

Deep Learning con grandes volúmenes de datos

Topic: Regression

Total: 40 pts (*en grupo de hasta 3 estudiantes*).

Fecha de entrega: 27 de noviembre de 2023, 21:00 vía gestion.ort.edu.uy

Se presenta un dataset que contiene información de alojamientos publicados en AirBnB con sus respectivos precios. El tamaño del dataset de train es de 1.5 Gb aproximadamente, y 0.5 Gb el de test. Este cuenta con 84 variables predictoras que se podrán utilizar como consideren adecuado.

El objetivo es asignar el precio correcto a los alojamientos listados.

Además del dataset se les provee una notebook (adjunta) conteniendo el script de carga de datos y un **modelo baseline** que corresponde a una arquitectura feed forward.

Consigna:

A - Participación en Competencia Kaggle: El objetivo de este punto es participar en la competencia de Kaggle y obtener como mínimo un Mean Absolute Error inferior a 70 puntos.

[Link a la competencia](#)

B - Utilización de Grid Search (o equivalente): Para cumplir con la búsqueda de modelos óptimos se debe realizar un grid search lo más abarcativo y metódico posible.

Recomendamos enfáticamente [Weights & Biases](#).

C - Investigar e implementar las siguientes técnicas:

- [Batch Normalization](#)
- [Gradient Normalization y/o Gradient Clipping](#)

Además como en todas las tareas se evaluará la prolijidad de la entrega, el preprocesamiento de datos, visualizaciones y exploración de técnicas alternativas.

