

Taller ANN (Regresión)

El objetivo de este taller es entrenar redes neuronales artificiales para tareas de regresión y maximizar su rendimiento. En cada conjunto de datos (Dataset) retira el 10% de los datos (de manera aleatoria) y cárgalo en un nuevo dataset para probar el rendimiento del modelo y realizar predicciones. El 90% restante divídelo (80-20 en entrenamiento y validación).

1. California Housing Prices

- **Descripción:** Este dataset contiene información sobre casas en California, donde el objetivo es predecir el valor medio de las viviendas en diferentes áreas basadas en características como el número de habitaciones, la ocupación, la proximidad al océano, etc.
- **Tarea:** Usar una ANN para predecir los precios de las casas basados en múltiples características numéricas y categóricas.
- **Fuente:** Viene con Scikit-learn y está disponible en California Housing Dataset.

Ejercicio:

- Cargar el dataset desde Scikit-learn.
- Normalizar los datos.
- Entrenar una ANN con varias capas densas (fully connected layers) para predecir el precio medio de las casas.
- Optimiza el rendimiento de la red neuronal usando los temas vistos en clase (no usar Optuna, Hyperopt o temas no vistos en clase).
- Grafica la curva de aprendizaje de la red neuronal que estás entrenando ¿Tu modelo se sobreajustó?

2. Boston Housing Prices

- **Descripción:** Este dataset contiene características sobre diversas áreas en Boston, como la tasa de criminalidad, la cantidad de habitaciones y el índice de contaminación, con el objetivo de predecir el precio de las casas en dichas áreas.
- **Tarea:** Utilizar una ANN para predecir el precio medio de las viviendas basado en las características del área.
- **Fuente:** Este dataset está disponible en Kaggle - Boston Housing Prices.

Ejercicio:

- Cargar el dataset desde Scikit-learn (o Kaggle).
- Preprocesar los datos.
- Entrenar una ANN para predecir el precio de las casas.
- Optimiza el rendimiento de la red neuronal usando los temas vistos en clase (no usar Optuna, Hyperopt o temas no vistos en clase).
- Grafica la curva de aprendizaje de la red neuronal que estás entrenando ¿Tu modelo se sobreajustó?

3. Air Quality Dataset

- **Descripción:** Este dataset registra los niveles de contaminación del aire y varias características relacionadas, como la temperatura, la humedad y la dirección del viento. El objetivo es predecir la concentración de contaminantes en función de estas variables.
- **Tarea:** Entrenar una ANN para predecir el nivel de contaminación basado en factores meteorológicos y de emisión.
- **Fuente:** Puedes encontrar el dataset en [UCI Machine Learning Repository - Air Quality](#).

Ejercicio:

- Descargar y cargar los datos del UCI Repository.
- Preprocesar las características, incluyendo normalización y manejo de valores faltantes.
- Entrenar una ANN para predecir el nivel de contaminación.

- Optimiza el rendimiento de la red neuronal usando los temas vistos en clase (no usar Optuna, Hyperopt o temas no vistos en clase).
- Grafica la curva de aprendizaje de la red neuronal que estás entrenando ¿Tu modelo se sobreajustó?

4. Concrete Compressive Strength

- **Descripción:** Este dataset contiene información sobre las propiedades de diferentes mezclas de concreto y el objetivo es predecir la resistencia del concreto a la compresión después de 28 días.
- **Tarea:** Usar una ANN para predecir la resistencia a la compresión del concreto basado en las proporciones de sus ingredientes (cemento, agua, arena, etc.).
- **Fuente:** Disponible en [UCI Machine Learning Repository - Concrete Compressive Strength](#).

Ejercicio:

- Cargar el dataset y dividirlo en entrenamiento y prueba.
- Entrenar una ANN con varias capas densas para predecir la resistencia a la compresión.
- Optimiza el rendimiento de la red neuronal usando los temas vistos en clase (no usar Optuna, Hyperopt o temas no vistos en clase).
- Grafica la curva de aprendizaje de la red neuronal que estás entrenando ¿Tu modelo se sobreajustó?