

# Selección de modelos y Regularización

19/01/2021

# Ejercicio 1

Modelos no regularizados

Boxcox Lamda=0.02

- Aditivo:  $= \log(\text{crim}) \sim$ .
- Step k=2:  $= \log(\text{crim}) \sim \text{zn} + \text{indus} + \text{nox} + \text{age} + \text{rad} + \text{ptratio} + \text{black} + \text{lstat} + \text{medv}$ .
- Step k=log(506)/regsubset:  $= \log(\text{crim}) \sim \text{zn} + \text{nox} + \text{age} + \text{rad} + \text{black} + \text{lstat}$ .

	Aditivo	Step k=2	Step k=log(n)/Regsubsets
ECM Ap	0.58	0.58	0.596
ECM Val	0.623	0.599	0.607
p	13	9	6
R <sup>2</sup>	0.875	0.875	0.872
R <sup>2</sup> Adj	0.872	0.873	0.871
AIC	1191.48	1184.10	1190.06
BIC	1254.88	1230.59	1223.87

Table 1: **Estadísticas e información adicional de los modelos no regularizados.** *Vemos que las estadísticas de los modelos no difieren en gran tamaño unas de otras, a pesar de contar con diferentes números de variables*

## Tasas Errores

Modelos	Error aparente	Error prueba
R. Aditiva	58.1	62.3
Step k=2	58.2	59.9
Step $k=\log(n)/\text{Regsubsets}$	59.6	60.7
Lasso	58.5	61.9
Ridge	60.4	63.2
Elasticnet	58.2	61.7

Table 2: **Tasas error en porcentajes de modelos ajustados.** *Las primeras tres tasas pruebas fueron calculadas por training/ test con 300 repeticiones, las últimas tres con cv  $k=10$ .*

# Modelos Regularizados

	TrnErr(Ap)	cv(k=10, B=1)	cvsd	betas(df)	lamda.min	lamda.lse
Lasso	0.587	0.608	0.033	9	0.021	0.124
Ridge	0.603	0.635	0.0347	13	0.184	0.425
Elasticnet	0.583	0.620	0.039	11	0.0183	0.206

Table 3: **Estadísticas e información adicional de modelos regularizados.** *Vemos que las cvsd son similares para todos los modelos pero difieren en las lamdas, siendo la más pequeña mínima para elasticnet y la más pequeña lse para Lasso.*

Modelos	Inter	zn	indus	chas1	nox	rm	age	dis	rad	tax	ptratio	black	lstat	medv
R. Aditiva	-3.73	-0.01	0.02	-0.04	3.84	-0.04	0.00	-0.00	0.14	0.00	-0.04	-0.00	0.03	0.01
Step k=2	-4.10	-0.01	0.02		3.86		0.00		0.14		-0.04	-0.00	0.03	0.00
Step k=log(n)	-4.83	-0.01			4.71		0.01		0.14			0.00	0.03	
Lasso		-0.01	0.02		3.94		0.01	-0.02	0.14		-0.02	0.00	0.02	
Ridge		-0.01	0.01	0.02	3.46	-0.02	0.01	-0.04	0.10	0.00	-0.02	0.00	0.03	0.01
Elasticnet		-0.01	0.02		3.86	-0.01	0.01	-0.02	0.14		-0.03	0.00	0.03	0.00

Table 4: **Valor de los coeficientes** asociados a los siete modelos seleccionados, recordemos que Step k=log(n) es el mismo modelo que Regsubsets.

## Ejercicio 2

- *ridge* ( $\alpha = 0$ )
- *lasso* ( $\alpha = 1$ )
- *elastic net* ( $\alpha = 0.5$ )

	<i>ridge</i>	<i>lasso</i>	<i>elastic net</i>
ECM aparente	0.028	0.042	0.037
ECM $cv(k=10, B=1)$	0.262	0.200	0.216
ECM $cv(k=10, B=100)$	0.262	0.210	0.220
$\lambda_{minECM}$	5.934	0.042	0.069

	<i>ridge</i>	<i>lasso</i>	<i>elastic net</i>
$\lambda_{minECM}$	5.93	0.042	0.069
Num. betas ( $\lambda_{minECM}$ )	4089	38	57
$\lambda_{1se}$	30.22	0.080	0.161
Num. betas ( $\lambda_{1se}$ )	4089	29	42
$\lambda_{maxdf}$	-	0.006	0.012
Num. betas ( $\lambda_{maxdf}$ )	-	72	90

Table 7: Información adicional de los modelos. Valores de  $\lambda$  que cumplen tres criterios (lambda.min, lambda.1se, dfmax) y su correspondiente número de coeficientes que son distintos de cero..

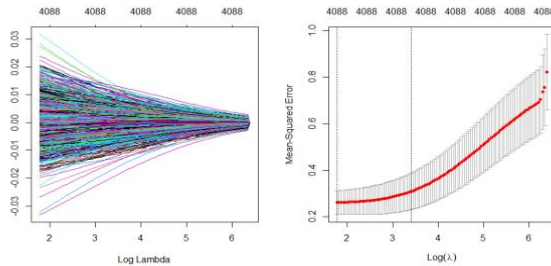
# Selección de Lambda

ridge: 6		lasso: 29		elastic net: 42	
xLYSC_at	-0.033	xYOAB_at	-0.784	xYOAB_at	-0.563
xYEBC_at	-0.032	xYEBC_at	-0.531	xYEBC_at	-0.497
xYFIT_at	0.032	xLYSC_at	-0.270	xLYSC_at	-0.396
xSPOVAA_at	0.028	xSPOVAA_at	0.247	xSPOVAA_at	0.283
xYBFI_at	0.027	xYQJU_at	0.213	xYQJU_at	0.198
xHUTP_at	-0.027	xYXLD_at	0.197	xYBFI_at	0.173

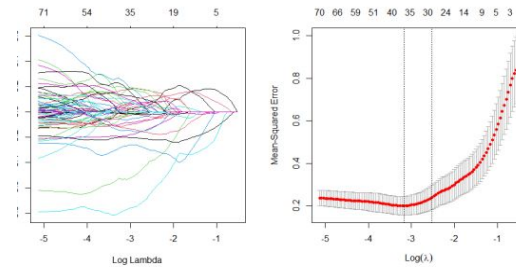
Tabla 6: Coeficientes con mayor valor absoluto para cada uno de los modelos y su variable asociada.

Decidimos priorizar el poder predictivo por lo cual se eligió el valor de lambda tal que minimiza el ECM..

Ridge



Lasso



Elastic net

