



Práctica 1 - Regresión

La práctica 1 de la asignatura *Machine Learning* consiste en la resolución de un problema de **Regresión**. El problema a resolver se describe a continuación.

Requisitos de la práctica

La práctica consta de los siguientes entregables:

1. Una libreta de *Python* en la cual se realizará el desarrollo de la práctica. La libreta incluirá todo el código de las operaciones realizadas para el preprocesamiento, entrenamiento y validación de los modelos, así como una sección de conclusiones en la cual se interpretarán no solo las medidas de calidad obtenidas por el modelo sino también el modelo en sí mismo.
2. Un vídeo corto de alrededor de 15 minutos en el cual se presenten los desarrollos, resultados y conclusiones obtenidas para el problema resuelto. Para la entrega del vídeo, se subirá a la nube y se entregará el enlace.

Será **imprescindible** realizar los siguientes procesos durante la resolución del problema:

- Preprocesamiento adecuado de los conjuntos de datos.

- Entrenamiento y validación de **todos** los modelos de regresión que se han visto en clase: regresión lineal, Ridge, Lasso y ElasticNet.
 - Optimización con *Grid Search* de los hiperparámetros de cada uno de los modelos anteriores.
 - Validación cruzada con *5 folds* de cada uno de los modelos entrenados, utilizando las siguientes métricas de calidad: *RMSE*, *MAE* y R^2 .
 - De cara a evaluar justamente, la semilla para todos aquellos métodos estocásticos será: `random_state=1337`.
-

Descripción del problema

Disponemos de un conjunto de datos con información variada sobre ventas de vehículos de segunda mano en el Reino Unido.

El problema de regresión a resolver consiste en predecir el precio de venta de un vehículo de segunda mano en función de sus características.

El conjunto de datos

Para esta práctica usaremos un conjunto de datos recopilados de listas de venta de coches usados. Se dispone de información sobre 100000 vehículos de segunda mano.

El conjunto de datos está formado por varios ficheros CSV, uno para cada fabricante de coches. Cada fila del archivo corresponde a una venta de un vehículo de segunda mano.

Las características (o *features*) de este conjunto de datos son las siguientes:

- `model` - modelo del vehículo
- `year` - año de matriculación
- `price` - precio de venta en libras esterlinas. **Variable objetivo**
- `transmission` - tipo de transmisión
- `mileage` - millas de uso
- `fuelType` - tipo de carburante
- `tax` - impuesto de circulación

- `mpg` - consumo del vehículo en millas recorridas por galón
- `engineSize` - cilindrada del motor en litros