Para que se usa la ganancia de informacion en arboles de decisión: para medir la capacidad predictiva de una variable

El algoritmo c4.5 no maneja bien clases desbalanceadas.

Las reglas generadas por el algoritmos prism cubren todos los ejemplos de entrenamiento

El clasificador de bayes

El método de bagging tiene como objetivo generar diversidades en los clasificadores

Que técnica es mas apropiada para imputar valores perdidos en una variable? Imputar en base a un predictor sobre esa variable

El algoritmo ipf sirve para reducir el ruido

Para hacer una selección de variables mas eficaz con pocas dimensiones es mejor hacer una búsqueda global

Ventaja de las técnicas envolventes de selección de características son mas eficaces

El algoritmo ENN de selección de instancias elimina las instancias clasificadas incorrectamente. Suaviza la frontera pero no elimina.

Entre las técnicas de discretizacion ¿Cuál de ellas es supervisada? Intervalos basados en entropía

Un orden sensato de preprocesamiento es ruido, valores perdidos, selección de instancias.