

Práctica 2. Regularización y Selección de Modelos

Objetivo

El objetivo es aplicar y comparar técnicas de regularización y validación cruzada para selección de modelos en una aplicación real de regresión lineal. Se trata de encontrar el mejor modelo polinómico para predecir el precio de un coche de segunda mano en función de tres atributos: antigüedad, kilómetros y potencia.

Estudio previo (es necesario prepararlo **por escrito, antes** de acudir a la práctica)

- A. Repasa las transparencias de clase y estudia las funciones auxiliares proporcionadas para esta práctica. Escribe el algoritmo de k-fold cross-validation para elegir el grado de una regresión polinómica o el valor del parámetro de regularización. Ten en cuenta que para resolver la regresión es necesario realizar la expansión polinómica de los atributos y su normalización.

Desarrollo de la práctica

Copia a tu directorio de trabajo los ficheros proporcionados, y comprueba que funcionan correctamente en Matlab. A continuación, escribe los programas necesarios para resolver la regresión polinómica con los datos de precios de coches de segunda mano, siguiendo los siguientes pasos:

1. **Selección de modelos mediante búsqueda heurística.**
 - a. Programa el algoritmo de k-fold cross-validation para la búsqueda del mejor modelo (grados de cada polinomio).
 - b. Empieza con todos los atributos con grado 1, y ve buscando el mejor grado para cada atributo en el orden que te parezca más adecuado. Para cada atributo, dibuja las curvas de evolución de los errores RMSE de entrenamiento y de validación, y analízalas. Fija el mejor grado obtenido y repite para el atributo siguiente.
 - c. Reentrena el modelo final seleccionado con todos los datos y evalúalo. Discute los resultados.
2. **Selección de modelos mediante búsqueda exhaustiva (*grid search*).** Programa la búsqueda exhaustiva del mejor modelo polinómico mediante k-fold cross-validation. Reentrena el mejor modelo con todos los datos y evalúalo. Discute los resultados.
3. **Regularización.** Programa la búsqueda de un modelo de regresión polinómica de grado 10 para los tres atributos, con regularización. Elige el mejor valor de λ mediante k-fold cross-validation. Dibuja las curvas de evolución de los errores RMSE de entrenamiento y de validación, y analízalas. Reentrena el mejor modelo con todos los datos y evalúalo. Discute los resultados.
4. Escribe en la memoria un breve **resumen de las conclusiones** principales obtenidas.

A entregar en Moodle

- Memoria de la práctica en un fichero `P2.pdf` con el estudio previo escaneado, los resultados de todos los apartados, su interpretación y las conclusiones que hayas obtenido.
- Programa `P2.m` que vaya ejecutando cada apartado y mostrando por pantalla los resultados obtenidos, junto con las funciones auxiliares que hayas programado, en un fichero `.zip`

Recuerda que:

- Tienes 5 días desde tu sesión de prácticas para depositarla en Moodle.
- Deberás defender la práctica en tu próxima sesión de prácticas.