

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. Se trata de encontrar el mejor modelo polinómico o con redes neuronales 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 teoría y estudia a fondo las funciones de Scikit-Learn que se indican en la última transparencia.

Desarrollo de la práctica

Copia los ficheros proporcionados, y comprueba que `P2demo.py` funciona bien. Observa las gráficas donde se muestran los datos. A continuación, escribe los programas necesarios para **entrenar, evaluar y comparar entre sí distintos modelos** de regresión para predecir el precio de los coches:

1. **Regresión lineal.** Como punto de partida, obtén un modelo de regresión lineal, evalúalo y discute los resultados. Analiza si sufre subajuste o sobreajuste.
2. **Selección del grado del polinomio.** Busca el mejor modelo polinómico con grado entre 1 y 10 por validación cruzada con k-fold. Dibuja las curvas de evolución de los errores RMSE de entrenamiento y de validación conforme varía el grado, y analízalas.
3. **Regresión polinómica regularizada.** Fija el grado del polinomio a 10, y aplica regresión regularizada. Busca el mejor valor del parámetro de regularización por validación cruzada con k-fold. Dibuja las curvas de evolución de los errores RMSE de entrenamiento y de validación conforme varía la regularización, y analízalas.
4. **Regresión con redes neuronales.** Prueba una red sin capa oculta, y otra con una capa oculta de 100 neuronas. Evalúalas y analiza y discute los resultados.
5. **Regresión con redes regularizada.** Con la red de una capa oculta de 100 neuronas, busca el mejor valor del parámetro de regularización por validación cruzada con k-fold. la búsqueda de un modelo de regresión polinómica de grado 10 para los tres atributos, con regularización. Dibuja las curvas de evolución de los errores RMSE de entrenamiento y de validación conforme varía la regularización, y analízalas.
6. Escribe en la memoria un **resumen de las conclusiones** principales obtenidas.

A entregar en Moodle

- Un notebook `P2.ipynb` con los programas de cada apartado, los resultados, su interpretación y las conclusiones que hayas obtenido.
- O, si lo prefieres, un fichero `P2.py` con el código y otro `P2.pdf` con los resultados de todos los apartados, su interpretación y las conclusiones que hayas obtenido.

Recuerda:

- Trae la práctica preparada para aprovechar la sesión de prácticas al máximo.
- Si te atascas, pregunta en la sesión o en tutorías.
- Debes citar correctamente todas las fuentes utilizadas.
- Tienes 5 días desde tu sesión para depositar la práctica en Moodle.
- Deberás defenderla en tu próxima sesión de prácticas.