

**Escuela de Ingeniería Informática  
Inteligencia Artificial**

Profesor: Guillermo Cabrera G.

**Fecha Entrega: Lunes 15 de Mayo 2023**

El siguiente proyecto corresponde al módulo de Regresión Lineal. El trabajo es grupal y los grupos deben ser de entre 4 y 5 personas.

El dataset a utilizar corresponde a los atributos de un conjunto de vinos portugueses, originalmente presentados en [1]. Los archivos “winequality-red.csv” y “winequality-white.csv” corresponde a datos de vinos rojos y blancos (respectivamente) de la viña, donde los atributos de los vinos (inputs) están explicados en un segundo archivo de texto llamado “winequality.names”.

Para estos dataset se solicita:

- Análisis exploratorio de los datos, que muestren las relaciones que existen entre los atributos y la variable que se busca predecir.
- Seleccione los 3 atributos que, a su juicio y basado solamente en el análisis exploratorio de datos realizado previamente, expliquen mejor la variable que se busca predecir (calidad del vino) y, por cada uno de ellos, ejecute una regresión lineal simple usando batch gradient descent y otra con stochastic gradient descent, y compárelas en función de sus métricas de evaluación. Considere métricas como  $R^2$  y  $R^2$  – ajustado en su análisis.
- Realice una regresión lineal usando el gradiente estocástico e indique las principales diferencias entre los resultados obtenidos con este algoritmo y aquellos encontrados con el algoritmo batch.

La entrega corresponde a un notebook donde, al menos, se incluye:

- EDA
- Imágenes de gráficos utilizados (con sus respectivos códigos)
- Tablas reportando los resultados obtenidos
- Reporte de métricas revisadas en clases
- Interpretación y discusión de resultados

## Referencias

[1] P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis. Modeling wine preferences by data mining from physicochemical properties. In Decision Support Systems, Elsevier, 47(4):547-553. ISSN: 0167-9236.