

Práctica 7: Valoración y selección de modelos de regresión

Módulo de Modelos Lineales.
Máster de Bioestadística, Universitat de València.

Miguel A. Martinez-Beneito

Tareas

1. Para el modelo de regresión del banco de datos `ISLR::Auto` lleva a cabo las siguientes actividades.
 - Valora, cuando sea necesario, la transformación de las covariables que te parezca oportuna para incluirlas en un modelo de regresión que explique `mpg`.
 - Considera todas las variables del banco de datos `Auto`, salvo `mpg` y `name`, como covariables de un modelo lineal sin interacciones para explicar `mpg`. Incluye las variables en dicho modelo con la transformación que te haya parecido adecuada en el apartado anterior. Selecciona las variables del modelo que estimes oportunas mediante un procedimiento de búsqueda exhaustiva de modelos. Compara los resultados obtenidos según el criterio de comparación de modelos (R^2 ajustado, C_p y BIC).
 - ¿Cambiaría el modelo elegido si emplearas cualquiera de los métodos stepwise introducidos en la parte teórica?
 - ¿Es anidado el conjunto de modelos devuelto por la selección exhaustiva de variables? ¿coincide dicho conjunto de modelos para alguno de los métodos stepwise implementados? ¿Qué método stepwise ofrece el conjunto de modelos más similar al ajuste exhaustivo?
 - Para las variables que hayas elegido en los apartados anteriores, considera la interacción de cada una de ellas con la variable `origin`. Utiliza el método exhaustivo de selección de variables para evaluar qué interacciones deberías mantener en el modelo. Compara los resultados obtenidos con lo que obtendrías evaluando la introducción, una a una, de las interacciones consideradas en este apartado ¿Con qué modelo te quedarías finalmente?
 - Para el modelo que hayas determinado en el apartado anterior valora la hipótesis de linealidad para todas las covariables (numéricas) incluidas en el modelo (salvo las interacciones) ¿Te sugieren los resultados obtenidos alguna forma de mejorar el modelo? Si es así considera dicho modelo.
 - Para el modelo que hayas determinado en el apartado anterior valora las hipótesis del modelo de regresión lineal ¿Consideras adecuado dicho modelo de regresión? ¿Encuentras algún vehículo que pudiera haber condicionado el modelo de regresión?
 - Finalmente, plantea qué medida propondrías para mejorar el modelo de regresión para `mpg`.