

## UNIDAD TEMÁTICA 3: Algoritmos Lineales

### Trabajo de Aplicación 2

#### Caso de Estudio: Predicción del precio de venta de una casa, a partir de varios predictores

El dataset completo está disponible para descargar del repositorio de UCI:

<https://archive.ics.uci.edu/ml/machine-learning-databases/housing/>

Los objetivos de este **Trabajo de Aplicación** son:

1. Identificar cuáles atributos, entre los varios disponibles, son necesarios para predecir con exactitud la mediana de precios de una casa
2. Construir un modelo de regresión lineal múltiple para predecir la mediana de los precios utilizando los atributos más importantes
3. Evaluar la exactitud del modelo para predecir nuevos ejemplos

Debido a la naturaleza del enfoque de ajuste de funciones, una limitación importante que encontramos tiene que ver con la dimensionalidad. A medida que la cantidad de atributos o predictores crece, se reduce nuestra capacidad para obtener un buen modelo, pero además se agrega complejidad computacional y también se hace más difícil la interpretación del modelo.

Revisaremos aquí algunos métodos de selección de características – “*feature selection*” – que permitan reducir el número de predictores al mínimo posible sujeto a obtener un buen modelo.

Utilizando RapidMiner, veremos cómo realizar la preparación de los datos, la construcción del modelo y la validación. Finalmente revisaremos que se cumplan algunos requerimientos para asegurar que la regresión lineal se utiliza correctamente.

#### Ejercicio 1 (15 minutos + 3 de preguntas)

Descargar el dataset de UCI.

Analizar los atributos, y describir en un archivo de texto (para cada uno)

- su contexto y significado,
- tipos de datos y rangos
- distribuciones y outliers
- ¿cuál es la variable de salida?

RESPONDER PREGUNTAS PROYECTADAS EN PANTALLA