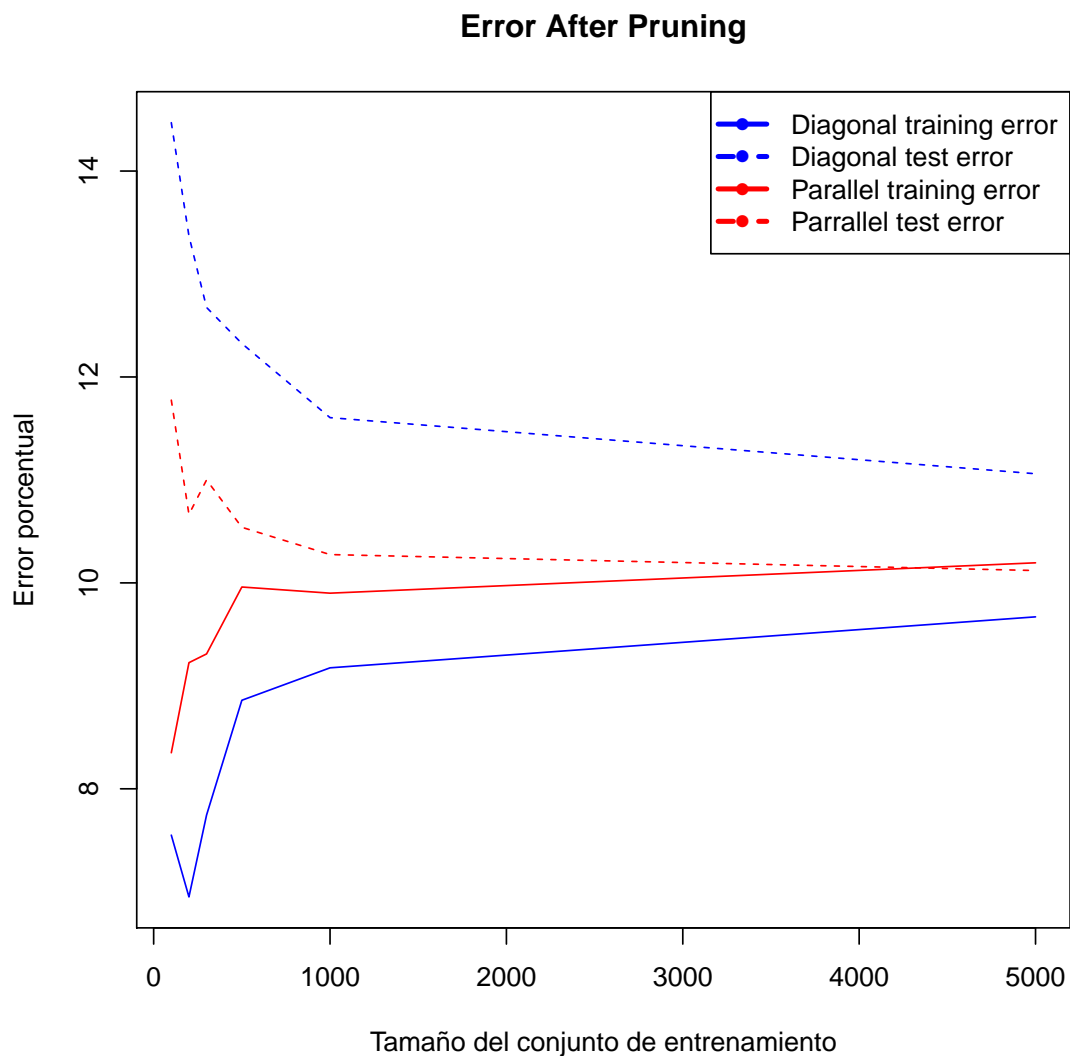
Observo que, si tomamos conjuntos de entrenamiento cada vez más grandes, c4.5 clasificará cada vez mejor a los puntos de conjuntos de test. Sin embargo, la curva parece llegar a un

comportamiento asintótico que indica que, llegado un cierto tamaño de  $n$ , el aumento de  $n$  no repercute en una reducción significativa del error. Es decir, tampoco es conviene usar conjuntos de entrenamiento inmensos.

Por otro lado, observo que en los ejemplos de training, el error va aumentando a medida que el conjunto de entrenamiento es mejor. Esto se da porque es más difícil para un árbol "fitear" todos los datos del conjunto de training cada vez más grande.

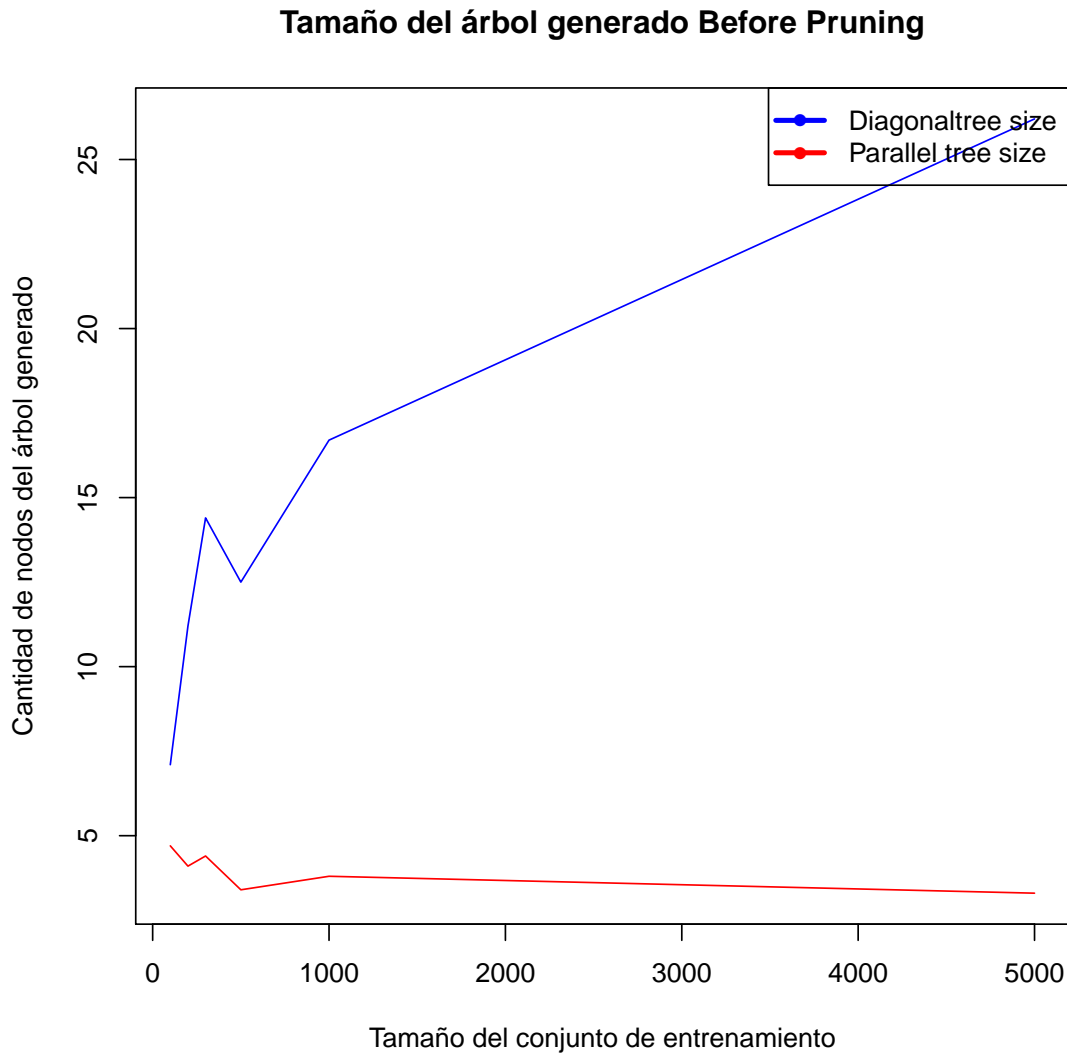
Una última cosa a apreciar es que, más allá de una diferencia pequeña, las curvas correspondientes al problema paralelo y al diagonal tienen la misma forma y ambos problemas parecen comportarse de forma muy parecida.

La siguiente gráfica muestra los mismos datos pero sobre el árbol after pruning. Como vemos, el pruning no muestra una gran diferencia con la gráfica before pruning. Esto sucede porque c4.5 casi no está podando nodos, probablemente porque considera que toda la información que aportan los nodos es valiosa y no aumenta la precisión del árbol eliminarlos. Tal vez para valores de  $n$  mayores, si considere más beneficioso podar nodos:



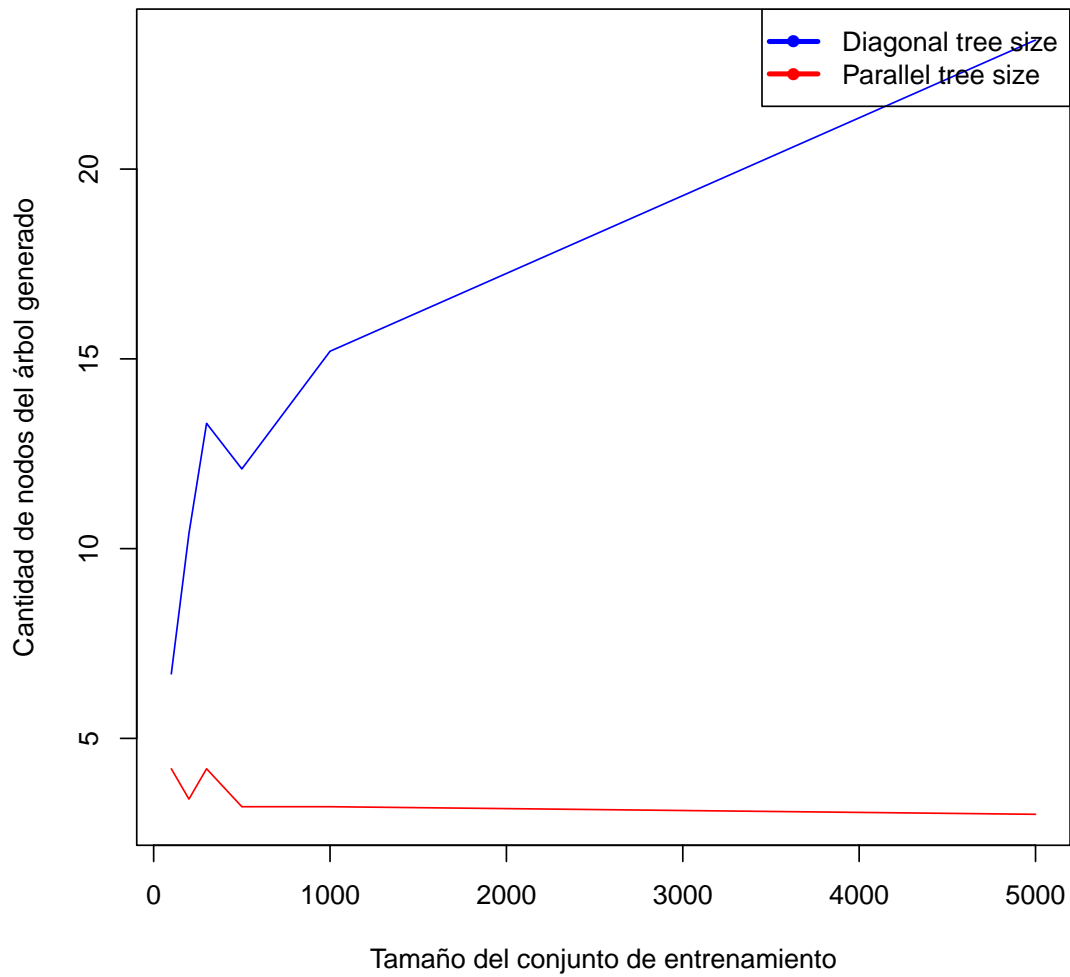
## 2.2. Tamaño del árbol resultante

Ahora analizaremos cómo varía el tamaño del árbol resultante cuando  $n$  varía:



En esta primera gráfica (before pruning) podemos observar que el árbol resultante se vuelve cada vez más complejo para el problema diagonal, si proveemos conjuntos de entrenamiento grandes. Por otro lado, el tamaño del árbol del problema paralelo permanece constante. Esto podría indicar que el problema paralelo es más "sencillo" que el problema diagonal, ya que su solución puede expresarse con un árbol muy sencillo y no necesita agregar más nodos para aumentar su precisión. De hecho, vimos en las primeras dos gráficas de este enunciado que el error porcentual sobre el conjunto de test era ligeramente menos para el caso paralelo.

**Tamaño del árbol generado After Pruning**

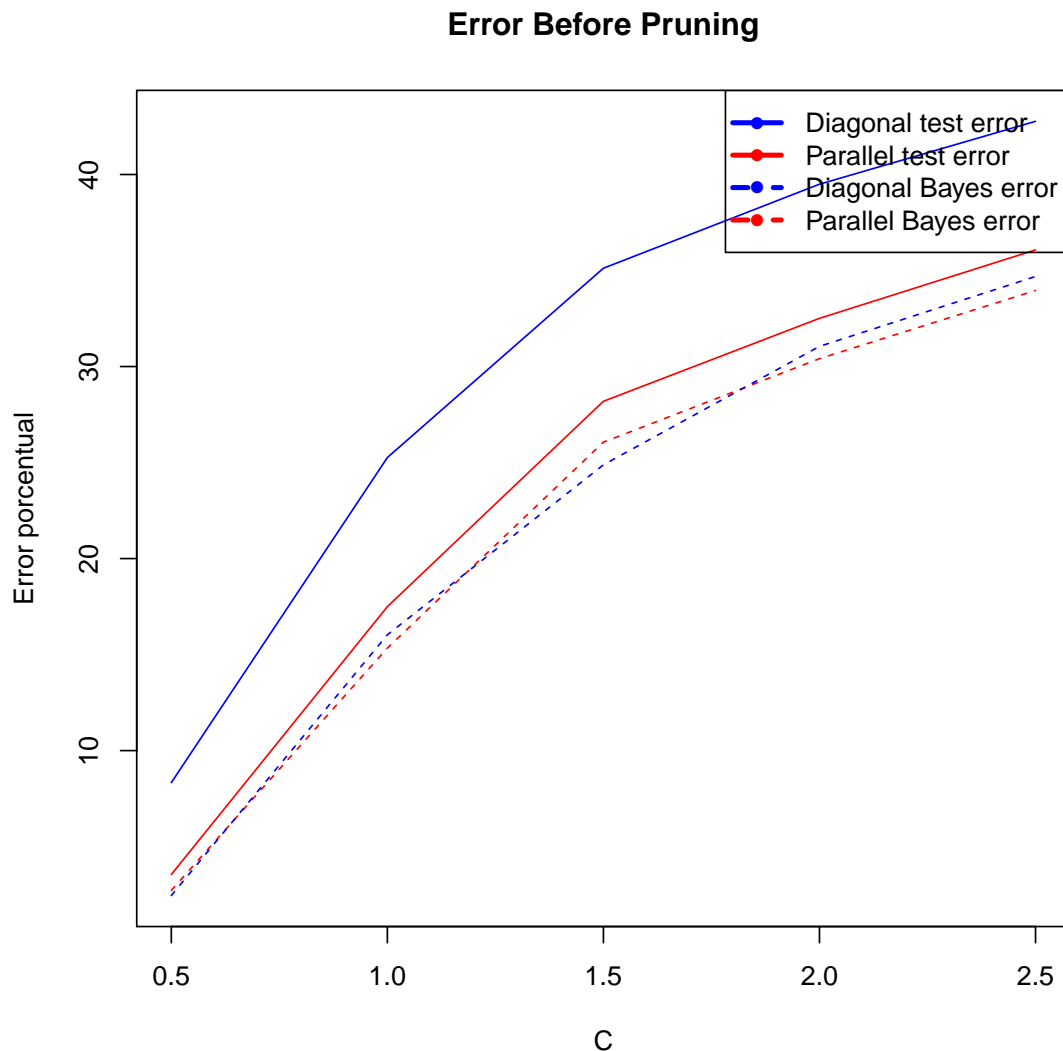


Nuevamente, el gráfico after pruning no muestra muchos cambios; solamente se aprecia que en el caso diagonal se podan algunos (pocos) nodos para tamaños de  $n$  grandes. Esto sucede porque no hay gran sobreajuste de los datos en el árbol antes de prunear.

### 3. Ejercicio 6: Resistencia al ruido

En este ejercicio haremos algo parecido al anterior, analizaremos el error cuando hacemos variar  $c$ . Recordemos que a mayor  $C$ , más mezclados estarán los datos (ruido).

En la siguiente gráfica se ve el error en función de  $C$ . También, para cada  $C$ , incluyo el error del clasificador ideal (Bayes). Este error lo calculé utilizando un conjunto de test grande, y viendo cuántas veces (porcentualmente) se equivoca este clasificador al clasificar los puntos del test. El clasificador elige siempre aquella clasificación con mayor probabilidad. Para seguir este criterio, calculé la distancia de cada punto al centro de las dos distribuciones normales y clasifiqué el punto con la clase cuyo centro es el más cercano.

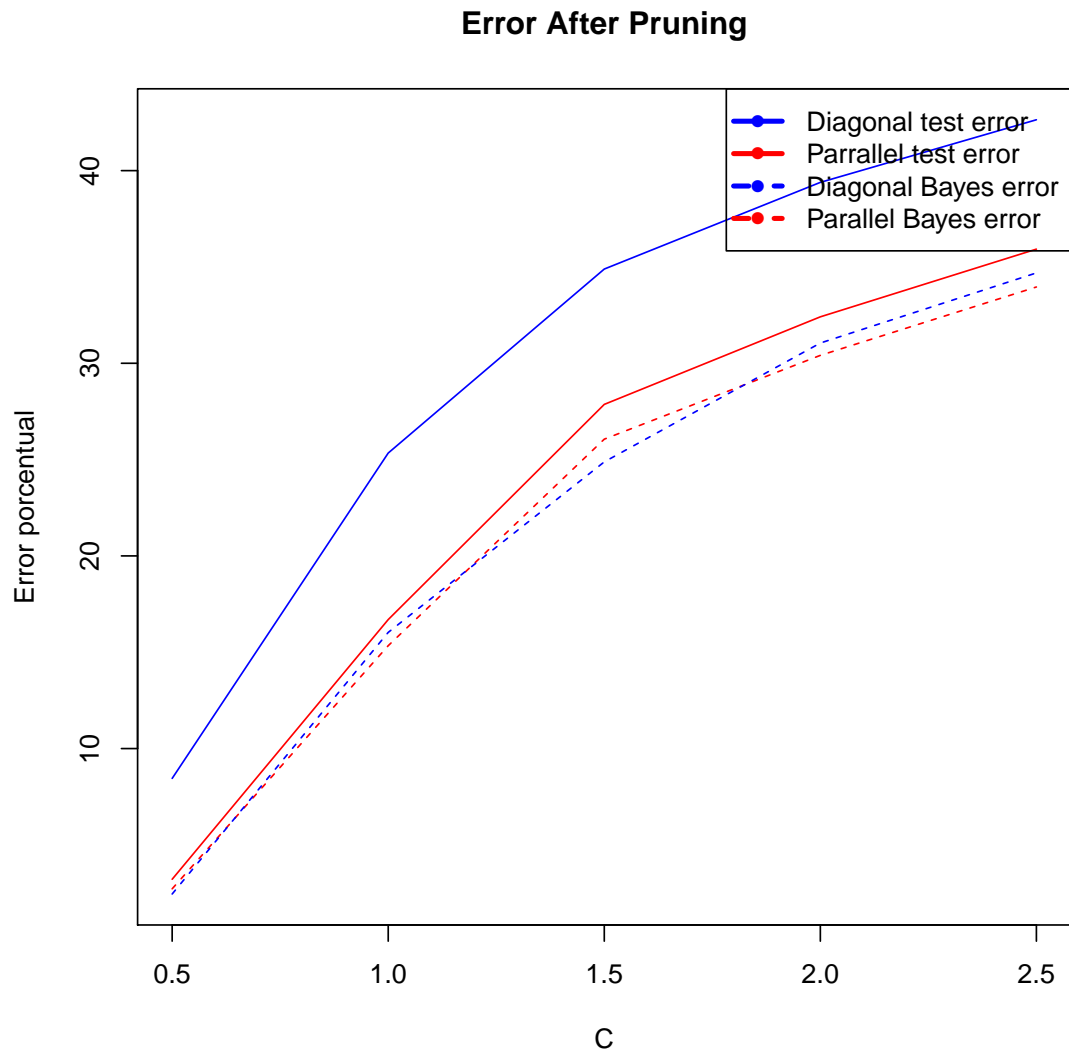


A mayor ruido, naturalmente, es mayor la cantidad de error que producen las predicciones. Sin embargo, si comparamos con el clasificador ideal, vemos que los crecimientos (las formas) de las



curvas son similares. Es decir, el error de c4.5 no crece de una forma descontrolada; sino que se comporta razonablemente bien frente al ruido. El caso paralelo es ligeramente mejor porque, como ya vimos, es un problema un poco más fácil que el diagonal. De todos modos, vemos que para el caso paralelo como el diagonal c4.5 se comporta forma similar frente al ruido.

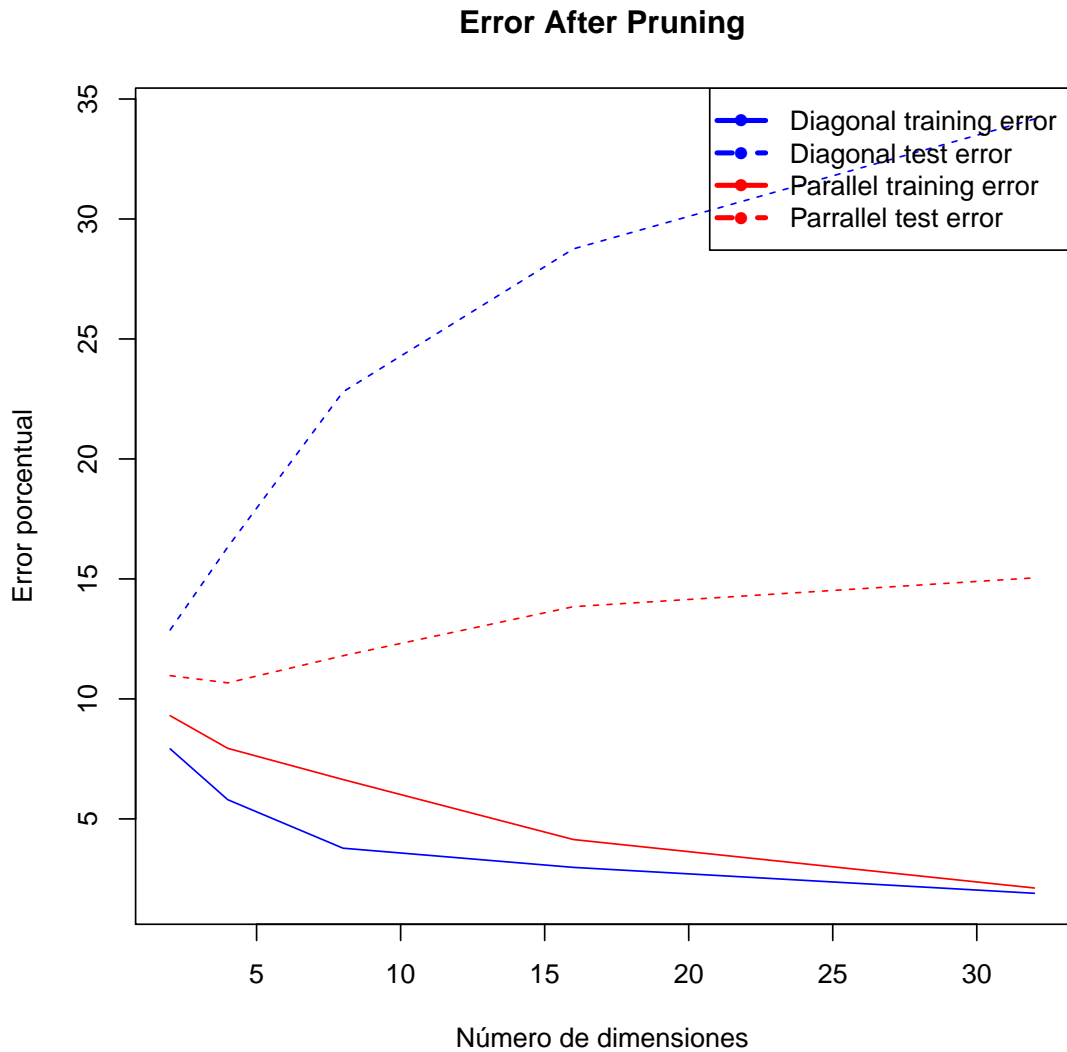
Una conclusión de esta gráfica entonces es que, en general, los árboles de decisión son robustos frente al ruido.



No hay diferencias perceptibles entre el caso before pruning y el after pruning.

## 4. Dimensionalidad

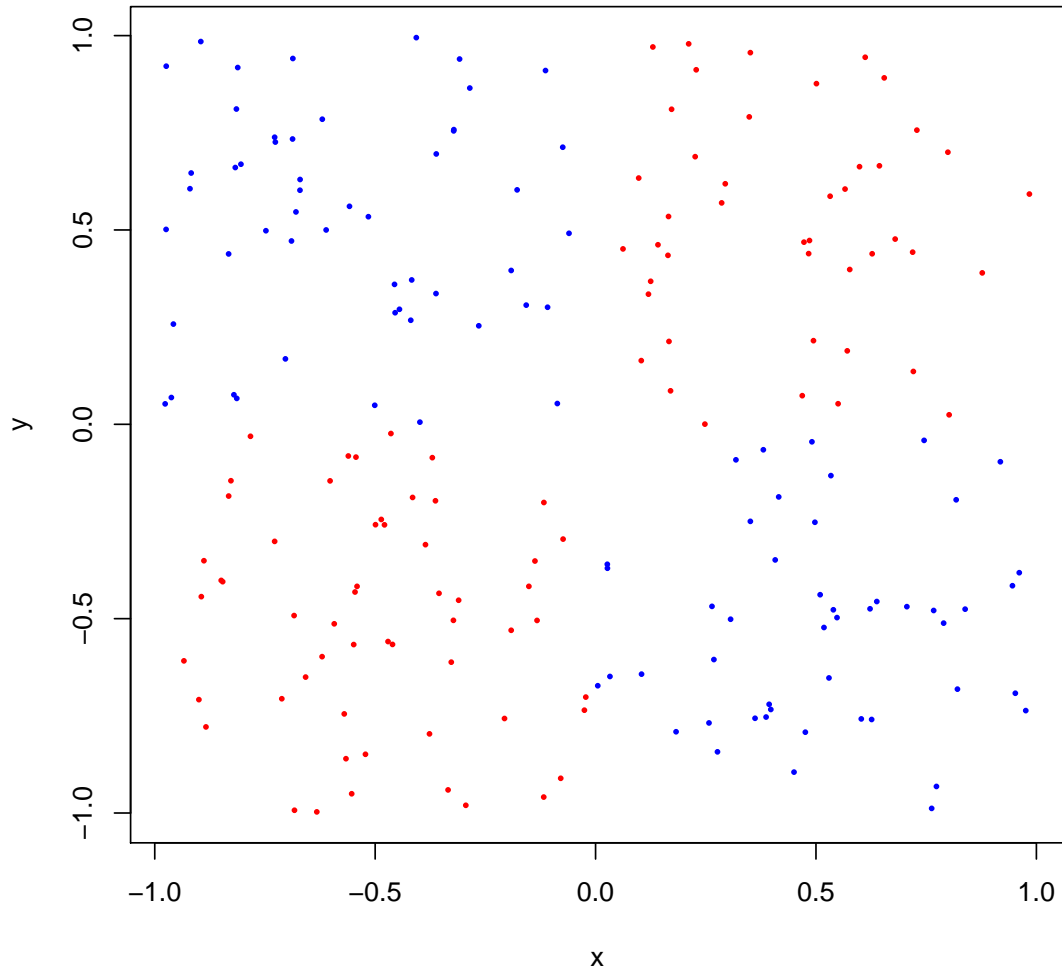
Finalmente, en este ejercicio analizaremos qué sucede con el error cuando variamos el parámetro  $d$  (cantidad de dimensiones):



Como podemos observar, cuando aumenta el número de dimensiones el error en el conjunto de training es cada vez menos; pero en el conjunto de test aumenta significativamente. Esto es un ejemplo de sobreajuste: El árbol generado es cada vez más complejo para ajustarse exactamente a los datos de training (y reduciendo su error); pero al extrapolarlos a un conjunto de test mayor no produce buenos resultados.

Nuevamente, vemos que el problema diagonal presenta un error mucho mayor, consecuencia de su naturaleza más compleja que el paralelo (recordemos que la dispersión de error del problema diagonal es proporcional a  $d$ ).

## 5. Problema XOR



El problema xor puede expresarse muy sencillamente con un árbol de decisión pequeño:

If  $X \geq 0 \cup Y \geq 0$  then clase 0  
If  $X \leq 0 \cup Y \leq 0$  then clase 0  
If  $X < 0 \cup Y > 0$  then clase 1  
If  $X > 0 \cup Y < 0$  then clase 1

Sin embargo, C4.5 es incapaz de generar este árbol, y clasifica a todos los puntos con un árbol constante. (A todos les asigna clase 0).

Decision Tree:  
0 (200.0/100.0)

Esto es consecuencia de que la creación del árbol se detiene prematuramente, dado que ningún

atributo individualmente exhibe una asociación significativa a la clase. (Information gain es 0 para ambos atributos)