

Aplicación de herramientas de Deep Learning en la gestión de datos de la empresa

Nombre:	
Fecha:	
Puntaje ideal:	Puntaje real:
Calificación Final	

Porcentaje corte nota 4,0	
---------------------------	--

INSTRUCCIONES GENERALES:

Estimado/a estudiante, a continuación, se presenta una evaluación parcial diseñada en función de establecer el nivel de conocimientos adquiridos a lo largo de la unidad 1, específicamente lo relacionado con la construcción de una red neuronal sencilla para realizar regresión a través de Python, según el contexto de la industria.

En esta evaluación tendrás que realizar la aplicación de una red neuronal implementada en lenguaje Python e incorporando el framework tensorflow y la librería keras, con el objetivo de realizar predicciones de precio de casas por medio de una red neuronal.

Una vez finalizada la última sesión de la unidad 1, contarás con un tiempo de 72 horas para la realización de esta evaluación, respetando el plazo establecido y siguiendo una serie de pautas que a continuación se le indican.

Ten presente las siguientes consideraciones:

- Esta evaluación representa un 30% de la evaluación total del curso.
- El desarrollo de la evaluación es de tipo individual.
- Revisa los criterios de evaluación para comprender la forma en que se evaluará cada ejercicio.
- El formato de entrega debe ser un archivo Notebook, cuya extensión es .ipynb con el siguiente nombre:
EP_U1_Nombre_Apellidos

Caso Práctico: Regresión lineal

En este ejercicio, trabajaremos con un conjunto de datos que se derivó del censo de EE. UU. de 1990, utilizando una fila por grupo de bloques censales. Un grupo de bloque es la unidad geográfica más pequeña para la que la Oficina del Censo de EE. UU. publica datos de muestra (un grupo de bloque suele tener una población de 600 a 3000 personas). Un hogar es un grupo de personas que residen dentro de una casa. Dado que la cantidad promedio de habitaciones y dormitorios en este conjunto de datos se proporciona por hogar, estas columnas pueden tomar valores sorprendentemente grandes para grupos de bloques con pocos hogares y muchas casas vacías, como centros vacacionales. Es un problema de Regresión. Las variables de entrada que describen a cada entrada son numéricas y tienen escalas variables. A continuación, se enumeran los ocho atributos para el conjunto de datos:

MedInc Ingreso promedio
HouseAge edad promedio de la casa
AveRooms Número promedio de habitaciones por hogar
AveBedrms número promedio de dormitorios por hogar
Population Población

Aplicación de herramientas de Deep Learning en la gestión de datos de la empresa

AveOccup	número promedio de miembros del hogar
Latitude	Latitud
Longitude	Longitud
MedInc	Ingreso promedio

La variable objetivo es el valor medio de la vivienda para los distritos de California, expresado en cientos de miles de dólares (\$100,000).

Características del conjunto de datos:

Número de atributos 8 atributos numéricos, predictivos y el objetivo

Este dataset contiene 20640 registros con atributos de tipo numéricos.

Tu labor será realizar las etapas de preprocesamiento, diseño, entrenamiento y evaluación de una red neuronal multicapa para un problema de regresión que permita predecir el precio de una casa, utilizando un dataset disponible en https://www.dcc.fc.up.pt/~ltorgo/Regression/cal_housing.html

A continuación, ejecute las siguientes operaciones con el dataset de casas de Boston.

1. Ingresa a Google Colab o Jupyter Notebook, carga y realiza la lectura del dataset de casas de Boston.
2. Carga las librerías pandas, matplotlib, numpy, sklearn, tensorflow y keras.
3. Diseña la arquitectura de red multicapa para resolver este ejercicio, argumentando la elección de capas y funciones de activación consideradas en ella.
4. Entrena la red.
5. Evalúa la red y establece sus resultados, para ello, elija la métrica más conveniente e interprete sus valores.
6. Representa gráficamente los resultados obtenidos.
7. Realiza una predicción de posible precio de la casa usando el modelo e interpreta su resultado.

Para importar los datos del dataset puedes usar directamente el siguiente código en Python:

```
from sklearn.datasets import fetch_california_housing
df = fetch_california_housing(as_frame=True)
```

A continuación, podrás revisar la rúbrica donde se especifican cada uno de los aspectos que serán evaluados a partir del encargo indicado.

Categoría	Puntaje	Descripción
Dominio Sobresaliente	4	Se observa un manejo extraordinario del indicador, se considera como la máxima para cualificar ese componente.
Excelente Dominio (ED)	3	Dominio esperado para el indicador, se considera como el punto óptimo para cualificar como competente.
Alto dominio (AD)	2	Se observan algunas dificultades o errores para el completo dominio del indicador.
Dominio Aceptable (DA)	1	Suficiencia de logro en el dominio del indicador, se considera como el mínimo aceptable para cualificar como competente.

Aplicación de herramientas de Deep Learning en la gestión de datos de la empresa

Dominio en Proceso (DP)	0	Se observan dificultades y/o errores para cualificar como competente en el dominio del indicador.
-------------------------	---	---

Criterios de evaluación	CATEGORÍAS			
	ED 3	AD 2	DA 1	DP 0
Carga los datos a procesar	Realiza la carga correcta del dataset a partir de la base de datos asignada al caso	Reconoce las instrucciones para la carga de la base de datos asignada a la actividad, pero no logra asignarla correctamente al dataset	N/A	No carga la base de datos asignada al caso
Importa correctamente las librerías necesarias para la implementación de la red neuronal multicapa	Importa correctamente las 6 librerías necesarias para la implementación de la red neuronal multicapa	Importa correctamente entre 3 y 5 librerías necesarias para la implementación de la red neuronal multicapa	Importa correctamente entre 1 y 2 librerías necesarias para la implementación de la red neuronal multicapa	No importa las librerías necesarias para la implementación de la red neuronal multicapa
Diseña una arquitectura de red multicapa, contemplando las capas y funciones de activación necesarias para la resolución del problema	Diseña correctamente la red multicapa considerando una arquitectura con capas necesarias con las funciones de activación correctas para la resolución del problema	Diseña la red multicapa considerando algunas de las capas ocultas necesarias con las funciones de activación correctas para la resolución del problema	Diseña la red multicapa considerando solo una capa oculta o propone algunas capas, pero con funciones de activación incorrectas para la resolución del problema	No diseña la red multicapa necesaria para la resolución del problema
Argumenta la selección de capas y funciones de activación contempladas en el diseño de la red multicapa propuesta	Argumenta correctamente la selección de todas las capas y funciones contempladas en el diseño de la red multicapa propuesta	Argumenta correctamente la selección de algunas de las capas y/o funciones contempladas en el diseño de la red multicapa propuesta	Argumenta correctamente la selección de solo una de las capas y/o funciones contempladas en el diseño de la red multicapa propuesta	No argumenta la selección de capas y/o funciones contempladas en el diseño de la red multicapa propuesta
Aplica el entrenamiento a la red diseñada estableciendo los parámetros de entrenamientos más convenientes	Aplica correctamente el entrenamiento a la red diseñada estableciendo, todos los	Aplica el entrenamiento a la red diseñada, estableciendo la mayoría de los parámetros convenientes	Aplica el entrenamiento a la red diseñada, pero con parámetros de entrenamiento	No aplica el entrenamiento a la red diseñada

Aplicación de herramientas de Deep Learning en la gestión de datos de la empresa

	parámetros convenientes		poco convenientes	
Evaluación de la red multicapa, identificando métricas adecuadas (perdida y precisión) e interpretando sus valores	Evalúa de manera correcta la red, identificando métricas adecuadas e interpretando sus valores	Evalúa la red multicapa con métricas adecuadas, pero no interpreta los valores.	Evalúa solo con una métrica la red multicapa interpretando su valor	No realiza la evaluación de la red multicapa
Grafica la pérdida y la precisión obtenida de la evaluación del modelo propuesto	Representa correctamente las gráficas de pérdidas y precisión obtenida de la evaluación del modelo propuesto	Representa correctamente solo uno de las dos gráficas solicitadas obtenida de la evaluación del modelo propuesto	Representa alguna de las gráficas solicitadas, pero no se ajusta a la evaluación del modelo propuesto	No representa las gráficas de pérdidas y precisión de la evaluación del modelo propuesto
Realiza una predicción para una casa por medio del modelo elaborado, interpretando el resultado obtenido	Realiza una predicción correcta del precio de la casa por medio del modelo elaborado interpretando de forma correcta el resultado obtenido	Aplica los comandos para la predicción del precio de la casa por medio del modelo, pero no interpreta de manera correcta el resultado obtenido	Aplica algunos comandos, pero no logra la predicción del precio de la casa por medio del modelo elaborado.	No realiza una predicción de la posible precio de la casa por medio del modelo elaborado.
Puntaje por categoría:				
Puntaje total:				