

Trabajo.3: Programación

Fecha límite de entrega: 19 de Mayo 2017

Valoración: 12.5 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 DE MODELOS LINEALES

Este ejercicio se centra en el ajuste de un modelo lineal a conjuntos de datos dadas con el objetivo de obtener el mejor predictor posible. En todos los casos los pasos a desarrollar serán aquellos que nos conduzcan al ajuste y selección del mejor modelo y a la estimación del error E_{out} del modelo final. Cómo mínimo se habrán de analizar y comentar los siguientes pasos **sobre un problema de clasificación y otro de regresión**:

1. Comprender el problema a resolver.
2. Preprocesado los datos: por ejemplo categorización, normalización, etc.
3. Selección de clases de funciones a usar.
4. Definición de los conjuntos de training, validación y test usados en su caso.
5. Discutir la necesidad de regularización y en su caso la función usada para ello.
6. Definir los modelos a usar y estimar sus parámetros e hiperparámetros.
7. Selección y ajuste modelo final.
8. Discutir la idoneidad de la métrica usada en el ajuste
9. Estimación del error E_{out} del modelo lo más ajustada posible.
10. Discutir y justificar la calidad del modelo encontrado y las razones por las que considera que dicho modelo es un buen ajuste que representa adecuadamente los datos muestrales.

Las clases de funciones de las transformaciones no lineales pueden definirse a partir de las potencias y los productos de potencias de las variables originales. Si se usan otras transformaciones de las variables iniciales, como \log , $\sqrt{}$, \sin , etc deberá justificarse el interés de dicha elección.

Bases de datos a usar

- <https://archive.ics.uci.edu/ml/datasets/optical+recognition+of+handwritten+digits>
- <https://archive.ics.uci.edu/ml/datasets/Airfoil+Self-Noise>

Alternativamente estos ficheros de datos estarán disponibles también en la web de decsai.

Recomendación: desarrollar un código en Python/R lo suficientemente general que permita ser reusado, en su mayor parte, en el desarrollo del Proyecto Final. Se recomienda escribir funciones que permitan ser reusadas.