

Práctica 1 Grupo 4

Iria Lago Portela

Mario Picáns Rey

Javier Kniffki

David Bamio Martínez

Ejercicios

1. Obtener un árbol de decisión que permita clasificar las observaciones (universidades) en privadas (`Private="Yes"`) o públicas (`Private="No"`).
 - a. Seleccionar el parámetro de complejidad de forma automática, siguiendo el criterio de un error estándar de Breiman et al. (1984).
 - b. Representar e interpretar el árbol resultante.
 - c. Evaluar la precisión, de las predicciones y de las estimaciones de la probabilidad, en la muestra de test.
2. Realizar la clasificación anterior empleando Bosques Aleatorios mediante el método "`rf`" del paquete `caret`.
 - a. Considerar 300 árboles y seleccionar el número de predictores empleados en cada división `mtry = c(1, 2, 4, 6)` mediante validación cruzada, con 10 grupos y empleando el criterio de un error estándar de Breiman.
 - b. Representar la convergencia del error en las muestras OOB en el modelo final.
 - c. Estudiar la importancia de las variables y el efecto de las principales empleando algún método gráfico (para la interpretación del modelo).
 - d. Evaluar la precisión de las predicciones en la muestra de test y comparar los resultados con los obtenidos con el modelo del ejercicio anterior.
3. Realizar la clasificación anterior empleando SVM mediante la función `ksvm()` del paquete `kernlab`,
 - a. Ajustar el modelo con las opciones por defecto.
 - b. Ajustar el modelo empleando validación cruzada con 10 grupos para seleccionar los valores "óptimos" de los hiperparámetros, considerando las posibles combinaciones de `sigma = c(0.01, 0.05, 0.1)` y `C = c(0.5, 1, 10)` (sin emplear el paquete `caret`; ver Ejercicio 3.1 en *03-bagging_boosting-ejercicios.html*).
 - c. Evaluar la precisión de las predicciones de ambos modelos en la muestra de test y comparar también los resultados con los obtenidos en el ejercicio anterior.