

Trabajo de evaluación Temas 3-5
Fundamentos de Estadística y Programación en R

Fichero de datos Cargar el paquete: AppliedPredictiveModeling

Ejecutar: data(solubility)

Se cargan seis ficheros de datos: solTrainX, solTraintransX, solTrainY, solTestX, solTestXtrans, solTestY

Construir un fichero "solTrainred" eliminando de "solTraintransX" todas las variables "FP..." que ocupan las 208 primeras posiciones.

Construir el fichero "sol" añadiendo a "solTrainred" la variable "solTrainY".

Las variables de "solTrainred" serán las variables regresoras y "solTrainY" la variable respuesta.

Descripción del trabajo

- Cree un fichero en código R (comentado), denominado nombredelalumno.R en el que se obtenga respuesta a cada uno de las cuestiones planteadas a continuación y envíelo mediante adjunto a jlmorero@us.es.

Nota: también podría ser un fichero Rmarkdown (Rmd).

- Fecha límite: 5 de mayo

Cuestiones

1. Determina el modelo con una y dos variables regresoras que mejor explica la variable respuesta. Denominaremos a estos modelos M1 y M2, respectivamente.
2. En los modelos M1 y M2
 - (a) Gráfico de dispersión de los regresores frente a la respuesta.
 - (b) Obtener los EMC de los parámetros, los intervalos de confianza para los parámetros y los p-valores asociados a los test.
 - (c) ¿Qué conclusiones se obtienen de los p-valores resultantes?
3. En el modelo M2, determina la estimación del valor medio de la respuesta, el intervalo de confianza y el intervalo de predicción para (10, 25) de los regresores. (Justificar que (10, 25) es un valor apropiado para realizar lo que se pide)

Presentar los datos resultantes en una tabla del tipo

	Estimación	IC(95%)	IP(95%)
(a, b)			

Interpreta la diferencia en amplitud de IC(95%) e IP(95%).

4. Sea M2int el modelo resultante de añadir a M2 la interacción entre sus dos regresores.

Construye la siguiente tabla,

	M1	M2	M2int
R^2			
R^2 ajustado			
Residual standard error			

y comenta los resultados.

5. Construye el modelo de regresión lineal con todos los regresores disponibles en el fichero (Mtodas).
 - (a) Interpretar los resultados obtenidos en comparación con los de los modelos considerados en los apartados anteriores.
 - (b) Calcula los vif (factores de inflación de la varianza) de los regresores. ¿Qué conclusión se obtiene?
 - (c) Obtén los gráficos de diagnósticos y comenta los resultados
6. Determina el modelo resultante de una regresión paso a paso hacia adelante. Sea MHA el modelo resultante.
 - (a) Sean v1 y v2 las dos primeras variables que entran en MHA.

Realiza un gráfico para ver como varía R^2 ajustado en los modelos MHA, $MHA+(v1)^2$, $MHA+(v1)^3$, $MHA+(v2)^2$, $MHA+(v2)^3$ y $MHA+(v1 \times v2)$.

7. Puedes añadir cualquier otro análisis que consideres de interés.