

CENTRO DE ENSEÑANZA TÉCNICA Y SUPERIOR



Escuela de Ingeniería en Ciencias Computacionales

Ingeniería en Ciencias Computacionales

Materia: Aprendizaje De Máquina

Presenta:

Pablo Díaz 030343

Kevin Huerta 30502

Gerardo Hernández 29902

Diego Garibay 30046

Profesor: Ulises Orozco Rosas

Practica #6: Selección de modelo lineal
y regularización

Tijuana, B.C., 25 de noviembre de 2020

Aprendizaje de Máquina Práctica 06

Pablo Díaz - 30343 | Kevin Huerta - 30502 | Diego Zuazo - 30046 | Gerardo Hernandez - 29902

Abstract—Comprender y utilizar las funciones básicas en el lenguaje de programación R y Python para realizar distintos tipos de métodos para ajustar modelos lineales en específico son: Subset Selection , Shrinkage y Dimension Reduction

I. INTRODUCCIÓN

Este reporte tiene como objetivo explicar las metodologías utilizadas para realizar los ejercicios de la practica numero 6 , con el uso de dos lenguajes de programación: Python y R. El objetivo de esta practica es hacer uso de los distintos tipos de métodos para ajustar modelos lineales haciendo énfasis en: Subset Selection , Shrinkage y Dimension Reduction. La practica se estructura de la siguiente manera: primero se explica los fundamentos requeridos para la practica , después la metodología , después se muestran los resultados. Finalmente se presentan nuestras conclusiones sobre los temas y la practica.

II. FUNDAMENTOS

Los fundamentos teóricos de la implementación elaborada en esta practica aborda principalmente otros métodos de ajustes lineales aparte de los que ya hemos visto el mas notable siendo el método de least squares. Y por qué podríamos querer utilizar otro procedimiento de ajuste en lugar del least squares? Como se vera, los procedimientos de ajuste alternativos pueden producir una mejor precisión de predicción e interpretabilidad del modelo los métodos en cuestión son los siguientes: Subset Selection , Shrinkage y Dimension Reduction. Por la naturaleza de R, estos son mas sencillos de implementar en dicho lenguaje.

A. Subset Selection

Este método involucra identificar un subconjunto de P predictores que se creen que están relacionados con la respuesta. Para posteriormente ajustar el modelos usando los mínimos cuadrados.

B. Shrinkage

Este método involucra ajustar el modelo involucrando todos los P predictores disponibles. Las estaciones de los coeficientes convergen hacia cero relativamente sobre los mínimos cuadrados.

C. Dimension Reduction

Este método involucra proyectar los P predictores en M dimensiones. Donde $M \leq P$. Esto es logrando calculando M diferentes combinaciones lineales o proyecciones de las variables. Donde dichas M proyecciones son usadas para predecir el ajusta del modelo de regresión lineal por mínimos cuadrados.

III. METODOLOGÍA

La metodología del equipo fue la misma que en veces anteriores, nos dividimos en dos grupos para la implementación del código, un grupo de R y un grupo de Python. Al igual que en practicas pasadas, los equipos se alternaron de lenguaje para practicar. Algunas funciones no existían en python y era muy complicada hacerlas, así que decidimos recurrir a buscar en internet funciones que nos ayudarán a completar el trabajo. Esto lo hicimos debido a que no habían librerías que nos ayudarán y si lo hacíamos todo desde 0, el grado de complejidad iba a aumentar demasiado. Considerando que es un trabajo en equipo, al igual que en las practicas pasadas, tuvimos que trabajar en equipo para encontrar el enfoque adecuada para algunas actividades. De forma mas puntual, la metodología fue la siguiente:

A. Ejercicio número 1

Para el primer ejercicio se aplico el método de "best subset selection" al conjunto de datos Hitters. Se desea predecir el salario de un jugador de béisbol en base a varias estadísticas referente con su desempeño en el año anterior.

B. Ejercicio número 2

Para el ejercicio 2 se uso la funcion de regsubsets() para realizar forward stepwise o backward stepwise selection, usando como parametro method="forward" o method="backward"

C. Ejercicio número 3

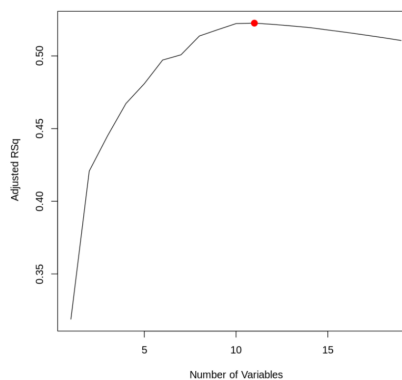
Para el último ejercicio usaremos Principal components regression que en R se traduce como rcp() y también usaremos Partial least squares regression. Para este ejercicio usaremos cross-validation para el remuestreo de las observaciones.

IV. RESULTADOS

Los resultados se encuentran plasmados de la misma forma que en practicas pasadas. Lo que se hace es que. en el caso de R, los resultados se muestran tal y como lo especifica la practica del libro. Ahora bien, en el caso de Python, estos se encuentran en un formato distinto a los presentados en R, pero abordan el mismo problema de la manera mas cercana que permite Python, teniendo en cuenta sus limitaciones al realizar problemas que R realiza con facilidad.

A. Ejercicios en R:

- 1) Actividad 1- Los resultados obtenidos para la Actividad 1 fueron

[illegible]
$$\begin{array}{r} 263 \cdot 20 \\ 0 \end{array}$$

```
#subset selection object  
Call: regsubsets.formula(salary ~ ., Hitters)  
19 Variables (and intercept)
```

	Forced In Forced out							
AtBat	FALSE	FALSE						
Hits	FALSE	FALSE						
HmRun	FALSE	FALSE						
Runs	FALSE	FALSE						
RBI	FALSE	FALSE						
walks	FALSE	FALSE						
Years	FALSE	FALSE						
CatBat	FALSE	FALSE						
Chits	FALSE	FALSE						
ChmRun	FALSE	FALSE						
CRuns	FALSE	FALSE						
CBRI	FALSE	FALSE						
Cwalks	FALSE	FALSE						
Leagueq	FALSE	FALSE						
Divisionw	FALSE	FALSE						
Putsouts	FALSE	FALSE						
Assists	FALSE	FALSE						
Errors	FALSE	FALSE						
NewLeagueq	FALSE	FALSE						

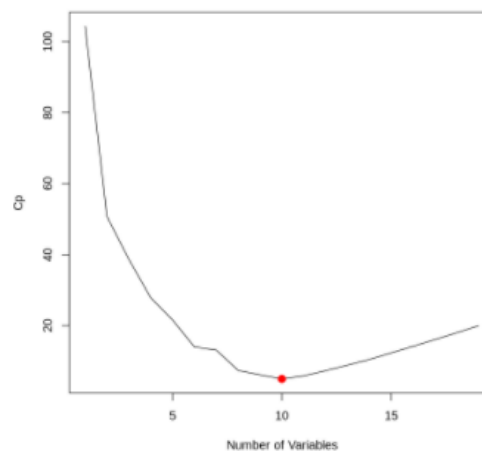
```
      1 subsets of each size up to 8  
Selection Algorithm: exhaustive
```

	AtBat	Hits	HmRun	Runs	RBI	walks	Years	CatBat	Chits	ChmRun	CRuns	CBRI
1 ()	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
2 ()	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
3 ()	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
4 ()	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
5 ()	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
6 ()	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
7 ()	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
8 ()	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "

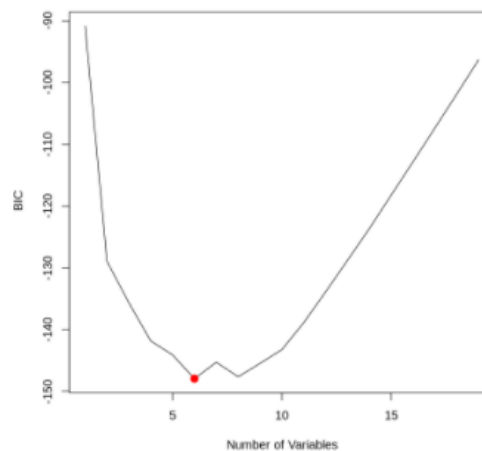
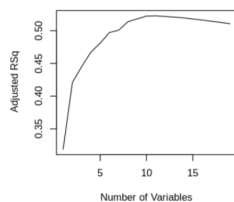
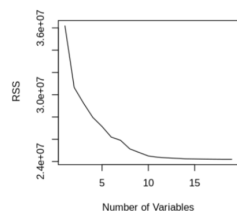
```
    Qualks Leagueq Divisionw Putsouts Assists Errors NewLeagueq
```

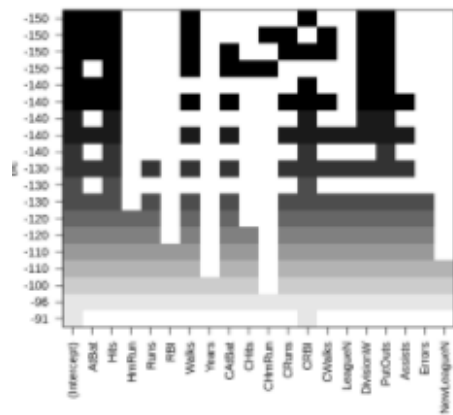
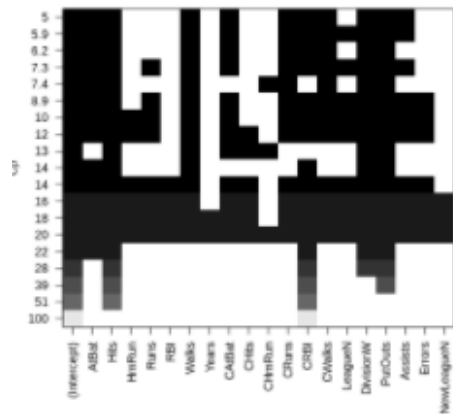
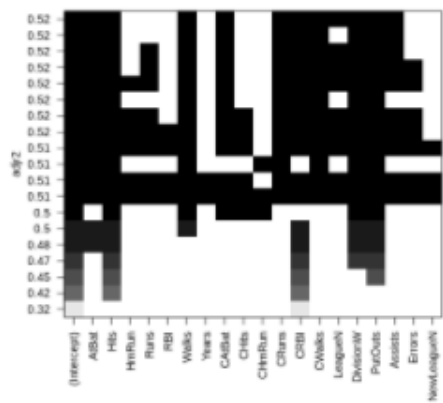
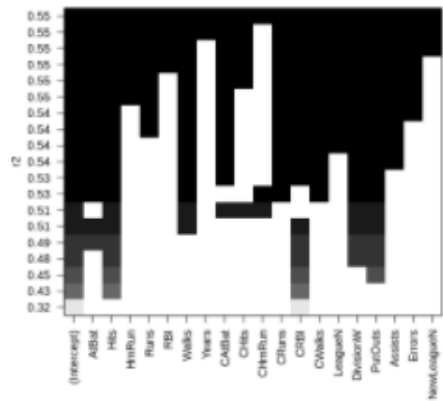
1 ()	" "	" "	" "	" "	" "	" "	" "
2 ()	" "	" "	" "	" "	" "	" "	" "
3 ()	" "	" "	" "	" "	" "	" "	" "
4 ()	" "	" "	" "	" "	" "	" "	" "
5 ()	" "	" "	" "	" "	" "	" "	" "
6 ()	" "	" "	" "	" "	" "	" "	" "
7 ()	" "	" "	" "	" "	" "	" "	" "
8 ()	" "	" "	" "	" "	" "	" "	" "

10
6



```
'which' · 'rsq' · 'rss' · 'adjr2' · 'cp' · 'bic' · 'outmat' · 'obj']
```





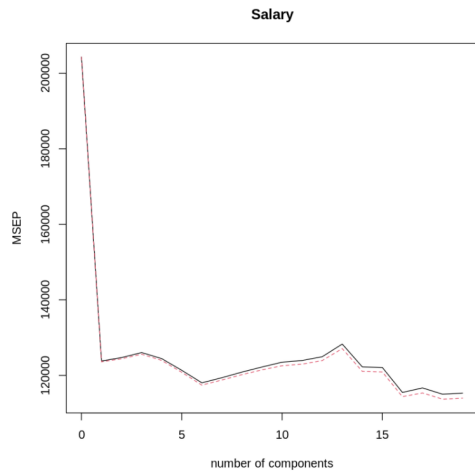
[illegible]

(Intercept)	15.443(11)553091	AzBat:	0.0771554748667925	Hits:	0.859115814031949	HvRun:	0.601031009746882	Runs:	1.06399006654778	RB:	0.879361053395972	Walks:	1.62444816279791	Years:	1.35254779526422	CatBat:
0.011346691152498	Cv:	0.0574666543981478	ChnRun:	0.4066801599484205	CRays:	0.114562244309134	CRB:	0.121165037984304	CWals:	0.0552929199108252	LeagueR:	22.094318947081	DivisionW:	-79.0403063979412	PctOuts:	
0.46640497115267	AsHr:	0.0594544807357670	Drmw:	0.000000000000000	BndL	0.10487399956914										

3) Actividad 3- Los resultados obtenidos para la Actividad 3 fueron los siguientes:

VALIDATION: RMSEP									
Cross-validated using 10 random segments.									
	(Intercept)	1 comps	2 comps	3 comps	4 comps	5 comps	6 comps		
CV	452	351.9	353.2	355.0	352.8	348.4	343.6		
adjCV	452	351.6	352.7	354.4	352.1	347.6	342.7		
	7 comps	8 comps	9 comps	10 comps	11 comps	12 comps	13 comps		
CV	345.5	347.7	349.6	351.4	352.1	353.5	358.2		
adjCV	344.7	346.7	348.5	350.1	350.7	352.0	356.5		
	14 comps	15 comps	16 comps	17 comps	18 comps	19 comps			
CV	349.7	349.4	339.9	341.6	339.2	339.6			
adjCV	348.0	347.7	338.2	339.7	337.2	337.6			

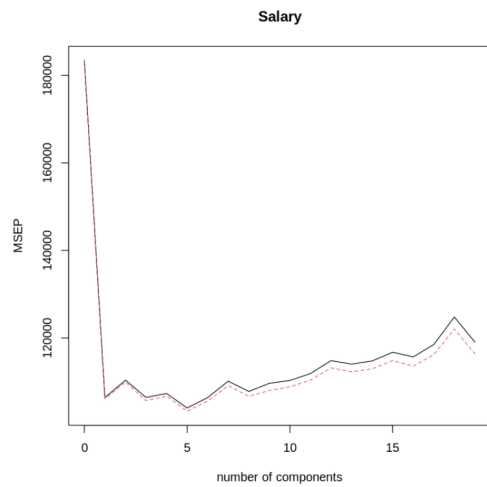
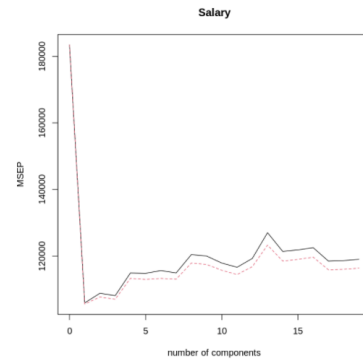
TRAINING: % variance explained									
	1 comps	2 comps	3 comps	4 comps	5 comps	6 comps	7 comps	8 comps	
X	38.31	60.16	70.84	79.03	84.29	88.63	92.26	94.96	
Salary	40.63	41.58	42.17	43.22	44.90	46.48	46.69	46.75	
	9 comps	10 comps	11 comps	12 comps	13 comps	14 comps	15 comps		
X	96.28	97.26	97.98	98.65	99.15	99.47	99.75		
Salary	46.86	47.76	47.82	47.85	48.10	50.40	50.55		
	16 comps	17 comps	18 comps	19 comps					
X	99.89	99.97	99.99	100.00					
Salary	53.01	53.85	54.61	54.61					



Data: X dimension: 131 19
Y dimension: 131 1
Fit method: kernelpls
Number of components considered: 19

VALIDATION: RMSEP
Cross-validated using 10 random segments.
(Intercept) 1 comps 2 comps 3 comps 4 comps 5 comps 6 comps
CV 428.3 325.5 329.9 328.8 339.0 338.9 340.1
adjCV 428.3 325.0 328.2 327.2 336.6 336.1 336.6
7 comps 8 comps 9 comps 10 comps 11 comps 12 comps 13 comps
CV 339.0 347.1 346.4 343.4 341.5 345.4 356.4
adjCV 336.2 343.4 342.8 340.2 338.3 341.8 351.1
14 comps 15 comps 16 comps 17 comps 18 comps 19 comps
CV 348.4 349.1 350.0 344.2 344.5 345.0
adjCV 344.2 345.0 345.9 340.4 340.6 341.1

TRAINING: % variance explained
1 comps 2 comps 3 comps 4 comps 5 comps 6 comps 7 comps 8 comps
X 39.13 48.80 60.09 75.07 78.58 81.12 88.21 90.71
Salary 46.36 50.72 52.23 53.03 54.07 54.77 55.05 55.66
9 comps 10 comps 11 comps 12 comps 13 comps 14 comps 15 comps
X 93.17 96.05 97.08 97.61 97.97 98.70 99.12
Salary 55.95 56.12 56.47 56.68 57.37 57.76 58.08
16 comps 17 comps 18 comps 19 comps
X 99.61 99.70 99.95 100.00
Salary 58.17 58.49 58.56 58.62



145367.722827519

Data: X dimension: 263 19
Y dimension: 263 1
Fit method: kernelpls
Number of components considered: 2
TRAINING: % variance explained
1 comps 2 comps
X 38.08 51.03
Salary 43.05 46.40

140751.276313081

Data: X dimension: 263 19
Y dimension: 263 1
Fit method: svdpc
Number of components considered: 7
TRAINING: % variance explained
1 comps 2 comps 3 comps 4 comps 5 comps 6 comps 7 comps
X 38.31 60.16 70.84 79.03 84.29 88.63 92.26
y 40.63 41.58 42.17 43.22 44.90 46.48 46.69

B. Ejercicios en Python:

4) Actividad 1- Los resultados obtenidos para la Actividad 1 fueron los siguientes:

	AtBat	Hits	HRun	Runs	RBI	Walks	Years	AtBat	Hits	HRun	Runs	RBI	Chalks	League	Division	PuTbts	Assists	Errors	Salary	NoLeague
0	293	66	1	30	29	14	1	293	66	1	30	29	14	A	E	446	33	20	NaN	A
1	315	81	7	24	38	39	14	3449	835	69	321	414	375	N	W	632	43	10	475.0	N
2	479	130	18	66	72	76	3	1624	457	63	224	266	263	A	W	880	82	14	480.0	A
3	496	141	20	65	78	37	11	5628	1575	225	828	838	354	N	E	200	11	3	500.0	N
4	321	87	10	39	42	30	2	396	101	12	48	46	33	N	E	805	40	4	91.5	N

(263, 20)

Best 1 features: ['CRBI']
 Best 2 features: ['Hits', 'CRBI']
 Best 3 features: ['Hits', 'CRBI', 'PutOuts']
 Best 4 features: ['Division[T.W]', 'Hits', 'CRBI', 'PutOuts']
 Best 5 features: ['Division[T.W]', 'AtBat', 'Hits', 'CRBI', 'PutOuts']
 Best 6 features: ['Division[T.W]', 'AtBat', 'Hits', 'Walks', 'CRBI', 'PutOuts']
 Best 7 features: ['Division[T.W]', 'Hits', 'Walks', 'CatBat', 'CHits', 'PutOuts']
 Best 8 features: ['Division[T.W]', 'AtBat', 'Hits', 'Walks', 'CHmRun', 'CRuns', 'CWalks', 'PutOuts']

```
Best 1 Features: ['CRBI']
Best 2 Features: ['CRBI', 'Hits']
Best 3 Features: ['CRBI', 'Hits', 'PutOuts']
Best 4 Features: ['CRBI', 'Hits', 'PutOuts', 'Division']
Best 5 Features: ['CRBI', 'Hits', 'PutOuts', 'Division', 'Walks']
Best 6 Features: ['CRBI', 'Hits', 'PutOuts', 'Division', 'Walks', 'CatBat']
Best 7 Features: ['CRBI', 'Hits', 'PutOuts', 'Division', 'Walks', 'CatBat', 'CHits']
Best 8 Features: ['CRBI', 'Hits', 'PutOuts', 'Division', 'Walks', 'CatBat', 'CHits', 'CHmRun']
Best 9 Features: ['CRBI', 'Hits', 'PutOuts', 'Division', 'Walks', 'CatBat', 'CHits', 'CHmRun', 'CRuns']
Best 10 Features: ['CRBI', 'Hits', 'PutOuts', 'Division', 'Walks', 'CatBat', 'CHits', 'CHmRun', 'CRuns', 'CWalks']
Best 11 Features: ['CRBI', 'Hits', 'PutOuts', 'Division', 'Walks', 'CatBat', 'CHits', 'CHmRun', 'CRuns', 'CWalks', 'PutOuts']
Best 12 Features: ['CRBI', 'Hits', 'PutOuts', 'Division', 'Walks', 'CatBat', 'CHits', 'CHmRun', 'CRuns', 'CWalks', 'PutOuts', 'Errors']
Best 13 Features: ['CRBI', 'Hits', 'PutOuts', 'Division', 'Walks', 'CatBat', 'CHits', 'CHmRun', 'CRuns', 'CWalks', 'PutOuts', 'Errors', 'Assists']
Best 14 Features: ['CRBI', 'Hits', 'PutOuts', 'Division', 'Walks', 'CatBat', 'CHits', 'CHmRun', 'CRuns', 'CWalks', 'PutOuts', 'Errors', 'Assists', 'League']
Best 15 Features: ['CRBI', 'Hits', 'PutOuts', 'Division', 'Walks', 'CatBat', 'CHits', 'CHmRun', 'CRuns', 'CWalks', 'PutOuts', 'Errors', 'Assists', 'League', 'NewLeague']
Best 16 Features: ['CRBI', 'Hits', 'PutOuts', 'Division', 'Walks', 'CatBat', 'CHits', 'CHmRun', 'CRuns', 'CWalks', 'PutOuts', 'Errors', 'Assists', 'League', 'NewLeague', 'Division']
Best 17 Features: ['CRBI', 'Hits', 'PutOuts', 'Division', 'Walks', 'CatBat', 'CHits', 'CHmRun', 'CRuns', 'CWalks', 'PutOuts', 'Errors', 'Assists', 'League', 'NewLeague', 'Division', 'PutOuts']
Best 18 Features: ['CRBI', 'Hits', 'PutOuts', 'Division', 'Walks', 'CatBat', 'CHits', 'CHmRun', 'CRuns', 'CWalks', 'PutOuts', 'Errors', 'Assists', 'League', 'NewLeague', 'Division', 'PutOuts', 'Assists']
Best 19 Features: ['CRBI', 'Hits', 'PutOuts', 'Division', 'Walks', 'CatBat', 'CHits', 'CHmRun', 'CRuns', 'CWalks', 'PutOuts', 'Errors', 'Assists', 'League', 'NewLeague', 'Division', 'PutOuts', 'Assists', 'Errors']
Best 20 Features: ['CRBI', 'Hits', 'PutOuts', 'Division', 'Walks', 'CatBat', 'CHits', 'CHmRun', 'CRuns', 'CWalks', 'PutOuts', 'Errors', 'Assists', 'League', 'NewLeague', 'Division', 'PutOuts', 'Assists', 'Errors', 'NewLeague']
```

```
Best 10 Features: ['AtBat', 'Hits', 'Walks', 'Years', 'CRBI', 'CHits', 'CHmRun', 'CRuns', 'CWalks', 'PutOuts', 'Assists', 'Errors', 'NewLeague']
Best 11 Features: ['AtBat', 'Hits', 'Walks', 'Years', 'CRBI', 'CHits', 'CHmRun', 'CRuns', 'CWalks', 'PutOuts', 'Assists', 'Errors', 'NewLeague', 'Division']
Best 12 Features: ['AtBat', 'Hits', 'Walks', 'Years', 'CRBI', 'CHits', 'CHmRun', 'CRuns', 'CWalks', 'PutOuts', 'Assists', 'Errors', 'NewLeague', 'Division', 'PutOuts']
Best 13 Features: ['AtBat', 'Hits', 'Walks', 'Years', 'CRBI', 'CHits', 'CHmRun', 'CRuns', 'CWalks', 'PutOuts', 'Assists', 'Errors', 'NewLeague', 'Division', 'PutOuts', 'Assists']
Best 14 Features: ['AtBat', 'Hits', 'Walks', 'Years', 'CRBI', 'CHits', 'CHmRun', 'CRuns', 'CWalks', 'PutOuts', 'Assists', 'Errors', 'NewLeague', 'Division', 'PutOuts', 'Assists', 'Errors']
Best 15 Features: ['AtBat', 'Hits', 'Walks', 'Years', 'CRBI', 'CHits', 'CHmRun', 'CRuns', 'CWalks', 'PutOuts', 'Assists', 'Errors', 'NewLeague', 'Division', 'PutOuts', 'Assists', 'Errors', 'NewLeague']
Best 16 Features: ['AtBat', 'Hits', 'Walks', 'Years', 'CRBI', 'CHits', 'CHmRun', 'CRuns', 'CWalks', 'PutOuts', 'Assists', 'Errors', 'NewLeague', 'Division', 'PutOuts', 'Assists', 'Errors', 'NewLeague', 'Division']
Best 17 Features: ['AtBat', 'Hits', 'Walks', 'Years', 'CRBI', 'CHits', 'CHmRun', 'CRuns', 'CWalks', 'PutOuts', 'Assists', 'Errors', 'NewLeague', 'Division', 'PutOuts', 'Assists', 'Errors', 'NewLeague', 'Division', 'PutOuts']
Best 18 Features: ['AtBat', 'Hits', 'Walks', 'Years', 'CRBI', 'CHits', 'CHmRun', 'CRuns', 'CWalks', 'PutOuts', 'Assists', 'Errors', 'NewLeague', 'Division', 'PutOuts', 'Assists', 'Errors', 'NewLeague', 'Division', 'PutOuts', 'Assists']
Best 19 Features: ['AtBat', 'Hits', 'Walks', 'Years', 'CRBI', 'CHits', 'CHmRun', 'CRuns', 'CWalks', 'PutOuts', 'Assists', 'Errors', 'NewLeague', 'Division', 'PutOuts', 'Assists', 'Errors', 'NewLeague', 'Division', 'PutOuts', 'Assists', 'Errors']
Best 20 Features: ['AtBat', 'Hits', 'Walks', 'Years', 'CRBI', 'CHits', 'CHmRun', 'CRuns', 'CWalks', 'PutOuts', 'Assists', 'Errors', 'NewLeague', 'Division', 'PutOuts', 'Assists', 'Errors', 'NewLeague', 'Division', 'PutOuts', 'Assists', 'Errors', 'NewLeague']
```

Forward stepwise select Features on training set, based on RSS.

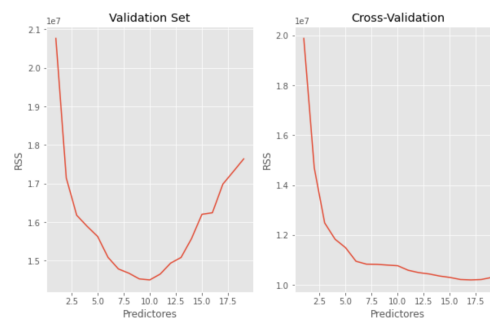
Mean RSS over numbers of Features:

```
# of x_j
1 2.076650e+07
2 1.714303e+07
3 1.617621e+07
4 1.588872e+07
5 1.562742e+07
6 1.508433e+07
7 1.478177e+07
8 1.467291e+07
9 1.454376e+07
10 1.448917e+07
11 1.465036e+07
12 1.493468e+07
13 1.508186e+07
14 1.557007e+07
15 1.620186e+07
16 1.628166e+07
17 1.688057e+07
18 1.739889e+07
19 1.763883e+07
```

El mejor modelo es el que tiene 10 características:
 ['CRBI', 'Hits', 'Division', 'Years', 'AtBat', 'Walks', 'PutOuts', 'CWalks', 'CRuns', 'RBI', 'Errors', 'Assists', 'HmRun', 'Runs', 'NewLeague', 'League', 'CHmRun', 'CatBat', 'CHits']

of x_j

```
1 1.988162e+07
2 1.469190e+07
3 1.247908e+07
4 1.182492e+07
5 1.148818e+07
6 1.094343e+07
7 1.082744e+07
8 1.081981e+07
9 1.079096e+07
10 1.076345e+07
11 1.058229e+07
12 1.048883e+07
13 1.043509e+07
14 1.035223e+07
15 1.029475e+07
16 1.021250e+07
17 1.019581e+07
18 1.021123e+07
19 1.029496e+07
dtype: float64
```



```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 263 entries, 1 to 321
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype
---  -
0   League_A        263 non-null    uint8
1   League_N        263 non-null    uint8
2   Division_E      263 non-null    uint8
3   Division_W      263 non-null    uint8
4   NewLeague_A     263 non-null    uint8
5   NewLeague_N     263 non-null    uint8
dtypes: uint8(6)
memory usage: 3.6 KB
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 263 entries, 1 to 321
Data columns (total 19 columns):
#   Column          Non-Null Count  Dtype
---  -
0   AtBat           263 non-null    int64
1   Hits            263 non-null    int64
2   HmRun           263 non-null    int64
3   Runs            263 non-null    int64
4   RBI             263 non-null    int64
5   Walks           263 non-null    int64
6   Years           263 non-null    int64
7   CatBat          263 non-null    int64
8   CHits           263 non-null    int64
9   CHmRun          263 non-null    int64
10  CRuns           263 non-null    int64
11  CRBI            263 non-null    int64
12  CWalks          263 non-null    int64
13  PutOuts         263 non-null    int64
14  Assists         263 non-null    int64
15  Errors          263 non-null    int64
16  League_N        263 non-null    uint8
17  Division_W      263 non-null    uint8
18  NewLeague_N     263 non-null    uint8
dtypes: int64(16), uint8(3)
memory usage: 35.7 KB
```

- 5) Actividad 2- Los resultados obtenidos para la Actividad 2 fueron los siguientes:
- 6) Actividad 1- Los resultados obtenidos para la Actividad 2 fueron los siguientes:

Ridge regression coefficients:

AtBat	0.098658
Hits	0.446094
HmRun	1.412107
Runs	0.660773
RBI	0.843403
Walks	1.008473
Years	2.779882
CAtBat	0.008244
CHits	0.034149
CHmRun	0.268634
CRuns	0.070407
CRBI	0.070060
CWalks	0.082795
League	0.104747
Division	-0.003739
PutOuts	0.268363
Assists	4.241051
Errors	-30.768885
NewLeague	4.123474

dtype: float64

MSE = 106216.52238005561

AtBat	0.055838
Hits	0.934879
HmRun	0.369048
Runs	1.092480
RBI	0.878259
Walks	1.717770
Years	0.783515
CAtBat	0.011318
CHits	0.061101
CHmRun	0.428333
CRuns	0.121418
CRBI	0.129351
CWalks	0.041990
League	0.179957
Division	0.035737
PutOuts	-1.597699
Assists	24.774519
Errors	-85.948661
NewLeague	8.336918

dtype: float64

Ridge regression coefficients:

AtBat	1.317464e-10
Hits	4.647486e-10
HmRun	2.079865e-09
Runs	7.726175e-10
RBI	9.390640e-10
Walks	9.769219e-10
Years	3.961442e-09
CAtBat	1.060533e-11
CHits	3.993605e-11
CHmRun	2.959428e-10
CRuns	8.245247e-11
CRBI	7.795451e-11
CWalks	9.894387e-11
League	7.268991e-11
Division	-2.615885e-12
PutOuts	2.084514e-10
Assists	-2.501281e-09
Errors	-1.549951e-08
NewLeague	-2.023196e-09

dtype: float64

MSE = 172862.23580379886

The best ridge regularization Alpha = 0.5748784976988678
MSE = 99825.64896292731

The best Lasso regularization Alpha = 0.5748784976988678
MSE = 104960.65853895503

```

AtBat      0.000000
Hits       1.082446
HmRun      0.000000
Runs       0.000000
RBI        0.000000
Walks      2.906388
Years      0.000000
CAtBat     0.000000
CHits      0.000000
CHmRun     0.219367
CRuns      0.000000
CRBI       0.513975
CWalks     0.000000
League     0.368401
Division   -0.000000
PutOuts    -0.000000
Assists    0.000000
Errors     -89.064338
NewLeague  0.000000
dtype: float64

```

```

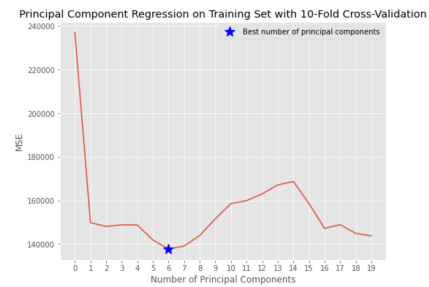
Explained Variance Ratio:
0      38.31 %
1      60.150000000000006 %
2      70.84 %
3      79.03 %
4      84.29 %
5      88.630000000000001 %
6      92.26 %
7      94.960000000000001 %
8      96.28 %
9      97.25 %
10     97.97 %
11     98.64 %
12     99.14 %
13     99.46 %
14     99.72999999999999 %
15     99.88 %
16     99.94999999999999 %
17     99.97999999999999 %
18     99.99 %
dtype: object

```

Shape (19, 19)

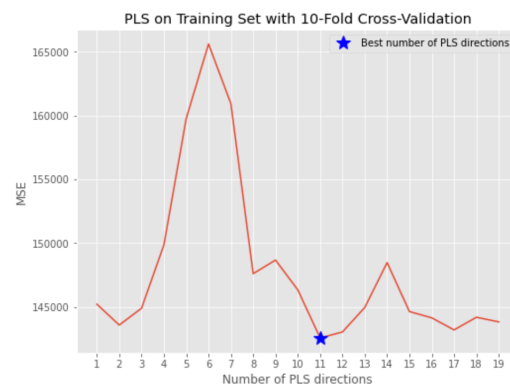
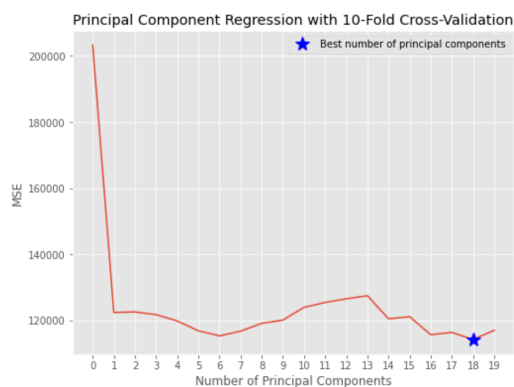
Head:

	0	1	2	3	4	5
0	0.198290	-0.383784	0.088626	0.031967	0.028117	-0.070646
1	0.195861	-0.377271	0.074032	0.017982	-0.004652	-0.082240
2	0.204369	-0.237136	-0.216186	-0.235831	0.077660	-0.149646
3	0.198337	-0.377721	-0.017166	-0.049942	-0.038536	-0.136660
4	0.235174	-0.314531	-0.073085	-0.138985	0.024299	-0.111675



7) Actividad 3- Los resultados obtenidos para la Actividad 3 fueron los siguientes:

Test set MSE = 111994.42273636989



V. CONCLUSIONES

En conclusión, como se a demostrado en las practicas anteriores los temas visto en esta practican resultan ser de gran utilidad. Nos es fascinante la facilidad con la que podemos implementar estos métodos en R , por ejemplo el método de forward o backward stepwise puede ser implementado usando una misma funcion y simplemente cambiando un parámetro. Al realizar las actividades nos ayuda a comprender los temas debido a que los ejercicios de la practica consisten en implementar los métodos. Al entender mas estos temas nos expande otra área de conocimiento respecto a la materia , y amplia nuestras habilidades en ajustar modelos. Finalmente al realizar la practica no solo concebimos estos temas ,sino que también estamos reforzando nuestras habilidades de programar y desarrollar problemas de aprendizaje maquina en R y Python.