

UNIDAD TEMÁTICA 6: Ensamblés de Algoritmos

Trabajo de Aplicación 2

Ejercicio 1 Random Forest

Este algoritmo funciona de manera similar a los otros modelos de ensambles; el usuario tiene que especificar la cantidad de árboles de base a ser construidos.

Como el modelo interno de base es siempre un árbol de decisión, no tiene una especificación de subproceso interna.

Todos los parámetros específicos de los árboles se configuran a nivel del operador RandomForest.

El parámetro clave que especifica el número de árboles de base es *Number of Trees*.

Realización:

- En un nuevo proceso, incorpora el dataset Iris y un operador “Split Validation”
- Dentro del “Split Validation”, en la sección de modelo agrega un operador “Random Forest”,
- Configura los parámetros del Random Forest
 - Documenta los valores que eliges y justifícalos
 - Se sugiere comenzar con 10 árboles para observar los resultados.
- Ejecuta el proceso y observa los resultados. Podrás examinar todos los **m** modelos generados (en este caso 10)
- Los resultados de los modelos base se integran utilizando votación sencilla.
- Documenta los resultados. Analiza comparativamente estos resultados con los obtenidos en el Ejercicio 1.

Compara los resultados con los obtenidos en los ejercicios del TA1.

Describe los resultados comparados en un POSTESR, para discusión entre equipos.

Remite toda la tabla comparativa, producto de los 3 ejercicios, a la tarea de la webasignatura, junto con todos los modelos desarrollados.

Ejercicio 2 Adaboost

Este algoritmo funciona de manera similar a los otros modelos de ensambles.

En RapidMiner el operador “AdaBoost” es un subproceso, dentro del cual debemos especificar el modelo base a utilizar.

A nivel del operador, solamente hay un parámetro para configurar: “iteraciones”, que tiene un valor por defecto de 10

Internamente se puede utilizar cualquier modelo de base, que seleccionaremos de acuerdo a las características del problema.

Realización:

Parte 1

- En un nuevo proceso, incorpora el dataset “Crowdsourced Mapping Data Set”, de UCI.
- Analiza el problema y las características de los atributos.
- agrega un operador “AdaBoost”,
- Configura el operador AdaBoost
 - Modelo de base – en principio, Árbol de decisión
 - Iteraciones: en principio, 10
- Ejecuta el proceso y observa los resultados. Podrás examinar todos los modelos generados y cómo van evolucionando.
- Ejecuta el proceso para diferentes cantidades de iteraciones, y documenta los resultados.
- Prueba con algún modelo de base diferente. ¿Cuál sería bueno probar, dadas las características del dataset?
- Documenta los resultados. Analiza comparativamente estos resultados en función de las iteraciones y del algoritmo de base.

Parte 2

- Agrega un operador “Gradient Boosted Trees” y repite todo el proceso, comparando los resultados.
- ¿Cómo funciona este operador? Documentalo!

Compila todos los resultados obtenidos y remite toda la tabla comparativa a la tarea de la webassignatura, junto con todos los modelos desarrollados.