

PRÁCTICA 4:

Fecha límite de entrega: 6 de junio 2018

Valoración: 25 puntos

NORMAS DE DESARROLLO Y ENTREGA DE TRABAJOS

Para este trabajo como para los demás es obligatorio presentar un informe escrito con sus valoraciones y decisiones adoptadas en el desarrollo de cada uno de los apartados. Incluir en el informe los gráficos generados. También deberá incluirse una valoración sobre la calidad de los resultados encontrados. (obligatorio en pdf). **Sin este informe se considera que el trabajo NO ha sido presentado.**

Normas para el desarrollo de los Trabajos: EL INCUMPLIMIENTO DE ESTAS NORMAS SIGNIFICA PERDIDA DE 2 PUNTOS POR CADA INCUMPLIMIENTO.

- El código se debe estructurar en un único script R con distintas funciones o apartados, uno por cada ejercicio/apartado de la práctica.
- Todos los resultados numéricos o gráficas serán mostrados por pantalla, parando la ejecución después de cada apartado. No escribir nada en el disco.
- El path que se use en la lectura de cualquier fichero auxiliar de datos debe ser siempre "datos/nombre_fichero". Es decir, crear un directorio llamado "datos" dentro del directorio donde se desarrolla y se ejecuta la práctica.
- Un código es apto para ser corregido si se puede ejecutar de principio a fin sin errores.
- NO ES VÁLIDO usar opciones en las entradas. Para ello fijar al comienzo los parámetros por defecto que considere que son los óptimos.
- El código debe estar obligatoriamente comentado explicando lo que realizan los distintos apartados y/o bloques.
- Poner puntos de parada para mostrar imágenes o datos por consola.
- ENTREGAR SOLO EL CODIGO FUENTE, NUNCA LOS DATOS.
- **Forma de entrega:** Subir los ficheros .R y .pdf ... según requerido en la definición de la entrega en el Tablón docente de CCIA. Y en un zip todos ficheros en DECSAI.

1. AJUSTE DEL MEJOR MODELO

Este ejercicio se centra en el ajuste de los modelos no-lineales SVM y Random Forest a un conjunto de datos. El objetivo es encontrar la mejor solución posible con dichos algoritmos y discutir sus pros y contras sobre el conjunto de datos.

Se habrá de justificar cada uno de los pasos dados para conseguir la solución aportada. Los puntos de discusión señalados en el trabajo.3 deben de servir como guía.

Se usará para ello la BBDD del repositorio de la UCI Student Performance Data Set (<https://archive.ics.uci.edu/ml/datasets/student+performance>)

OBJETIVOS:

- Abordar el problema de clasificación binaria definido por la umbralización de la variable $G3$. Clase-1: $G3 \geq 10$ (pasa); Clase-2: $G < 10$ (suspende);
- Usar como métrica de evaluación la curva ROC y la matriz de confusión.