

Curso: Redes Neuronales

Examen Final

Se desea desarrollar un modelo que permita inferir de manera automática la calidad de un vino a partir de una muestra. Para esto, se ha construido un dataset con información de mediciones fisicoquímicas para un conjunto de muestras de vino blanco y tinto provenientes de diferentes viñedos de la región de San Juan. Para cada muestra se dispone de información de 12 variables: *fixed_acidity*, *volatile_acidity*, *citric_acid*, *residual_sugar*, *chlorides*, *free_sulfur_dioxide*, *total_sulfur_dioxide*, *density*, *pH*, *sulphates*, *alcohol*. Además, se indica el nivel de calidad del vino (*quality*, evaluada por un conjunto de sommelier experimentados) y si se trata de una muestra de vino blanco (*white_wine*). La calidad del vino se encuentra representada por un índice que toma valores enteros en el rango 1 – 5, siendo 5 el valor de calidad para un vino de exportación. Este dataset se provee como archivo CSV bajo el nombre *winequality.csv*.

En base a esta información, se solicita que:

- Construya un modelo neuronal capaz de inferir el índice de calidad del vino a partir de las variables fisicoquímicas medidas. Se recomienda la exploración de diferentes arquitecturas y la optimización de los hiperparámetros para lograr un buen desempeño. Recuerde realizar un particionado adecuado de los datos, que le permita llevar adelante todas las etapas del proceso.
- Evalúe la capacidad de generalización del modelo desarrollado, empleando alguna estrategia de validación cruzada.
- Redacte un informe (no más de 6 páginas A4) con la descripción del proceso seguido para la construcción del modelo neuronal, incluyendo hipótesis planteadas y justificación de las decisiones tomadas durante el proceso (arquitectura, particionado de datos, ajuste de hiperparámetros, validación, etc.). El informe deberá contener la matriz de confusión y las figuras de mérito adecuadas para analizar y sustentar su trabajo. Se valorará la capacidad de síntesis, por lo que se recomienda ser conciso en las explicaciones.
- Implemente una función que permita utilizar el modelo ya entrenado. La misma debe ser capaz de leer un archivo **input** en formato CSV (similar al provisto para el examen), y debe retornar un archivo **output** en formato CSV como salida (una única columna), con el índice de calidad de cada vino contenido en el archivo **input**. Evalúe la capacidad de generalización del modelo desarrollado empleando el dataset *winequality.test.csv*. Adjunte la función y el archivo de salida al informe (los índices de calidad de este dataset serán provistos una vez finalizado el examen).
- Evalúe si es posible identificar, a partir de las variables fisicoquímicas (o un subconjunto de ellas), si una muestra corresponde a vino tinto. Describa brevemente el análisis realizado. En base a los resultados alcanzados: ¿Es necesario contar con la variable *white_wine* para construir el modelo de inferencia para el índice de calidad de vino?. Explique.