



UNIVERSIDAD DE GUADALAJARA

CENTRO UNIVERSITARIO DE CIENCIAS EXACTAS E INGENIERÍA
DEPARTAMENTO DE CIENCIAS COMPUTACIONALES

Seminario de Solución de Problemas de Sistemas Basados en Conocimiento

Práctica No. 7

Nombre: Hurtado González Edgar Arturo
Código: 212597894

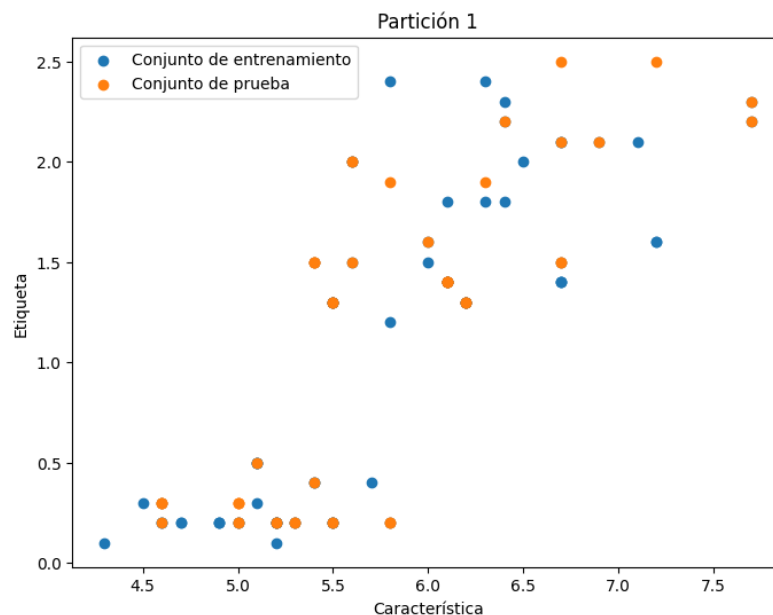
5 formas de Partición de la Información

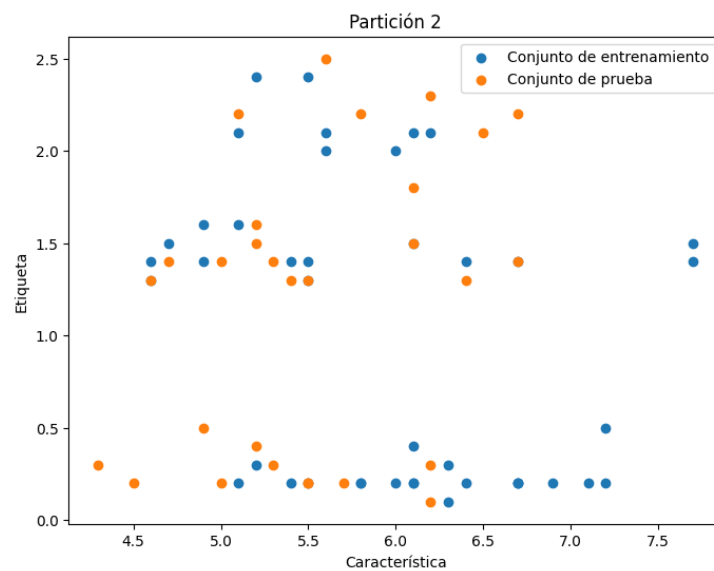
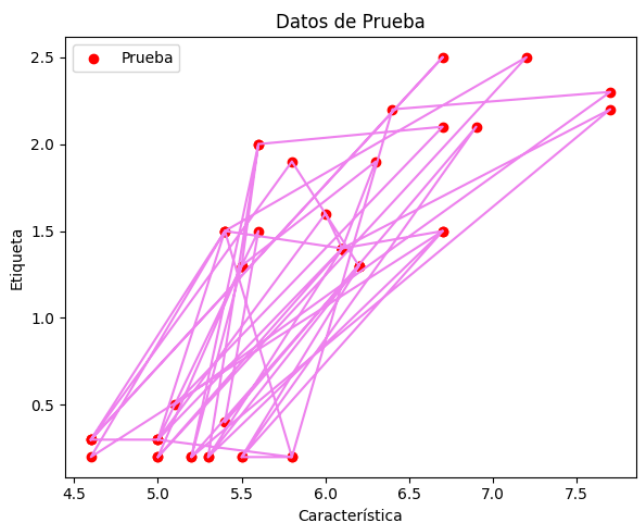
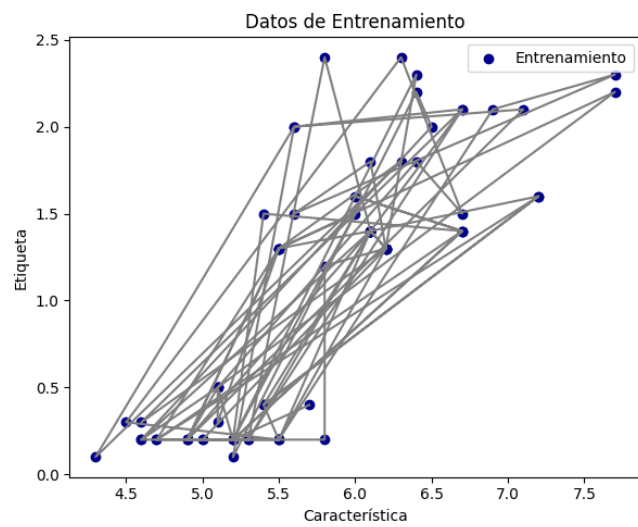
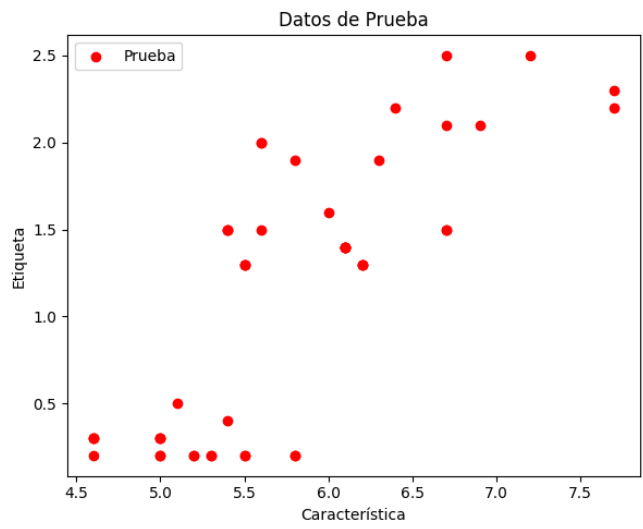
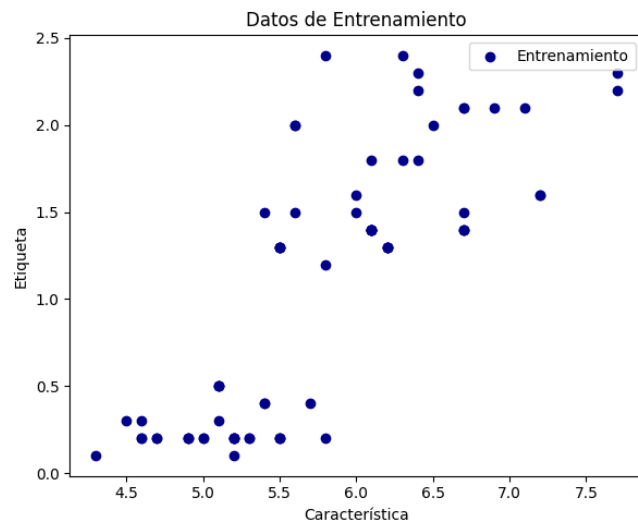
En este código lo que se realiza son 5 diferentes formas de Particionar Información, esto es interesante ya que para aplicación de diferentes métodos para poder manipular los datos de dicha información muchas veces es necesario el particionarla correctamente para poder aplicar dicha metodología de forma precisa.

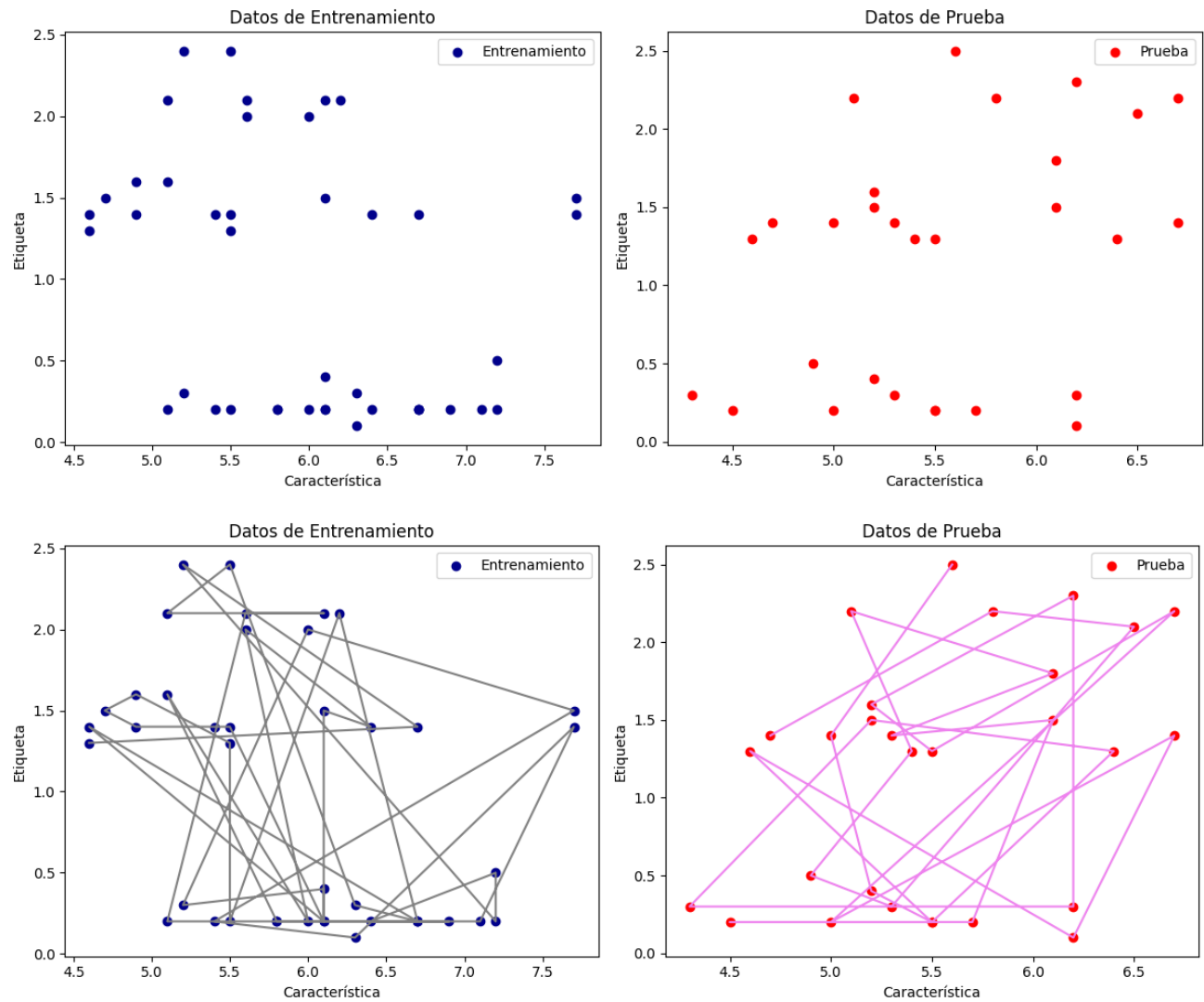
Los 5 métodos con los que he realizado esta Partición de la información son:

****Partición Aleatoria:****

> Este método implica dividir aleatoriamente el conjunto de datos en conjuntos de entrenamiento y prueba. La división se realiza de forma aleatoria y puede variar cada vez que se realiza el proceso de partición.

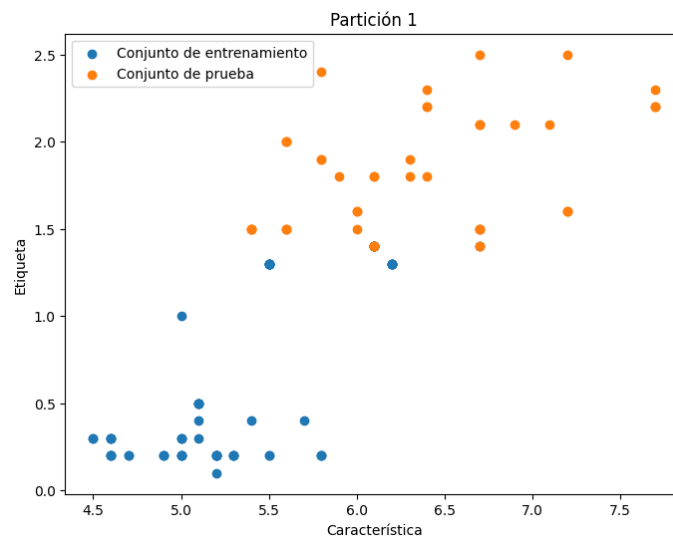


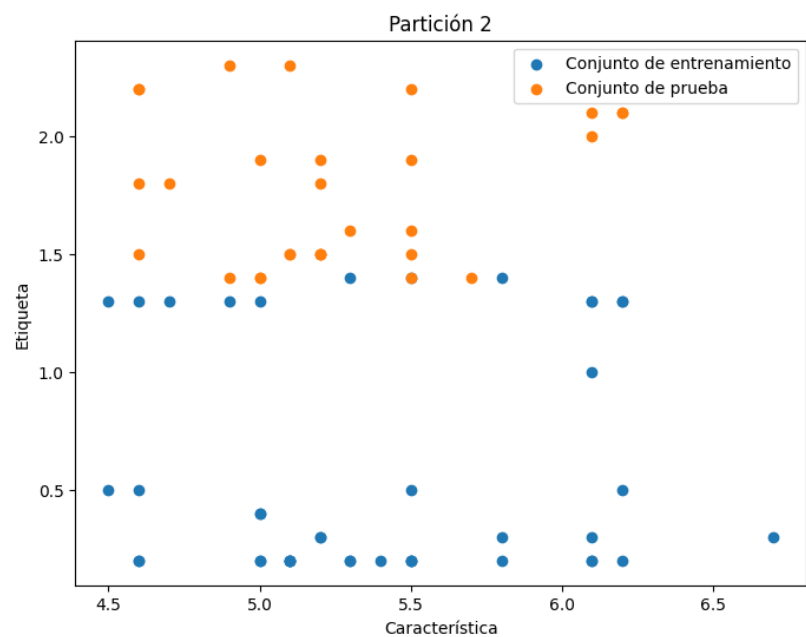
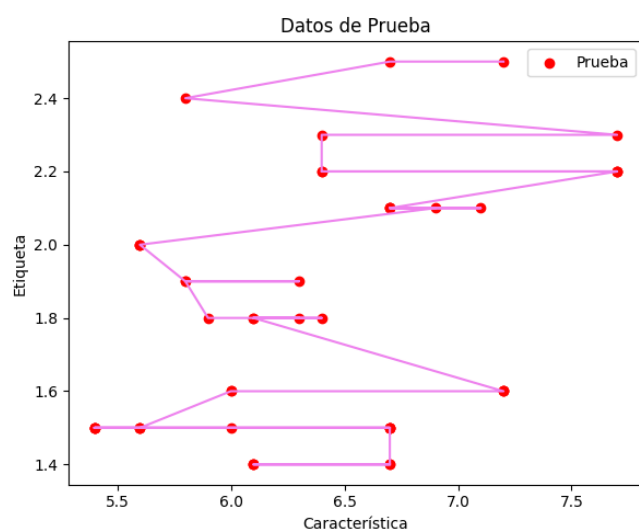
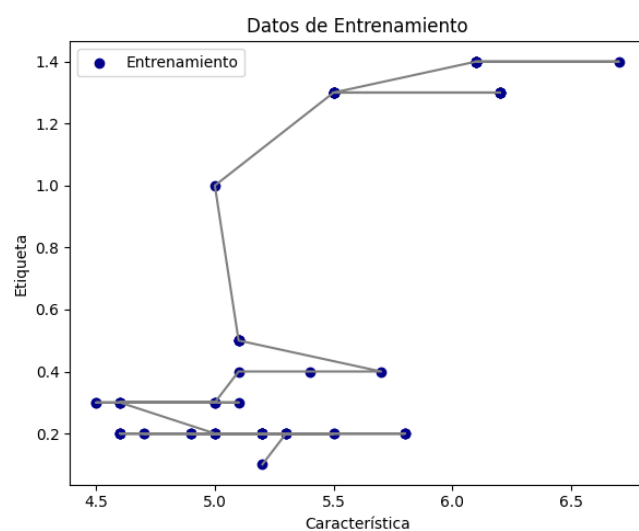
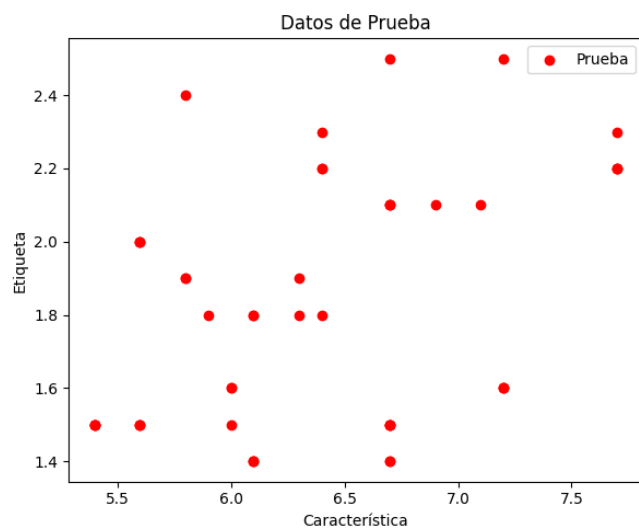
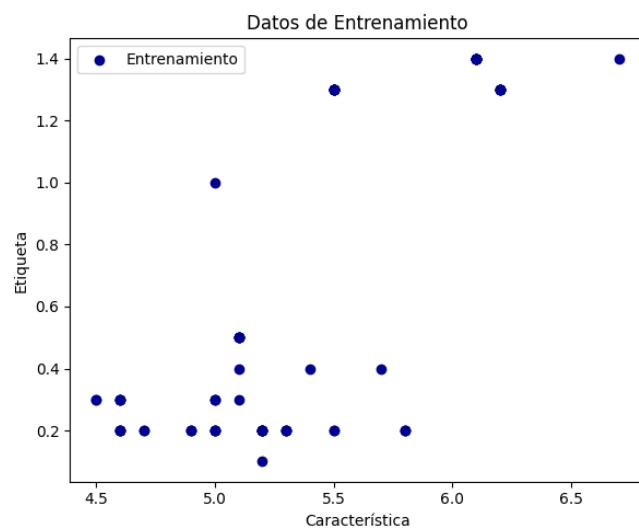


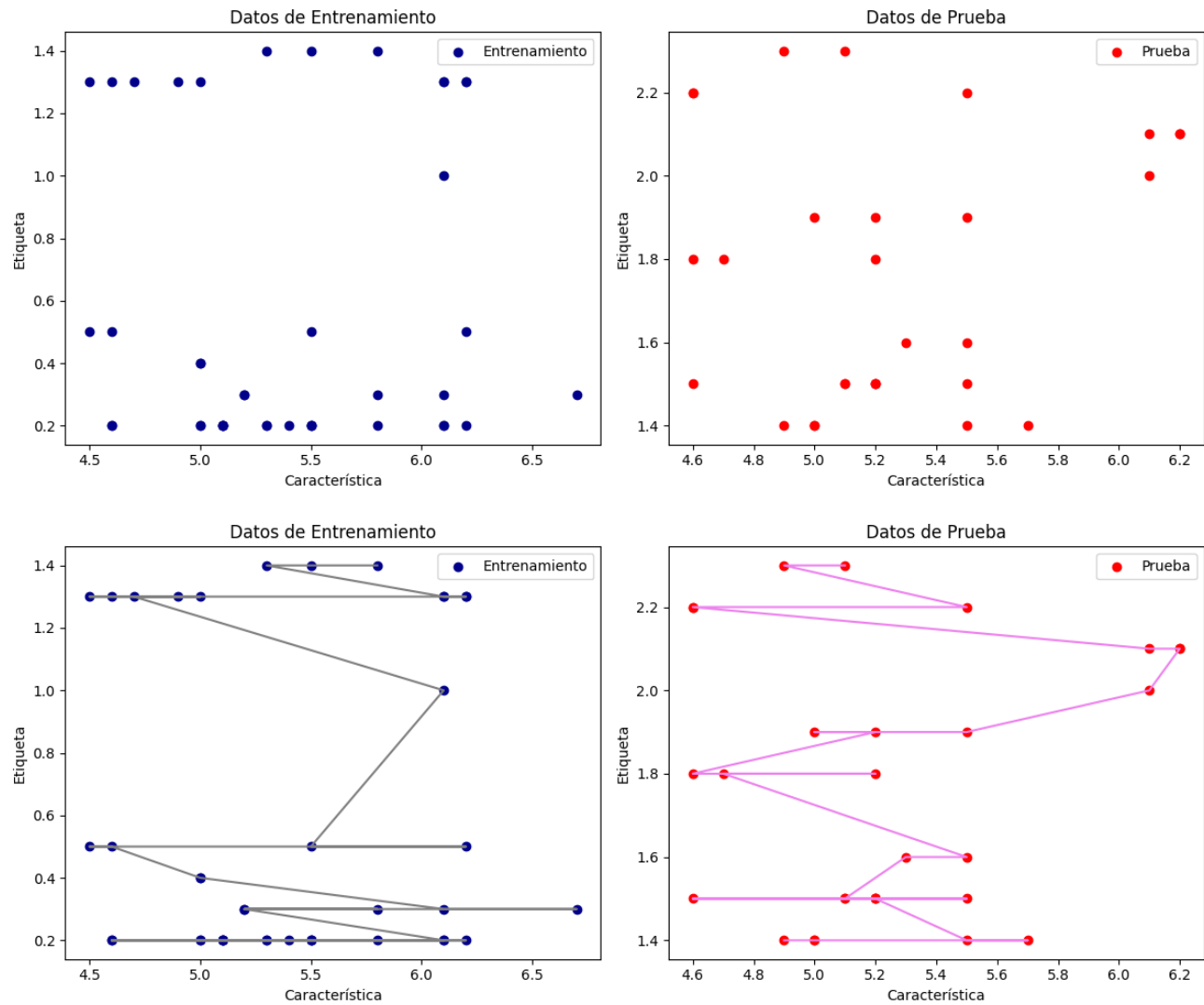


****Partición por Densidad:****

> Esta técnica ordena los datos según las etiquetas o clases y luego divide el conjunto de datos en conjuntos de entrenamiento y prueba manteniendo esa distribución ordenada. El objetivo es mantener la proporción de las clases en ambos conjuntos.

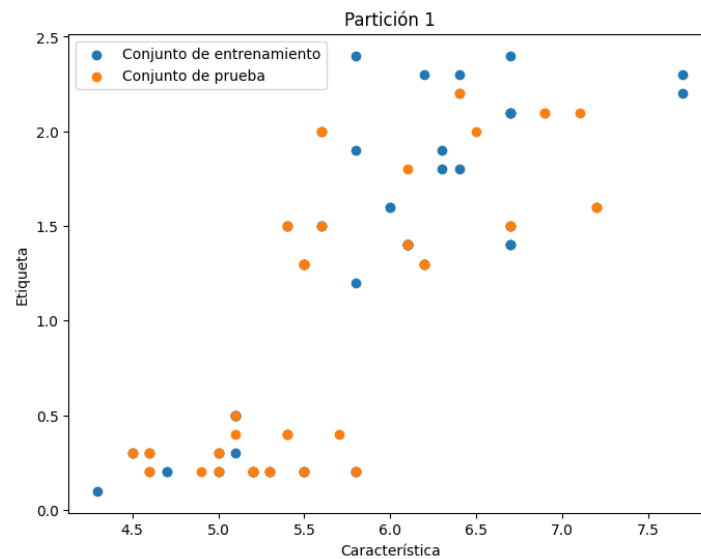


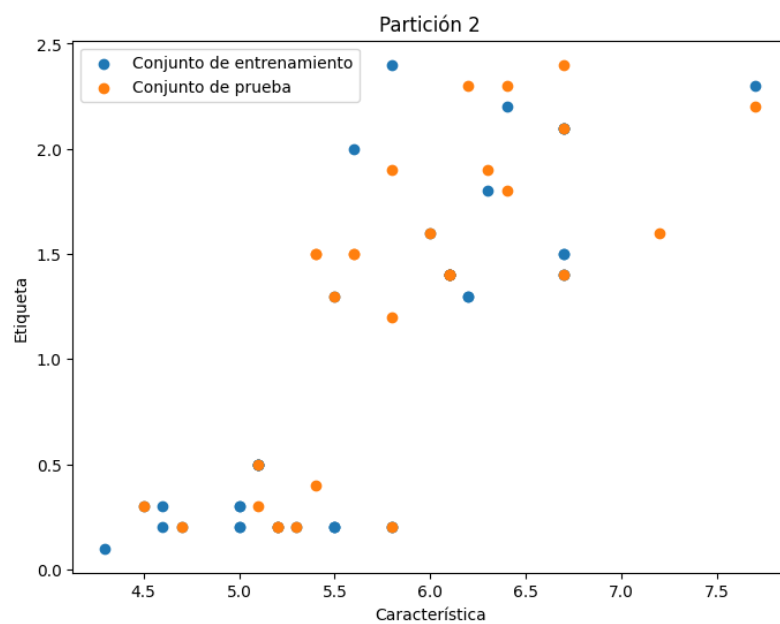
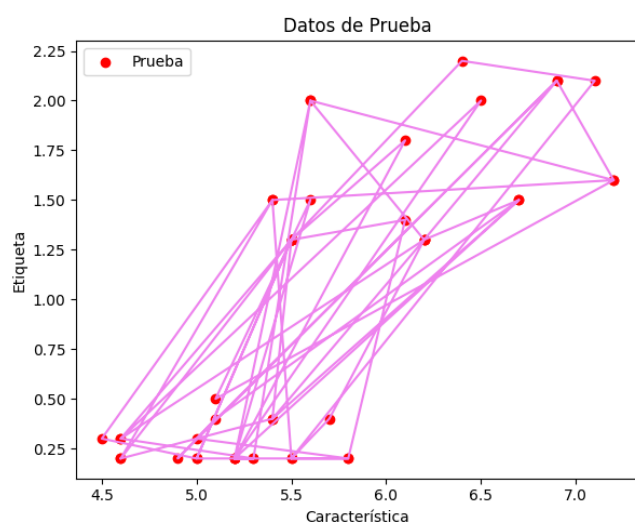
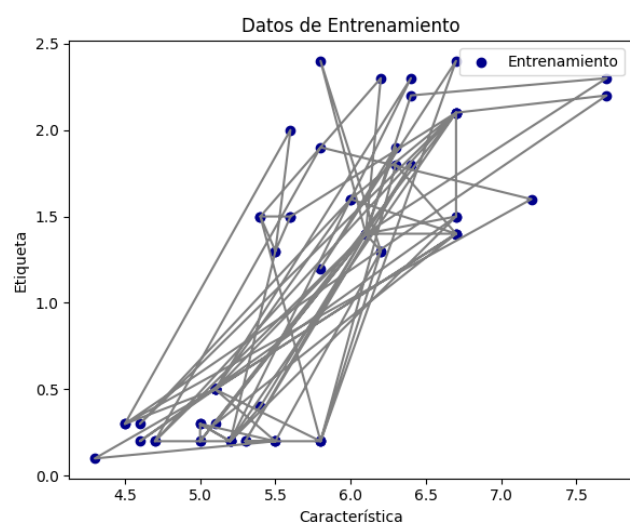
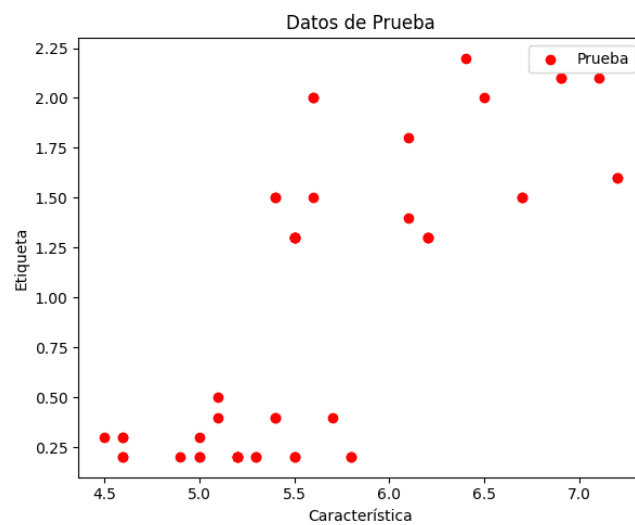
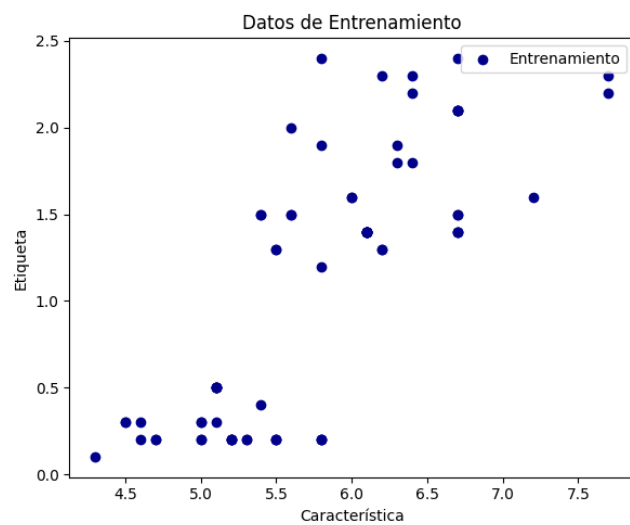


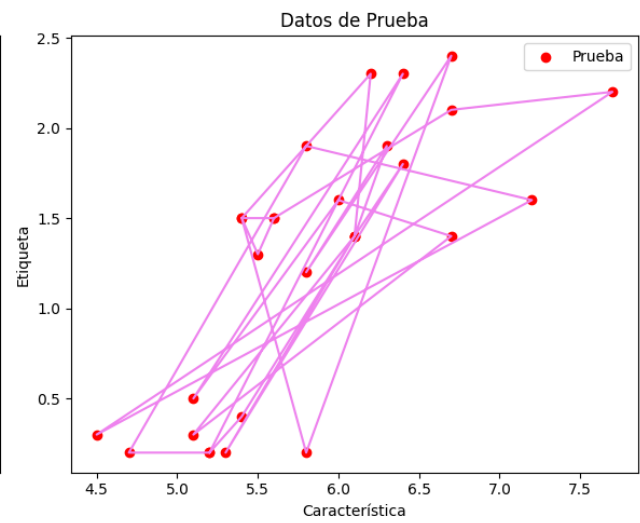
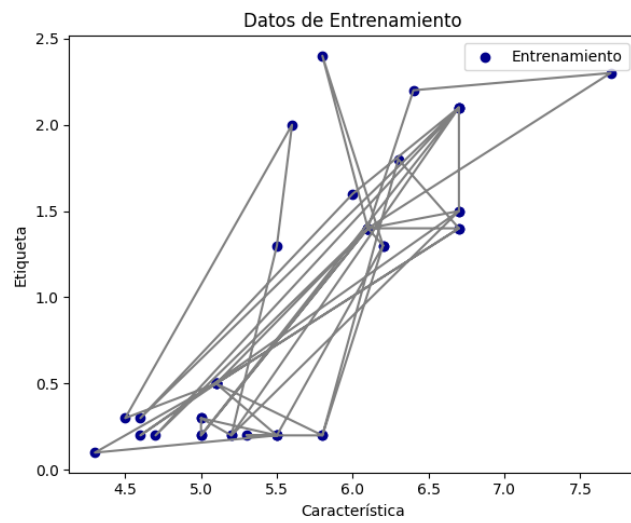
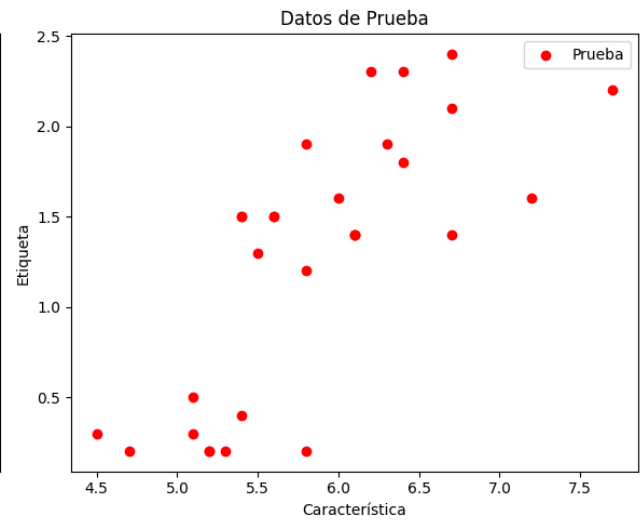
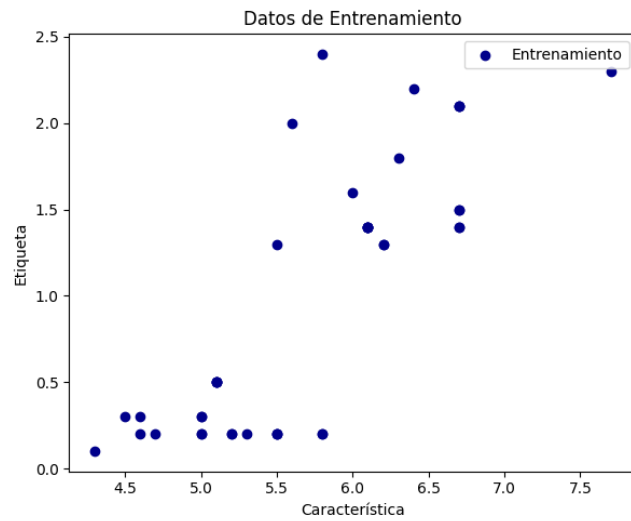


****Partición Secuencial:****

> En este enfoque, los datos se dividen secuencialmente en conjuntos de entrenamiento y prueba. Por ejemplo, se pueden tomar los primeros N porcentajes de datos para entrenamiento y el resto para prueba.

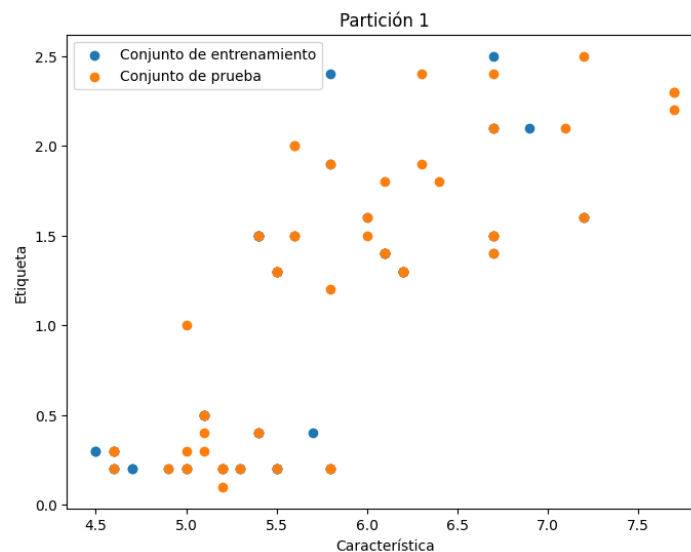


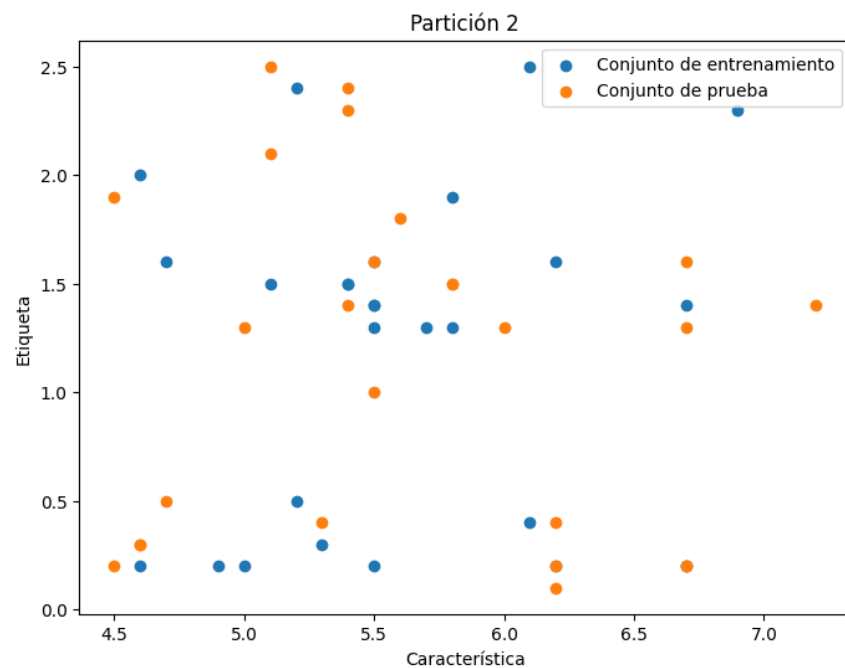
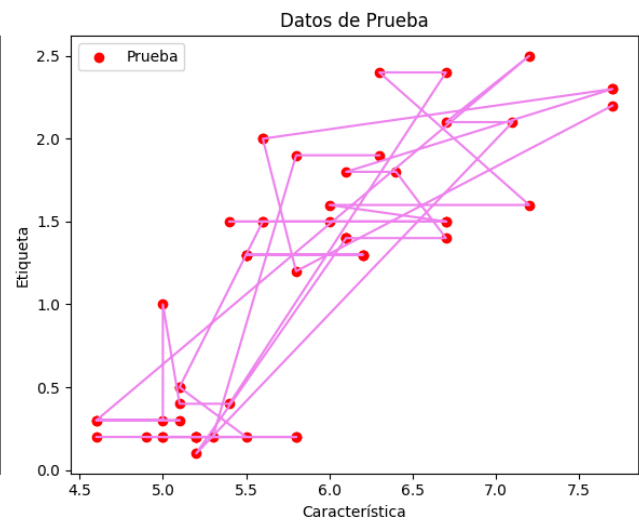
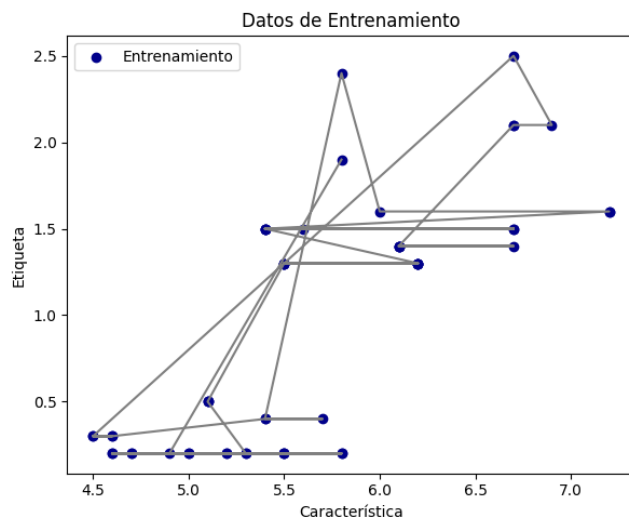
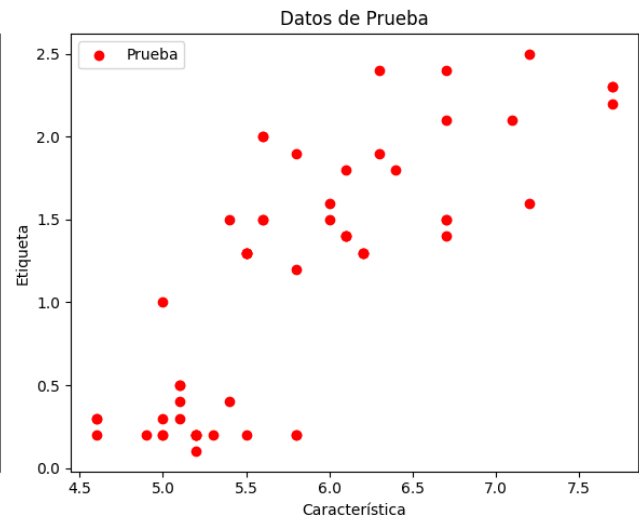
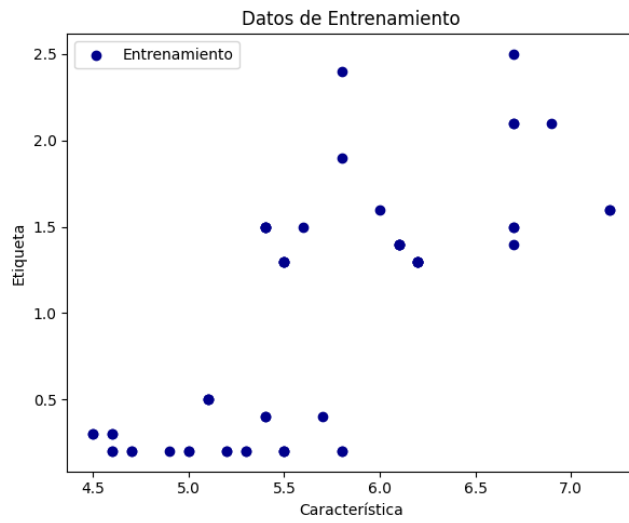


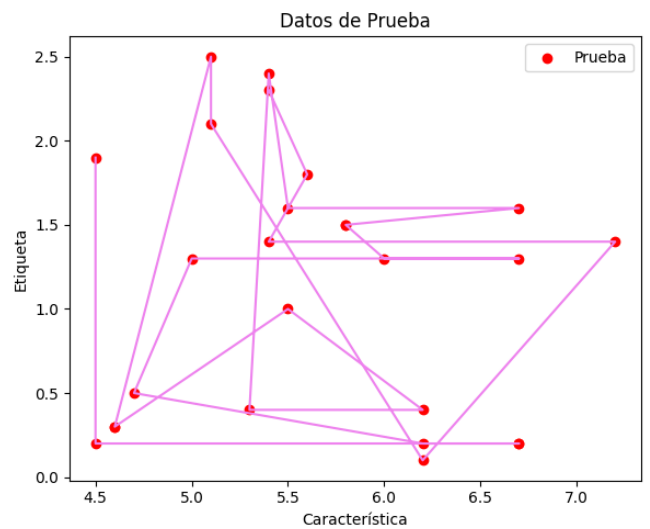
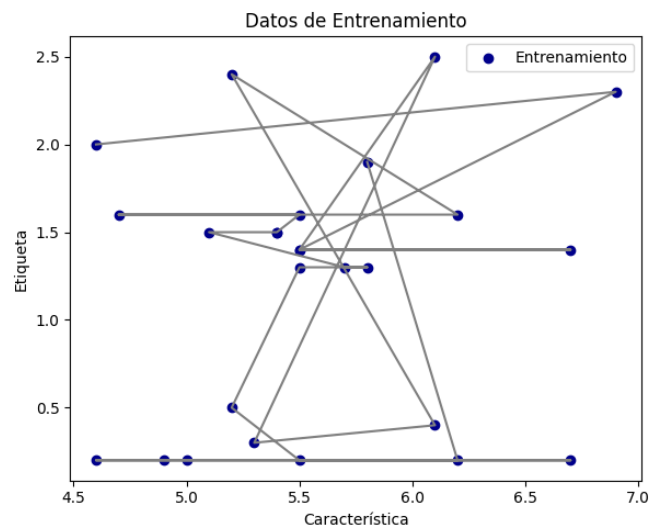
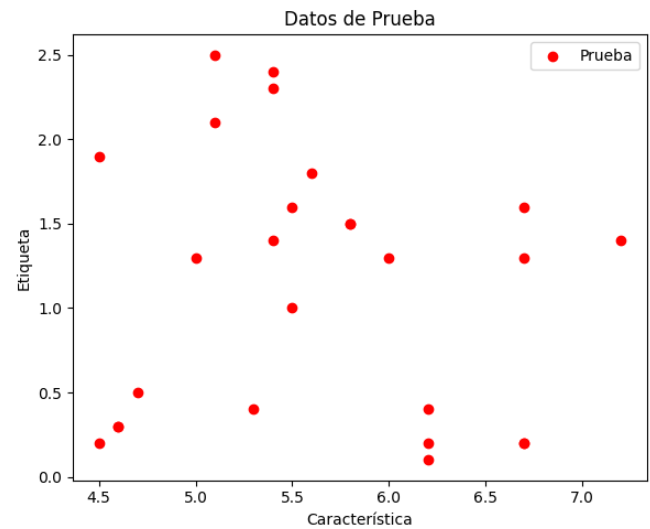
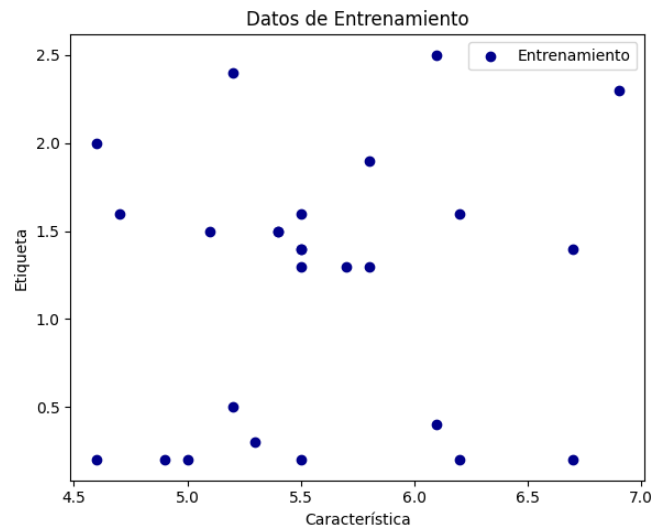


****Partición Estratificada:****

> Esta técnica es similar a la partición por densidad pero se enfoca específicamente en mantener la proporción de clases en ambos conjuntos de manera más estratégica. Es útil cuando se tienen conjuntos de datos desequilibrados, asegurando que la distribución de clases sea similar en ambos conjuntos.







****Partición Usando Validación Cruzada:****

> La validación cruzada es una técnica para evaluar la capacidad de generalización de un modelo. Uno de sus métodos, como k-fold cross-validation, divide el conjunto de datos en k subconjuntos, utilizando cada uno de ellos como conjunto de prueba mientras se entrena el modelo con el resto. Esto se repite k veces, lo que permite evaluar el rendimiento del modelo en diferentes conjuntos de prueba.

Estas técnicas de partición son fundamentales para evaluar modelos de machine learning, asegurando una buena generalización y minimizando sesgos en la evaluación del rendimiento del modelo. La elección de la técnica adecuada puede depender de factores como el tamaño del conjunto de datos, la distribución de clases y el tipo de problema de machine learning que se está abordando.