

Asignatura	Datos del alumno	Fecha
Técnicas de inteligencia Artificial	Apellidos:	
	Nombre:	

Actividad: Trabajando con redes neuronales y Deep Learning

Objetivos

Esta actividad te permitirá profundizar en la aplicación de técnicas de aprendizaje supervisado (regresión y clasificación) basadas en redes neuronales utilizando las librerías Keras y TensorFlow sobre Python. Para ello, habrás de elegir dos dataset en plataformas online como Kaggle, OpenML o Google Dataset Search en base a unos requisitos mínimos especificados y aplicar, en un primer caso, una red neuronal para la regresión comparada con otra técnica de regresión no basada en redes neuronales y, en un segundo caso, otra red neuronal diferente para la clasificación, de nuevo comparada con otra técnica de clasificación no basada en redes neuronales.

Descripción

En primer lugar, repasa los siguientes contenidos teórico-prácticos de la asignatura:

- ▶ Tema 2: “Python para la implementación de técnicas de inteligencia artificial”, haciendo especial hincapié en las secciones:
 - 2.6. Librerías útiles para el análisis de datos (apartados Keras, TensorFlow y Pandas).
 - 2.8. Importación de datos.
 - 2.9. Introducción a Machine Learning con librerías en Python.
- ▶ Tema 5: “Redes neuronales artificiales”, especialmente:
 - 5.7. Aplicaciones y ejemplos de implementación.
- ▶ Tema 6: “Deep Learning”, especialmente:
 - 6.12. Ejemplos de implementación.

Asignatura	Datos del alumno	Fecha
Técnicas de inteligencia Artificial	Apellidos:	
	Nombre:	

El anterior repaso y la ejecución de los ejemplos incluidos en los diferentes apartados te permitirá asegurar que cuentas en tu máquina computadora con el entorno de ejecución de Python y con todas las librerías necesarias para llevar a cabo esta actividad, además de haber realizado el ejemplo básico con TensorFlow 2.0, así como el tutorial oficial de dicha librería (secciones 2.6 y 6.12). Del mismo modo, permitirá asegurar que estás familiarizado con los conceptos de regresión lineal y regresión logística (recuerda, que, en este último caso, se trata de un clasificador).

Se recomienda encarecidamente, además, al menos seguir los dos siguientes tutoriales oficiales de TensorFlow 2.0:

- [TensorFlow 2.0: Conceptos básicos de AA con Keras: Regresión.](#)
- [TensorFlow 2.0: Conceptos básicos de AA con Keras: Clasificación de imágenes básica.](#)

La actividad está dividida en dos apartados: regresión y clasificación. Habrás de entregar un único informe de 4 páginas como máximo de extensión (Arial 11, interlineado 1,5) y dos archivos Python, uno para el problema de regresión y otro para el problema de clasificación. Ambos archivos han de poder ejecutarse de forma independiente el uno del otro. **El informe no debe incluir anexos, portada ni repetir el enunciado o la rúbrica de la actividad.** Cualquier página que incluyas será computada dentro de la extensión entregada. No se evaluará ninguna página a partir más allá de la cuarta página. **No es posible adjuntar ningún archivo adicional.** El propio código Python de los dos archivos ha de ingerir/descargar automáticamente los datos de entrenamiento y test correspondientes a cada uno y trabajar sobre ellos.

Puedes elegir libremente cómo repartes las 4 páginas de extensión del informe para cada uno de los apartados.

Asignatura	Datos del alumno	Fecha
Técnicas de inteligencia Artificial	Apellidos:	
	Nombre:	

En ambos apartados has de utilizar, al menos, las siguientes librerías:

- ▶ **Pandas** para la importación de datos.
- ▶ Keras sobre TensorFlow 2.0 (tensorflow.keras).
- ▶ Eres libre de utilizar otras librerías para trabajar con otros modelos que no sean las redes neuronales (p.ej., scikit-learn), para la muestra de gráficas (p.ej., matplotlib o seaborn) o para otras operaciones auxiliares, si así lo necesitaras.

APARTADO A (Regresión)

- ▶ Comenzamos con el ejercicio de regresión. En primer lugar, has de elegir un dataset de acceso público y disponible online. El dataset ha de ser apropiado para aplicar diferentes técnicas de regresión y compararlas entre sí.
- ▶ Puedes elegir un dataset en Kaggle, OpenML o Google Dataset Search (es un metabuscador, de modo que te puede llevar igualmente a los anteriores). Otras fuentes de datos son posibles, siempre que sean de acceso público y estén disponibles online sin necesidad de autenticarse, pues habrás de indicar dicha fuente y se comprobará su accesibilidad para evaluar la actividad.
- ▶ **El dataset escogido ha de contar, como mínimo, con 1 000 instancias y has de trabajar con todas ellas. Es tu elección elegir la fracción que empleas para entrenar los modelos y la fracción que utilizas para el test de los mismos, siempre que utilices en total al menos 1 000 instancias en todos los modelos.**
- ▶ El problema de regresión a resolver contará con, al menos, una variable/atributo (obviamente numérico) a la salida, y **que dependa de, al menos, 6 variables/atributos de entrada.**
- ▶ Describe el conjunto de datos utilizado y explica el problema de regresión a resolver (no son necesarios contenidos teóricos, sino explicar qué relaciones tratas de comprobar y con qué métodos).

Asignatura	Datos del alumno	Fecha
Técnicas de inteligencia Artificial	Apellidos:	
	Nombre:	

- ▶ Utiliza la librería **pandas** para la importación del dataset. **No es posible adjuntar ningún archivo adicional.** El propio código Python del archivo aportado ha de ingerir/descargar automáticamente los datos de entrenamiento y test correspondientes y trabajar sobre ellos. **Preprocesa los datos si fuera necesario.**
- ▶ Mediante Python y las librerías que consideres, analiza el dataset proporcionando una caracterización del mismo, mostrando **al menos** algunas de sus características en modo texto (mediante tablas o prosa) y **al menos** algunas de ellas en modo gráfico (p.ej., histogramas, diagramas de dispersión, diagramas de cajas y bigotes, etc.). **Las características y las gráficas incluidas han de provenir de la ejecución del código en Python que se aporte como respuesta.**
- ▶ **Normaliza los datos si fuera necesario.**
- ▶ **Elige al menos un método de regresión no basado en redes neuronales** (p.ej. regresión lineal, regresión polinómica, regresión logarítmica, SVR, random forest regression, etc.).
- ▶ **Incluye en el informe una explicación de los parámetros que consideres relevantes en cada ejecución.**
- ▶ Mediante Python y las librerías que consideres, entrena el modelo o modelos escogidos, dividiendo el dataset en datos de entrenamiento y datos de test previamente en base a tu criterio y pruébalos frente a dichos datos de test.
- ▶ **Elige al menos una arquitectura de red neuronal (describe en el informe las neuronas en la capa de entrada, las capas intermedias – **al menos dos capas intermedias** – y capa de salida, funciones de activación en cada caso) que permita realizar una **regresión**.**
- ▶ Mediante Python y utilizando **al menos Keras sobre TensorFlow 2.0 (tensorflow.keras)**, entrena el modelo o modelos de red neuronal escogidos, dividiendo el dataset en datos de entrenamiento y datos de test previamente en base a tu criterio y pruébalos frente a dichos datos de test.

Asignatura	Datos del alumno	Fecha
Técnicas de inteligencia Artificial	Apellidos:	
	Nombre:	

- ▶ Mediante Python y las librerías que consideres, **muestra los resultados** obtenidos por los diferentes algoritmos escogidos de forma gráfica y comparada/superpuesta.
- ▶ Incluye en el informe una **discusión sobre los resultados obtenidos**, argumenta con qué técnica se obtienen mejores resultados en base a las diferentes métricas que hayas escogido y **explica cómo se podrían mejorar los resultados obtenidos por las redes neuronales**, independientemente de que mejoren o no a los algoritmos no basados en redes neuronales.
- ▶ **Recuerda, incluye en el informe a entregar al menos lo siguiente:**
 - Descripción de la fuente de datos empleada.
 - Caracterización del dataset utilizado en modo texto y gráfico.
 - Parámetros relevantes utilizados en los diferentes algoritmos (en el caso de las redes neuronales, su arquitectura, al menos).
 - Resultados obtenidos por los diferentes algoritmos escogidos de forma gráfica y comparada/superpuesta.
 - Discusión de los resultados obtenidos y argumentos sobre cómo mejorar de dichos resultados.

APARTADO B (Clasificación)

- ▶ Seguimos, de forma análoga, con el ejercicio de clasificación. En primer lugar, has de elegir un dataset de acceso público y disponible online. El dataset ha de ser apropiado para aplicar diferentes técnicas de clasificación y compararlas entre sí.
- ▶ Puedes elegir un dataset en Kaggle, OpenML o Google Dataset Search. De nuevo, otras fuentes de datos son posibles, siempre que sean de acceso público y estén disponibles online sin necesidad de autenticarse, pues habrás de indicar dicha fuente y se comprobará su accesibilidad para evaluar la actividad.
- ▶ **El dataset escogido ha de contar, como mínimo, con 1 000 instancias y has de trabajar con todas ellas. Es tu elección elegir la fracción que empleas para**

Asignatura	Datos del alumno	Fecha
Técnicas de inteligencia Artificial	Apellidos:	
	Nombre:	

entrenar los modelos y la fracción que utilizas para el test de los mismos, siempre que utilices en total al menos 1 000 instancias en todos los modelos.

- ▶ El problema de clasificación a resolver contará con, al menos, una variable categórica de **al menos 6 clases posibles a la salida**, y **que dependa de, al menos, 6 variables/atributos de entrada**.
- ▶ Describe el conjunto de datos utilizado y explica el problema de clasificación a resolver (no son necesarios contenidos teóricos, sino explicar qué relaciones tratas de comprobar y con qué métodos).
- ▶ Utiliza la librería **pandas** para la importación del dataset. **No es posible adjuntar ningún archivo adicional**. El propio código Python del archivo aportado ha de ingerir/descargar automáticamente los datos de entrenamiento y test correspondientes y trabajar sobre ellos. **Preprocesa los datos si fuera necesario**.
- ▶ Mediante Python y las librerías que consideres, analiza el dataset proporcionando una caracterización del mismo, mostrando **al menos** algunas de sus características en modo texto (mediante tablas o prosa) y **al menos** algunas de ellas en modo gráfico (p.ej., histogramas, diagramas de dispersión, diagramas de cajas y bigotes, etc.). **Las características y las gráficas incluidas han de provenir de la ejecución del código en Python que se aporte como respuesta**.
- ▶ **Normaliza los datos si fuera necesario**.
- ▶ **Elige al menos un método de clasificación no basado en redes neuronales** (p.ej. regresión logística, árboles de decisión, reglas de clasificación, random forest, SVM, etc.).
- ▶ **Incluye en el informe una explicación de los parámetros que consideres relevantes en cada ejecución**.
- ▶ Mediante Python y las librerías que consideres, entrena el modelo o modelos escogidos, dividiendo el dataset en datos de entrenamiento y datos de test previamente en base a tu criterio y pruébalos frente a dichos datos de test.
- ▶ **Elige al menos una arquitectura de red neuronal (describe en el informe las neuronas en la capa de entrada, las capas intermedias – al menos dos capas**

Asignatura	Datos del alumno	Fecha
Técnicas de inteligencia Artificial	Apellidos:	
	Nombre:	

intermedias – y capa de salida, funciones de activación en cada caso – **al menos utiliza *relu* en algunas de las capas intermedias y utiliza *softmax* en la capa de salida**) que permita realizar una **clasificación**.

- ▶ Mediante Python y utilizando **al menos Keras sobre TensorFlow 2.0 (tensorflow.keras)**, entrena el modelo o modelos de red neuronal escogidos, dividiendo el dataset en datos de entrenamiento y datos de test previamente en base a tu criterio y pruébalos frente a dichos datos de test.
- ▶ Mediante Python y las librerías que consideres, **muestra los resultados** obtenidos por los diferentes algoritmos escogidos de forma gráfica y comparada/superpuesta.
- ▶ Incluye en el informe una **discusión sobre los resultados obtenidos**, argumenta con qué técnica se obtienen mejores resultados en base a las diferentes métricas que hayas escogido y **explica cómo se podrían mejorar los resultados obtenidos por las redes neuronales**, independientemente de que mejoren o no a los algoritmos no basados en redes neuronales.
- ▶ **Recuerda, incluye en el informe a entregar al menos lo siguiente:**
 - Descripción de la fuente de datos empleada.
 - Caracterización del dataset utilizado en modo texto y gráfico.
 - Resultados obtenidos por los diferentes algoritmos escogidos de forma gráfica y comparada/superpuesta.
 - Discusión de los resultados obtenidos y argumentos sobre cómo mejorar de dichos resultados.

NOTA: No está permitido entregar ninguno de los ejemplos ya resueltos en clase. Tampoco está permitido el plagio entre compañeros del mismo o anteriores cursos, ésta no es una actividad grupal. Tampoco está permitido el plagio de fuentes de Internet, libros, manuales o de cualquier otro tipo que no se citen apropiadamente. **Entregar un ejercicio que incurra en los anteriores escenarios implicará automáticamente un 0 como calificación de la actividad, sin perjuicio de que se**

Asignatura	Datos del alumno	Fecha
Técnicas de inteligencia Artificial	Apellidos:	
	Nombre:	

tomen medidas disciplinarias de acuerdo con la correspondiente normativa de la UNIR. Está permitido utilizar código, fracciones de código o ejemplos existentes en Internet u otras fuentes **siempre que se citen apropiadamente y su uso se limite a partes minoritarias de la actividad.**

Rúbrica

Trabajando con redes neuronales y Deep Learning (valor real: 5 puntos)	Descripción	Puntuación máxima (puntos)	Peso %
Criterio 1. Elección de dataset para la regresión	El dataset escogido para la aplicar la regresión cumple los requisitos mínimos especificados y el alumno lo demuestra en su análisis	1	10%
Criterio 2. Ejecución del código de regresión	El código Python utilizado para realizar el apartado de regresión es correcto (se ejecuta sin errores y cumple los objetivos mínimos marcados)	1.5	15%
Criterio 3. Calidad del código de regresión	El código Python utilizado para realizar el apartado de regresión está debidamente comentado y formateado	1	10%
Criterio 4. Elección del dataset para la clasificación	El dataset escogido para la aplicar la clasificación cumple los requisitos mínimos especificados y el alumno lo demuestra en su análisis	1	10%
Criterio 5. Ejecución del código de clasificación	El código Python utilizado para realizar el apartado de clasificación es correcto (se ejecuta sin errores y cumple los objetivos mínimos marcados)	1.5	15%
Criterio 6. Calidad del	El código Python utilizado para realizar el apartado de clasificación está debidamente comentado y formateado	1	10%

Asignatura	Datos del alumno	Fecha	
Técnicas de inteligencia Artificial	Apellidos:		
	Nombre:		
código de clasificación			
Criterio 7. Comprensión de conceptos	La justificación aportada en el informe de resultados demuestra la comprensión de conceptos por parte del alumno	2	20%
Criterio 8. Presentación del informe	Redacción correcta, clara, sin faltas de ortografía y ajustada a la extensión máxima	1	10%
		10	100 %

Extensión máxima de la actividad: un único informe de 4 páginas como máximo de extensión (Arial 11, interlineado 1,5) y dos archivos Python, uno para el problema de regresión y otro para el problema de clasificación. Ambos archivos han de poder ejecutarse de forma independiente el uno del otro. **El informe no debe incluir anexos, portada ni repetir el enunciado o la rúbrica de la actividad.** Cualquier página que incluyas será computada dentro de la extensión entregada. No se evaluará ninguna página a partir más allá de la cuarta página. **No es posible adjuntar ningún archivo adicional.** El propio código Python de los dos archivos ha de ingerir/descargar automáticamente los datos de entrenamiento y test correspondientes a cada uno y trabajar sobre ellos.