

## Interpretación de la salida de Weka usando J48 con un 10-fold

**Si usamos Iris, que tiene 150 patrones y 3 clases:**

Para cada uno de los fold, saca un árbol diferente que se entrena por ejemplo con 135 patrones de train y 15 de test, por lo que se tendrían 10 árboles diferentes. La matriz de confusión la obtiene sumando los 15 patrones de test de cada uno de los 10-fold ( $15 \times 10 = 150$  patrones).

Cuando acaba todo el proceso ya se tienen resultados de rendimiento (CCR, F, etc), y lo que weka hace a continuación es entrenar con todos los datos para generar un árbol, por así decirlo, con más información que los anteriores, ya que ha entrenado con los 150 patrones en vez de con 135.

Si ahora lo que se quiere es crear un modelo que utilizar cuando te lleguen nuevos patrones, se debe usar ese árbol que ha sacado con los 150 patrones como modelo final.

**Cuestión:**

¿Entonces la matriz que reporta Weka no se ha obtenido con el modelo de árbol que aparece? ¿Se entiende entonces qué es una estimación de su rendimiento?

**Respuesta:**

Exacto, esa matriz se ha obtenido con el 10-fold, mientras que el modelo final se ha obtenido con todos los datos del conjunto. Sí, es una estimación de su rendimiento.