

Tarea-examen 5-6: selección de modelos y regularización

César Cossio Guerrero. Aldo Sayeg Pasos Trejo

Statistical Machine Learning
Profa. Guillermina Eslava
Posgrado en Ciencias Matemáticas
Universidad Nacional Autónoma de México

13 de mayo de 2020

Problema 1: Modelo normal

R^2	aic	bic	Aparent error	Non aparent error	Non apare error std
0.9373	1164.0844	1548.6992	0.4078	0.7629	0.0596

Cuadro 1: Estadísticas del modelo base. Los errores no aparentes se calcularon mediante validación cruzada con $k = 5$ y 100 repeticiones

Problema 1: Modelos seleccionados

	model	Dfs	details
1	step (AIC)	52	-
2	step (BIC)	32	-
3	Best subset	20	Selección hacia adelante (forward), iniciando en el mejor modelo de 1 variable y finalizando en el mejor con 20
4	Lasso	25	Se usó $\lambda_{1se} = 0.0286$
5	Ridge	92	Se usó $\lambda_{1se} = 0.5232$
6	Elastic Net	31	Se usó $\lambda_{1se} = 0.0521$
7	Random Forest	91	NBoot = 100, m = 91

Cuadro 2: Detalles de cada modelo

Problema 1: λ en modelos regularizados

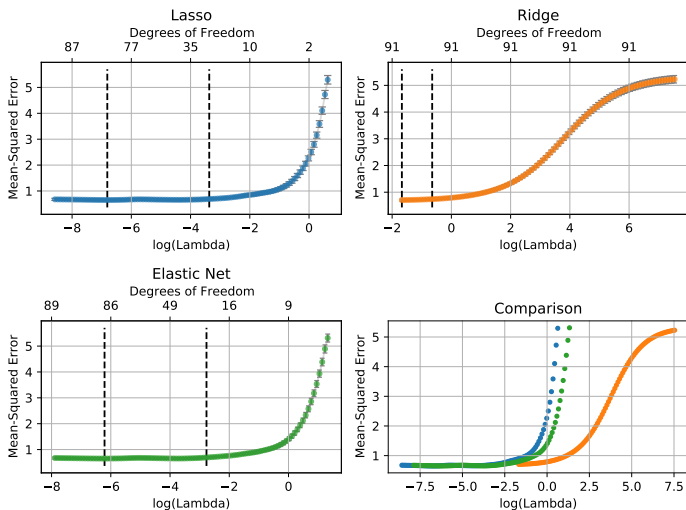


Figura 1: Error cuadrático medio para los modelos

Problema 1: Variables de cada modelo

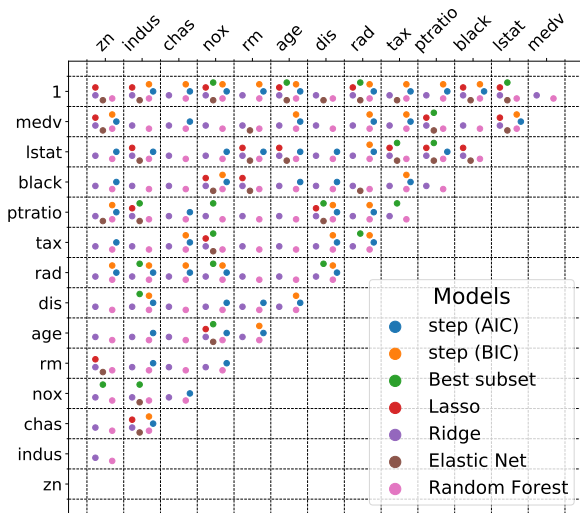


Figura 2: Variables para cada modelo

Problema 1: Errores de cada modelo

model	Dfs	R^2	Aparent error	Non apa- rent error	Non apa- rent error std
step (AIC)	52	0.9199	0.4245	0.5671	0.0255
step (BIC)	32	0.9107	0.4733	0.5564	0.0143
Best subset	20	0.8860	0.6043	0.6641	0.0129
Lasso	25	0.8871	0.5983	0.6749	0.0126
Ridge	92	0.8824	0.6234	0.7304	0.0125
Elastic Net	31	0.8881	0.5934	0.6772	0.0129
Random Forest	91	0.9929	0.0378	0.3039	0.0123

Cuadro 3: Errores para cada modelo

Problema 1: Modelos seleccionados por step

model	Selection outside CV	Selection outside CV std	Selection inside CV	Selection inside CV std
step (AIC)	0.5671	0.0255	1.5123	0.1992
step (BIC)	0.5564	0.0143	1.3623	0.3950

Cuadro 4: Errores incluyendo proceso de selección

Problema 2: Modelo Ridge

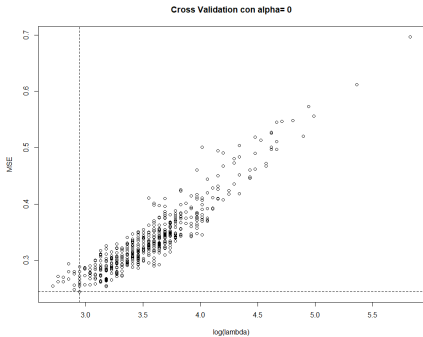
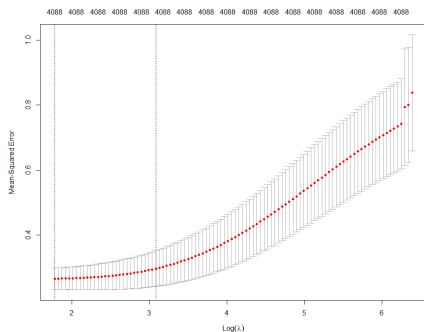


Figura 3

Problema 2: Modelo Elastic Net

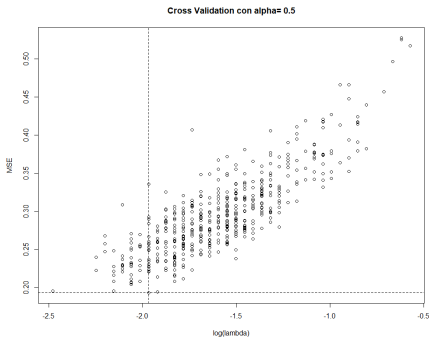
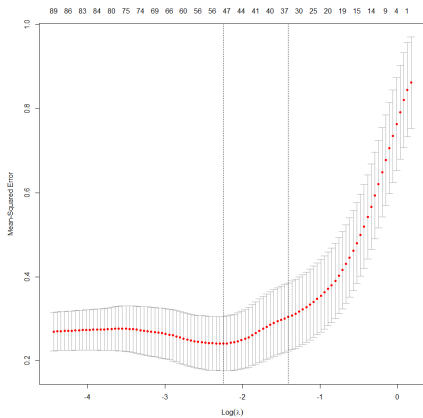


Figura 4

Problema 2: Modelo Elastic Net

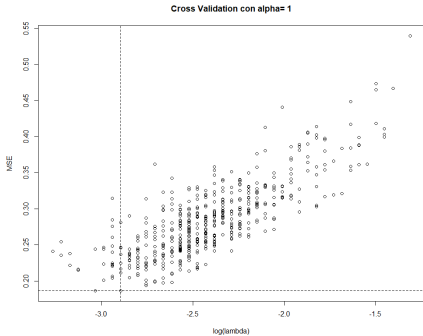
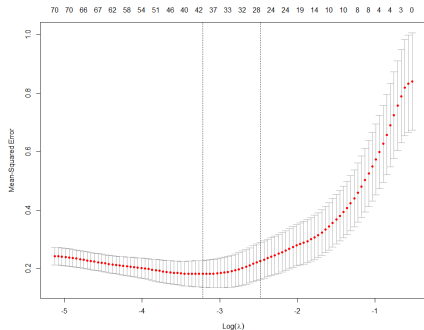


Figura 5

Problema 2: Modelos regularizados

Modelo	mse_{min}	mse_{1se}	mse_{RCV}	λ_{min}	λ_{1se}	λ_{RCV}	df_{min}	df_{1se}	df_{RCV}
Ridge	0.26	0.29	0.24	5.93	21.82	18.98	4089	4089	4089
Elastic Net	0.24	0.30	0.19	0.10	0.24	0.13	49	34	45
Lasso	0.18	0.22	0.18	0.039	0.08	0.05	40	29	34

Cuadro 5: Errores por validación cruzada

Problema 2: Errores de cada modelo

Modelo	Error aparente	Error no aparente	Error no aparente std
Ridge	0.079	0.29	0.027
Elastic Net	0.056	0.25	0.034
Lasso	0.056	0.24	0.028
Modelo nulo	0.83	0.86	0.018

Cuadro 6