Tarea 1
Máquinas de Aprendizaje

Matías Ramírez
Javier Rodríguez

Pregunta 1

...

Regresión Lineal Ordinaria (LSS)

Parte A: Cargar los Datos

1	-0.579818495	2.769459	50	-1.38629436	0	-1.38629436	6	0	-0.4307829	T
2	-0.994252273	3.319626	58	-1.38629436	0	-1.38629436	6	0	-0.1625189	T
3	-0.510825624	2.691243	74	-1.38629436	0	-1.38629436	7	20	-0.1625189	T
4	-1.203972804	3.282789	58	-1.38629436	0	-1.38629436	6	0	-0.1625189	T
5	0.751416089	3.432373	62	-1.38629436	0	-1.38629436	6	0	0.3715636	T
6	-1.049822124	3.228826	50	-1.38629436	0	-1.38629436	6	0	0.7654678	T
7	0.737164066	3.473518	64	0.61518564	0	-1.38629436	6	0	0.7654678	F
8	0.693147181	3.539509	58	1.53686722	0	-1.38629436	6	0	0.8544153	T
9	-0.776528789	3.539509	47	-1.38629436	0	-1.38629436	6	0	1.0473190	F
10	0.223143551	3.244544	63	-1.38629436	0	-1.38629436	6	0	1.0473190	F
11	0.254642218	3.604138	65	-1.38629436	0	-1.38629436	6	0	1.2669476	T

Parte B: Descripción de los Datos

	lcavol	lweight	age	lbph	svi	lcp	gleason	pgg45	lpsa
count	97.000000	97.000000	97.000000	97.000000	97.000000	97.000000	97.000000	97.000000	97.000000
mean	1.350010	3.628943	63.865979	0.100356	0.216495	-0.179366	6.752577	24.381443	2.478387
std	1.178625	0.428411	7.445117	1.450807	0.413995	1.398250	0.722134	28.204035	1.154329
min	-1.347074	2.374906	41.000000	-1.386294	0.000000	-1.386294	6.000000	0.000000	-0.430783
25%	0.512824	3.375880	60.000000	-1.386294	0.000000	-1.386294	6.000000	0.000000	1.731656
50%	1.446919	3.623007	65.000000	0.300105	0.000000	-0.798508	7.000000	15.000000	2.591516
75%	2.127041	3.876396	68.000000	1.558145	0.000000	1.178655	7.000000	40.000000	3.056357
max	3.821004	4.780383	79.000000	2.326302	1.000000	2.904165	9.000000	100.000000	5.582932

Parte C: Normalización de los Datos

Problema:

- Rango de valores distintos entre cada variable
- Naturaleza y distribución de cada columna es diferente del resto

Se normalizan los datos:

- Arreglar varianzas
- Centrar datos
- Crear modelos correctamente

Parte D: Creación del Modelo

- Columna de intercepto (unos)
  - Forma matricial
- Atributo *fit_intercept*
- Columna y no normalizada

Parte E: Pesos y Z-score de Predictores

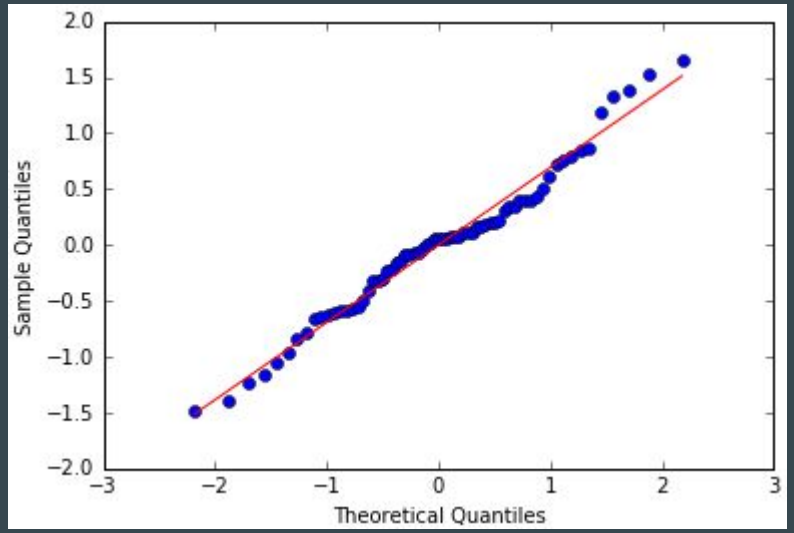
lcavol	0.676	5.32
lweight	0.262	2.727
age	-0.141	-1.384
lbph	0.209	2.038
svi	0.304	2.448
lcp	-0.287	-1.851
gleason	-0.021	-0.145
pgg45	0.266	1.723
intercept	2.465	27.359

Parte F: Estimación del Error del Modelo

Validación cruzada:

- k = 5: mse = 0.9565
- k = 10: mse = 0.7572

Parte G: Hipótesis de Normalidad



Pregunta 2

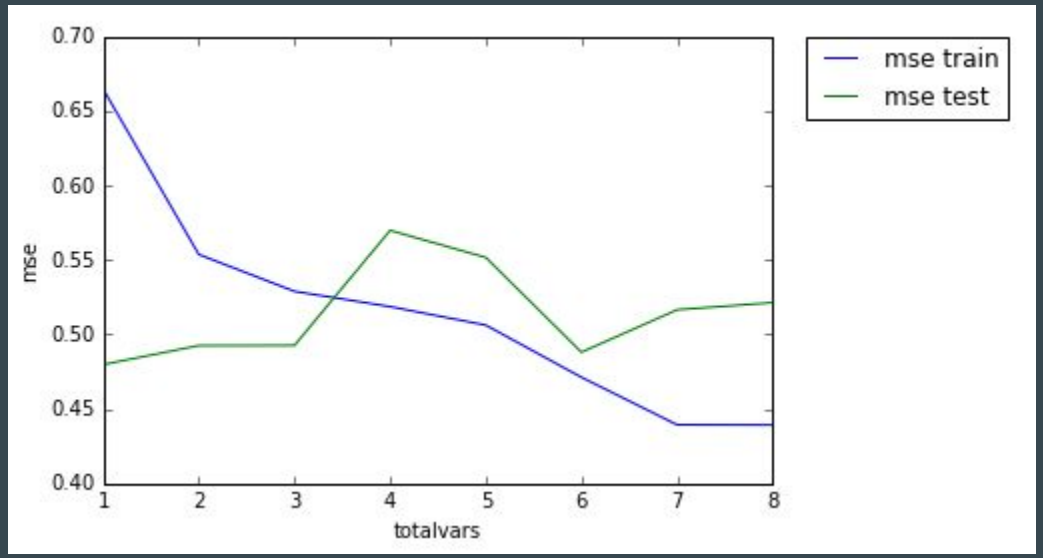
...

Selección de Atributos

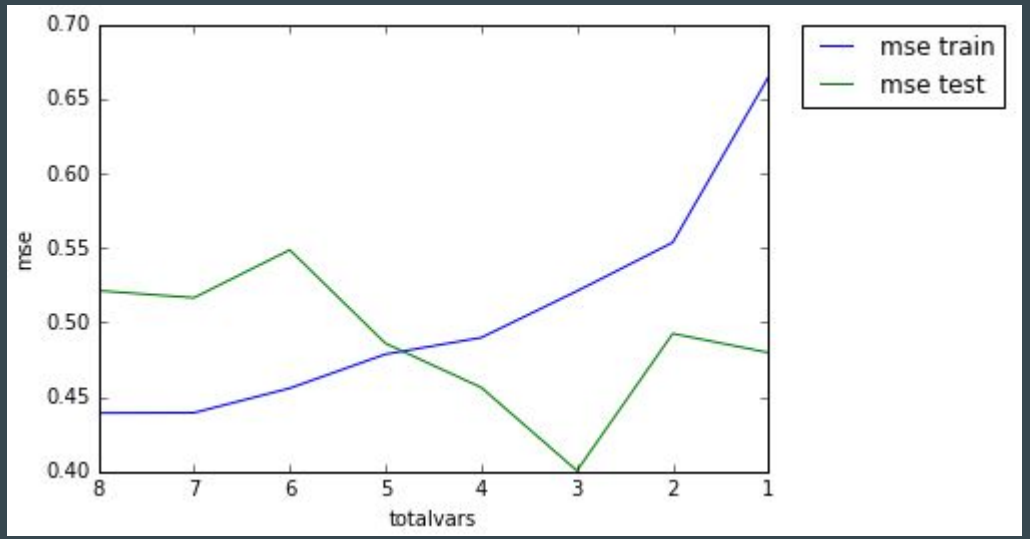
Parte A: Forward Step-wise Selection

Métodos:

- Cross-validation
- Z-score
- MSE



Parte B: Backward Step-wise Selection

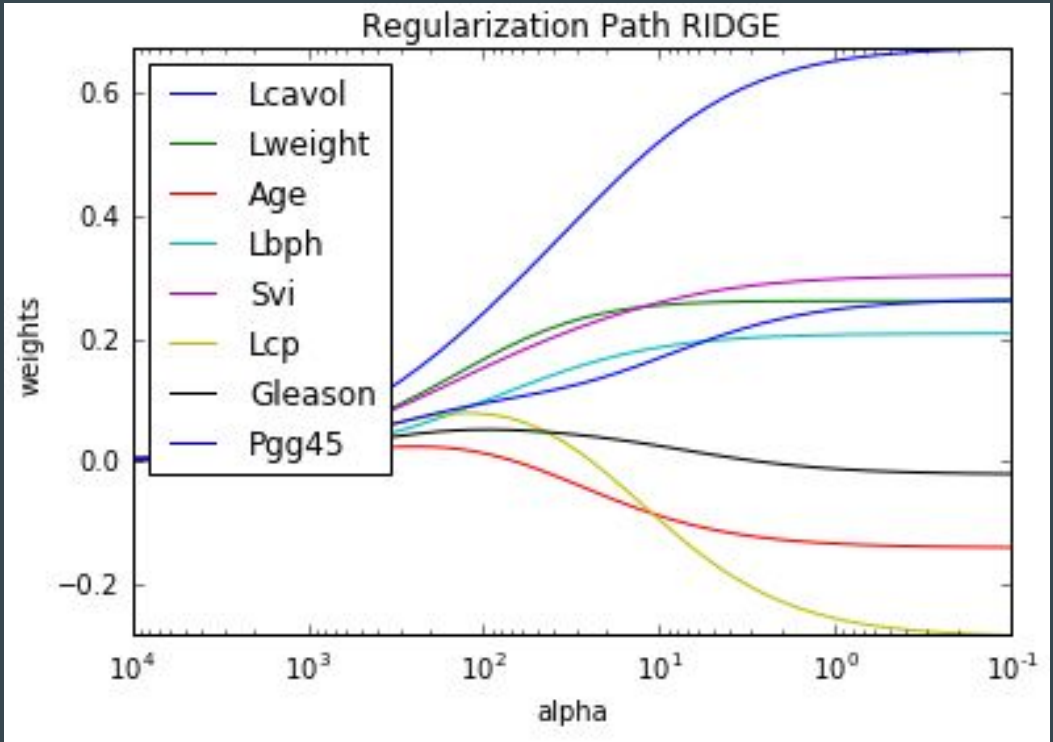


Pregunta 3

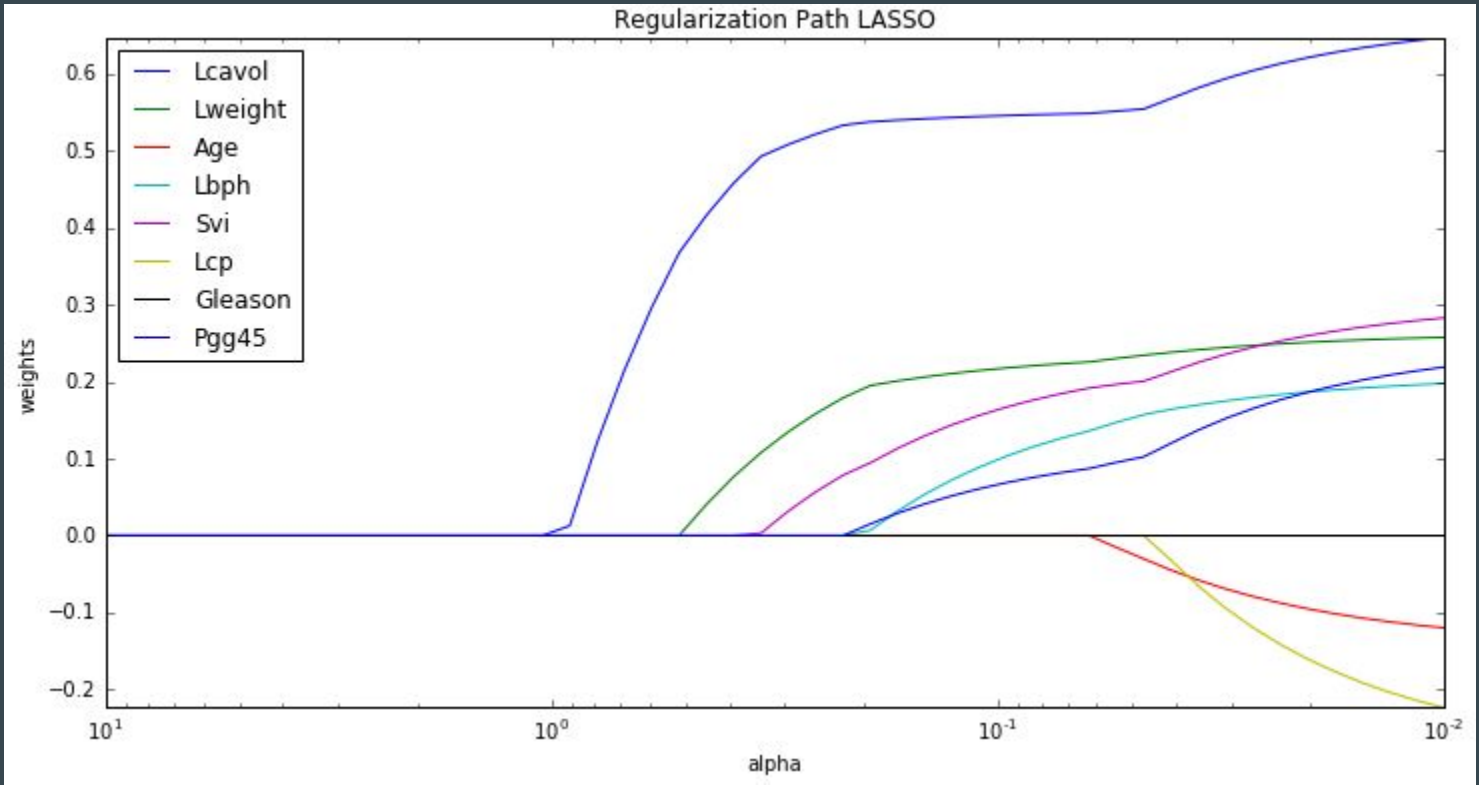
...

Regularización

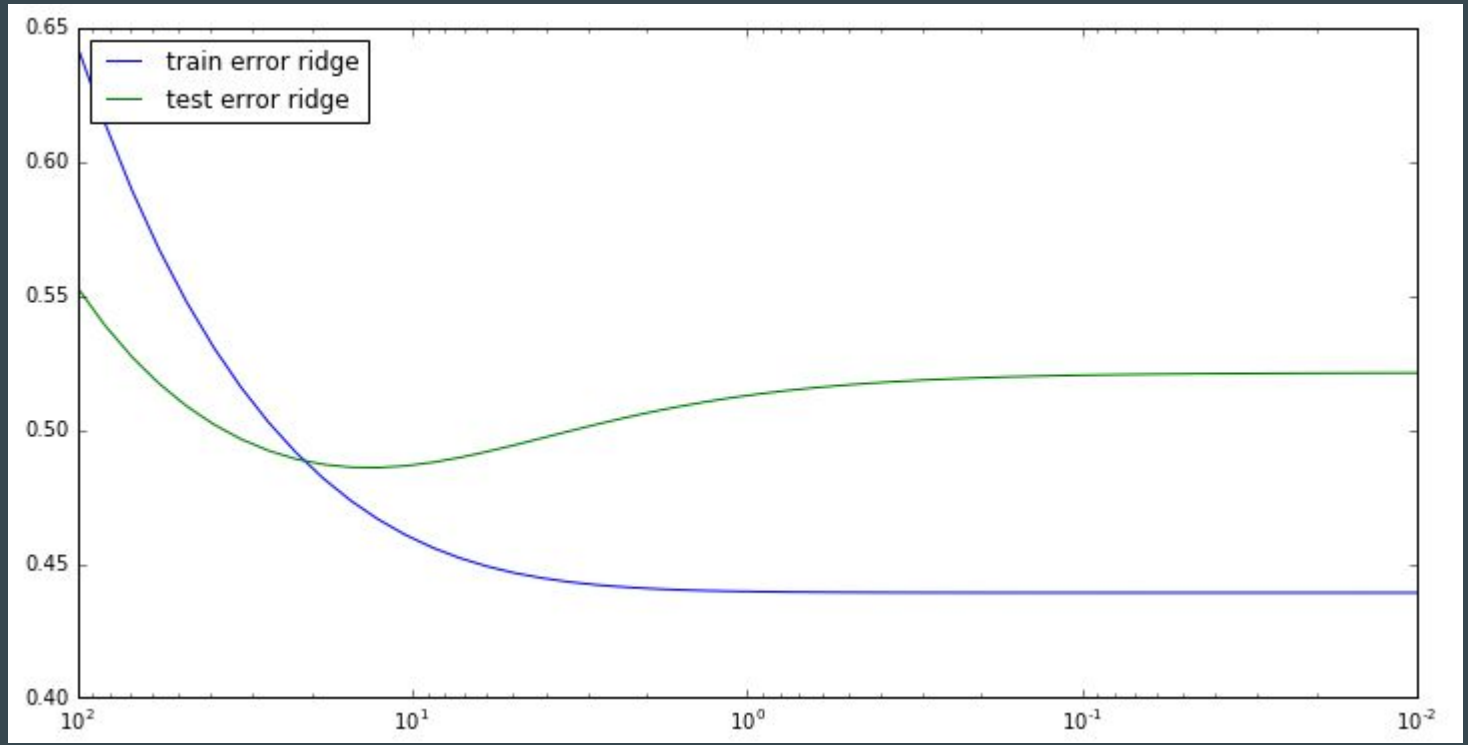
Parte A: Ridge Regression



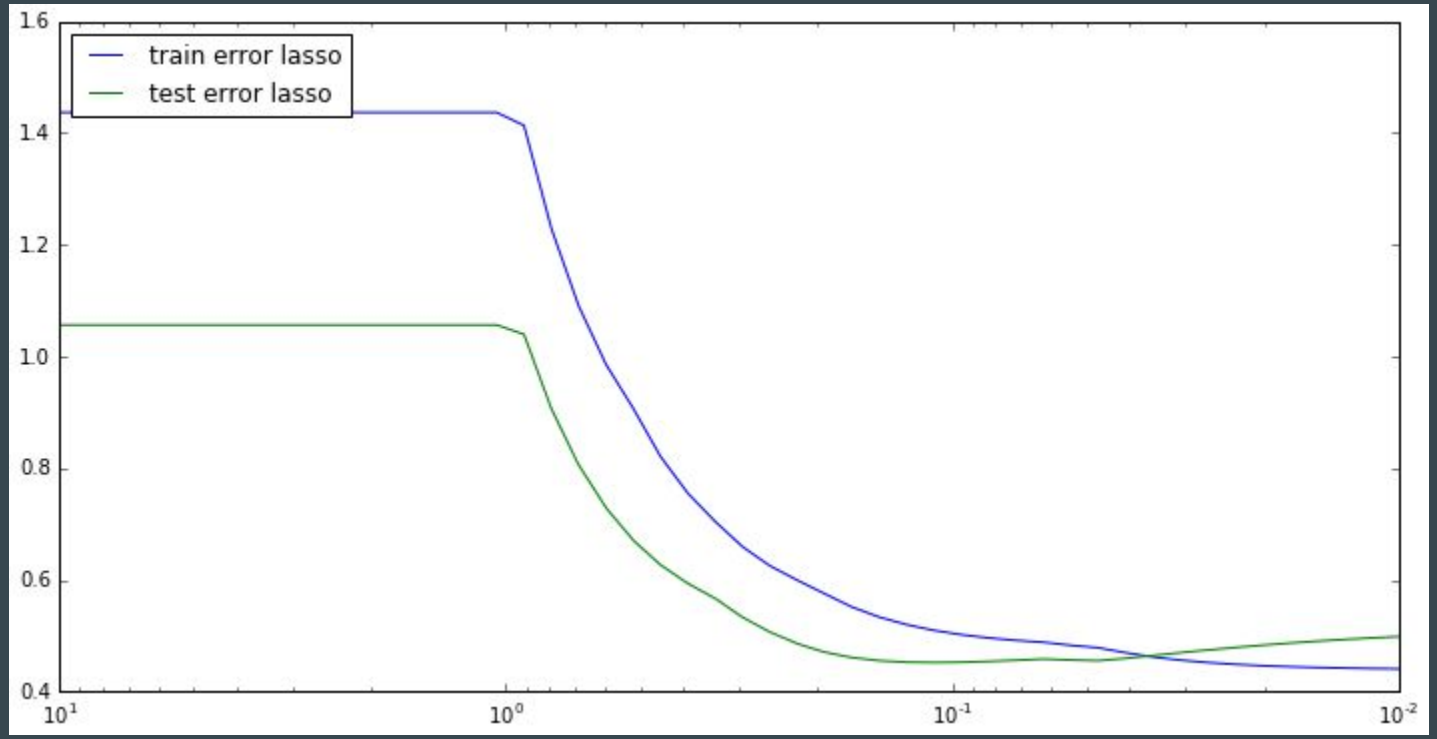
Parte B: Lasso Regression



Parte C: Ridge - Error en Función de Alpha



Parte D: Lasso - Error en Función de Alpha



Tarea 1
Máquinas de Aprendizaje

Matías Ramírez
Javier Rodríguez