



Iași, 16 noiembrie 2018

Tema 4*

Exercițiu 1

Se consideră setul de date "House" ce poate fi descărcat la adresa

<http://thor.info.uaic.ro/~cgatu/csia/res/house.dat> (Sursa: Long-Kogan Realty, Chicago, USA).

Cele 14 coloane corespund, în ordine, următoarelor variabile:

Y	PRICE	Selling price of house in thousands of dollars
X_1	BDR	Number of bedrooms
X_2	FLR	Floor space in sq.ft.
X_3	FP	Number of fireplaces
X_4	RMS	Number of rooms
X_5	ST	Storm windows (1 if present, 0 if absent)
X_6	LOT	Front footage of lot in feet
X_7	TAX	Annual taxes
X_8	BTH	Number of bathrooms
X_9	CON	Construction (0 if frame, 1 if brick)
X_{10}	GAR	Garage size (0 = no garage, 1 = one-car garage, etc.)
X_{11}	CDN	Condition (1 = 'need work', 0 otherwise)
X_{12}	L1	Location (L1 = 1 if property is in zone A, L1 = 0 otherwise)
X_{13}	L2	Location (L2 = 1 if property is in zone B, L2 = 0 otherwise).

1. Să se determine pentru fiecare dimensiune p de la 1 la $n = 13$ modelul de regresie având p variabile și care optimizează, respectiv, *Residual Sum of Squares* (RSS), R^2 , R_a^2 și C_p .
2. Generați graficele celor 4 criterii mai sus menționate pentru dimensiunile $1, \dots, n$.
3. Selectați, conform rezultatelor obținute la punctele anterioare cel mai bun (sub)model de regresie.

*Termen de prezentare: vineri, 7 decembrie 2018.