Profesor: Dr. Oldemar Rodríguez Rojas
Análisis de Datos 2
Fecha de Entrega: Domingo 30 de octubre a las 12 media noche
Instrucciones:

- Las tareas deben ser subida la Aula Virtual antes de las 6:00pm. Luego de esta hora pierde 20 puntos y cada día de retraso adicional perderá 20 puntos más.
- Las tareas son estrictamente individuales.
- Tareas idénticas se les asignará cero puntos.
- Todas las tareas tienen el mismo valor en la nota final del curso.
- Cada día de entrega tardía tendrá un rebajo de 20 puntos.

Tarea Número 9

1. [**40 puntos**] La tabla de datos `novatosNBA.csv` contiene diferentes métricas de desempeño de novatos de la NBA en su primera temporada. Para esta tabla, las 21 primeras columnas corresponden a las variables predictoras y la variable `Permanencia` es la variable a predecir, la cual indica si el jugador permanece en la NBA luego de 5 años. La tabla contiene 1340 filas (individuos) y 21 columnas (variables), con la tabla realice lo siguiente:

   a) Usando Bosques Aleatorios con el 80 % de los datos para la tabla aprendizaje y un 20 % para la tabla testing determine la mejor Probabilidad de Corte, de forma tal que se prediga de la mejor manera posible la categoría 0 de la variable Permanencia, pero sin desmejorar de manera significativa la precisión de la categoría 1.

   b) Repita el ejercicio anterior usando `XGBoosting`. ¿Cambió la probabilidad de corte? Explique.

2. [**40 puntos**] Utilizando nuevamente la tabla `novatosNBA.csv` realice lo siguiente:

   a) Compare todos los métodos predictivos vistos en el curso con esta tabla de datos utilizando la curva `ROC` y el área bajo la curva `ROC` Aquí interesa predecir en la variable `Permanencia`. Compare los métodos `SVM`, `KNN`, `Árboles`, `Bosques`, `ADA Boosting`, `eXtreme Gradient Boosting`, `Bayes`, `LDA` y `QDA`. Utilice los parámetros por defecto o los que usted mejor considere.

   b) ¿Qué se puede concluir?

3. [**20 puntos**] Dada la siguiente tabla:

Individuo	Clase	Score
1	P	0.68
2	N	0.16
3	N	0.85
4	P	0.21
5	N	0.58
6	N	0.66
7	N	0.80
8	N	0.29
9	N	0.30
10	P	0.51

a) Usando la definición de curva `ROC` calcule y grafique "a mano" la curva ROC, use un umbral $T = 0$ y un paso de 0.1. Es decir, debe hacerlo variando el umbral y calculando las matrices de confusión.

b) Verifique el resultado anterior usando el código visto en clase, denominado `PROGRAMA 1`.

c) Usando el algoritmo eficiente para la curva `ROC` calcule y grafique "a mano" la curva ROC, use un umbral $T = 0.1$ y un paso de 0.1.

d) Verifique el resultado anterior usando el código visto en clase para el algoritmo eficiente, `PROGRAMA 2`.

oldemar rodríguez

CONSULTOR en M1NƐR14 DƐ D4T0S