

Tarea 1

Análisis Exploratorio y Redes Neuronales Artificiales

Curso 2025– 2026

La entrega de la tarea se realizará a través de la plataforma **Aula Virtual** antes del día **17 de noviembre a las 23.55h**. Si se entregan ficheros comprimidos, deben tener formato **.zip**. No se evaluarán entregas con otros formatos (.rar, ...). Se **debe** entregar:

- **La base de datos utilizada.**
- El código que implementa la tarea 1. Este fichero debe comenzar con un comentario indicando el nombre y apellidos de los integrantes del grupo. Se debe identificar claramente la parte del código asociada a cada sección, y debe indicarse claramente la separación entre las secciones indicadas en este guión. Se debe explicar la **estrategia a seguir y justificar razonadamente** cada paso mediante comentarios en el fichero del código.
- Un **breve informe en formato pdf** (máximo de 4 hojas a una cara, sin contar anexos) **explicando resultados obtenidos particularizados a la base de datos de trabajo**. Debe **justificar la elección de la(s) técnica(s) aplicada(s), analizando el resultado y proponiendo mejoras**. El informe debe contener el nombre de los integrantes del grupo y seguir las **normas de estilo en la redacción de trabajos académicos** (véase material disponible en la pestaña “Otros Recursos” de Aula Virtual). El informe no debe contener explicación de las instrucciones, que en su caso debieran aparecer en el fichero de código.
- Los anexos son **optativos**, con una extensión máxima de 4 páginas.

En esta tarea, la puntuación asociada a la explicación de la estrategia, justificaciones y desarrollo del código corresponde a **5.5 puntos**, y la puntuación asociada al informe (documento .pdf) corresponde a **2.5 puntos**.

Tanto la base de datos como el informe y el fichero para ejecutar el código deben nombrarse como “GXX.extension”, donde “XX” corresponde al número de grupo asignado conforme al listado publicado en Aula Virtual, y “.extension” será una de las extensiones permitidas en función del tipo de fichero (.pdf, etc).

Adicionalmente, los integrantes del grupo deberán entregar un **único vídeo explicando y justificando**, de manera estructurada, el trabajo realizado. En el vídeo deben participar en la misma proporción de tiempo los dos integrantes del grupo en caso de grupos con dos integrantes. La duración mínima del vídeo es de 3 minutos, y la máxima de 4 minutos. En el vídeo se debe dedicar aproximadamente la misma duración a explicar el objetivo obligatorio y el objetivo creativo. El vídeo se debe poder reproducir utilizando el software Quick Time Player. Cada vídeo debe ser nombrado como **GXX_Tarea1**, donde XX será la denominación del grupo, conforme al listado publicado en Aula Virtual. El vídeo debe registrar la voz del estudiante; no debe contener explicación del código utilizado, sino la explicación e interpretación de los resultados. En esta tarea, **la puntuación máxima del vídeo es de 2 puntos**.

En caso de que el grupo esté formado por dos integrantes, sólo el “Integrante-1” del grupo (véase archivo publicado en Aula Virtual) debe subir a Aula Virtual el material de la tarea. En caso de que los dos integrantes realicen un envío, sólo se evaluará el trabajo enviado en último lugar.

Para el desarrollo de la tarea, cada grupo hará uso de la base de datos seleccionada, cuya referencia está disponible en el documento de asignación que aparece bajo el epígrafe “Relación de grupos grupos-proyectos (act:8 oct)”. A no ser que se indique lo contrario en el guión de cada una de las tareas proporcionada por los profesores, esta base de datos será utilizada por cada grupo en todas las tareas.

Es obligatorio utilizar las **librerías de *scikit-learn***, y **no utilizar PyTorch ni TensorFlow**.

Es obligatorio utilizar cuadernillo de Jupyter con todas las celdas ejecutadas y los **resultados finales visibles en el mismo cuadernillo**.

Evaluación

Cada tarea será evaluada en base a:

- *Calidad técnica*. La estructura y complejidad del análisis y de los experimentos realizados. Se debe describir adecuadamente el conjunto de datos, así como las decisiones tomadas para su preprocesamiento. Una tarea que incluya un análisis claro y estructurado de los resultados obtenidos con cada implementación y proporcione varias figuras de mérito, incluyendo una explicación adecuada de las mismas, tendrá una mayor puntuación. Se puede utilizar técnicas no presentadas durante el desarrollo de la asignatura siempre y cuando estén bien explicadas y su elección esté justificada. Es importante proporcionar las referencias consultadas, tanto en el código como en el documento escrito.
- *Comunicación*. Se tendrá en cuenta la claridad de las explicaciones y la estructura del código Python, del documento escrito y del vídeo. En la elaboración del vídeo se deben utilizar diapositivas para presentar y explicar los resultados obtenidos.

Sección I. Análisis descriptivo y exploratorio de la base de datos. Preprocesamiento (2 puntos)

Explique el objetivo que se aborda con el análisis de la base de datos considerada, identificando claramente si se trata de una tarea de clasificación o de regresión. Realice un análisis descriptivo y exploratorio de la base de datos haciendo uso, entre otras, de herramientas de visualización.

Teniendo en cuenta el resultado del análisis anterior, justifique razonadamente si es necesario realizar un preprocesamiento de los datos y, en caso afirmativo, en qué consistiría.

Discuta, apoyándose de un análisis adecuado, sobre la relación entre distintas características y entre las características y el *target*. Puede hacer uso del valor del coeficiente de correlación (y de otros similares) en función de la naturaleza de las variables consideradas.

Se deben utilizar un máximo de dos diapositivas en el vídeo para sintetizar el objetivo, el resultado del análisis descriptivo y exploratorio, y el pre-procesamiento realizado.

Puede incluir en un anexo de la memoria figuras o descripciones que considere relevantes, siempre que se encuentren adecuadamente referenciadas en la Sección I.

Sección II. Esquema lineal y selección de características (2 puntos)

Para abordar la tarea debe hacer uso, primero, de un esquema de regresión lineal o logístico (en función de si la tarea a resolver es de regresión o de clasificación, respectivamente) que considere todas las variables disponibles* y no haga uso de ningún término de regularización. Analice los coeficientes obtenidos y obtenga conclusiones sobre los resultados.

A continuación, aplique alguna técnica de regularización (indicando y justificando cuál ha elegido), seleccionando de forma adecuada el hiper-parámetro de regularización. Analizando los coeficientes obtenidos, justifique qué características son más relevantes. Discuta los resultados obtenidos y extraiga conclusiones, para lo cual debe hacer uso de varias figuras de mérito (al menos, 4).

Por último, utilice un método *wrapper* para seleccionar las características más informativas de entre las originalmente disponibles. Resuelva la tarea de clasificación/regresión (cada grupo, la tarea asociada) utilizando el subconjunto de características elegido (considerando las mismas observaciones que previamente) y un esquema lineal. Discuta los resultados y extraiga conclusiones.

***Nota-1:** Recordad que el Id (identificador) de la observación no es una característica informativa y no se debe usar como característica de entrada a un esquema de aprendizaje automático.

Sección III. Perceptrón Multicapa o MLP (Multi-Layer Perceptron) (2.5 puntos)

Explique si considera razonable abordar el diseño de un MLP con una única capa oculta de neuronas. Explique, aplicado a su conjunto de observaciones, a qué hace referencia la afirmación “El MLP es un aproximador universal”.

Discuta la necesidad o no de normalizar las características antes de realizar el entrenamiento del modelo.

- Entrene un MLP considerando **una única capa oculta** y distinto número de neuronas ocultas haciendo uso sólo de un esquema de aprendizaje “batch” y considerando suficientes épocas (o iteraciones). Explore el uso de distintas funciones de activación. Monitorice y represente la evolución del valor de la figura de mérito que considere más adecuada para abordar su tarea y detenga el proceso iterativo de aprendizaje en función del valor de esa figura de mérito. ¿Qué arquitectura ha elegido? Justifique su respuesta.

Utilice la arquitectura elegida para evaluar las prestaciones del modelo obteniendo los valores asociados a las 4 figuras de mérito anteriores. Comente los resultados obtenidos.

- Entrene un MLP considerando **más de una capa oculta** haciendo uso sólo de un esquema de aprendizaje “batch” y considerando suficientes épocas (o iteraciones). Explore el uso de **distintas funciones de activación en las neuronas**. Monitorice y represente la evolución del valor de la figura de mérito que considere más adecuada para abordar su tarea y detenga el proceso iterativo de aprendizaje en función del valor de esa figura de mérito. ¿Qué arquitectura ha elegido? Justifique su respuesta.

Utilice la arquitectura elegida para evaluar las prestaciones del modelo obteniendo los valores asociados a las 4 figuras de mérito anteriores. Comente los resultados obtenidos.

- Utilice un método *wrapper* que haga uso de un MLP para seleccionar las características más informativas de entre las originalmente disponibles. Resuelva la tarea de clasificación/regresión (cada grupo, la tarea asociada) utilizando el subconjunto de características elegido (considerando las mismas observaciones que previamente) y un esquema lineal. Discuta los resultados y extraiga conclusiones.

Sección IV. Comparación de resultados (1.5 puntos)

Compare los resultados obtenidos con los modelos diseñados en los apartados anteriores. Para ello, proporcione al menos una tabla con los valores obtenidos para las 4 (o más) figuras de mérito consideradas. Comente y justifique los resultados.

De los esquemas de aprendizaje automático utilizados en esta tarea, justifique razonadamente qué esquema y conjunto de características elegiría.

Explique **detallada y razonadamente** el procedimiento a seguir para obtener estadísticos de cada una de las figuras de mérito (por ejemplo, media y desviación típica) **sobre test**. **No debe implementarlo**, únicamente explicar **de forma detallada** en una sección del cuadernillo, considerando como comentario, cómo lo obtendría. Se debe incluir un esquema gráfico del procedimiento en un anexo, cuya extensión computará en el máximo número de páginas permitidas como anexo.

Nota: Todas las entregas se realizarán **en un fichero .zip** a través del Aula Virtual de la asignatura. **No se evaluarán entregas con otros formatos (.rar, ...).**

Profesora responsable: Inmaculada Mora Jiménez