Práctica. Regresión Polinomial

DRA. CONSUELO VARINIA GARCÍA MENDOZA

- I. Para este programa se utilizará el dataset datos.csv
- II. Divide datos.csv en 70% para entrenamiento y 30% para pruebas, con los parámetros shuffle=True y random_state =0
- III. Con las bibliotecas de scikit-learn realiza las siguientes regresiones:
 - Regresión lineal con OLS
 - Regresión polinomial de grado 2 con OLS
 - Regresión polinomial de grado 3 con OLS
 - Regresión lineal con SGD
 - Regresión polinomial de grado 2 con SGD
 - Regresión polinomial de grado 3 con SGD
- IV. Entrada
 - Archivo datos.csv, para SGD: número de iteraciones y α

V. Salida

- Una gráfica para cada regresión con los siguientes elementos:
 - √ distribución de los datos de prueba
 - ✓ (X_test, y_pred) // en el caso de regresión polinomial ordenar los datos antes de pintar la curva predicha
- Resumen de los resultados

VI. Bibliotecas de scikit-learn

- from sklearn.linear model import LinearRegression
- from sklearn.linear_model import SGDRegressor
- from sklearn.preprocessing import PolynomialFeatures
- from sklearn.metrics import mean squared error, r2 score
- import operator

Dataset

X	У
-3.292157	-44.608971
-0.936214	0.6231784
-3.602674	-58.873048
-0.8502653	-6.0575889
2.3096566	-3.9671278
1.5678693	-0.0586592
-0.2831132	-3.0897405
0.6684103	-2.0747978
-2.4822372	-26.06787
1.0607969	-6.1552183
9.6589694	272.59856
-0.5933086	-6.6282606
-4.8092639	-114.04718
1 8627244	-5 4067249

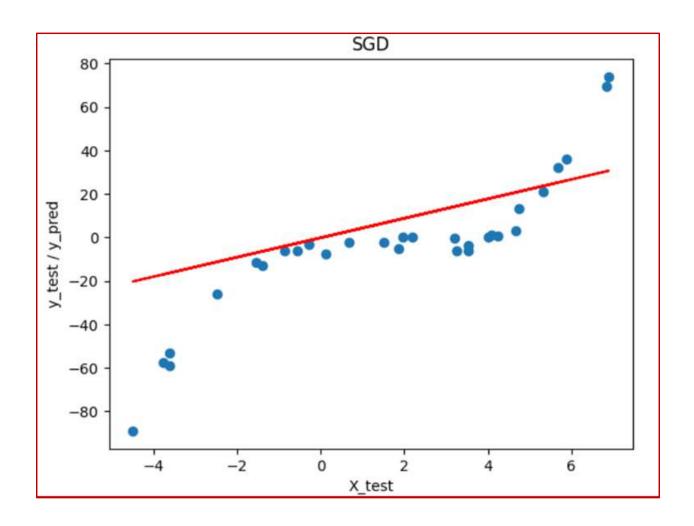


Figura 1. Ejemplo de la gráfica de la regresión con SGD

even the transfer of the second	mse r2
Regresión lineal con OLS	4.163454e+09 6.807057e-01
Regresión polinomial de grado 2 con OLS	3.650732e+10 -1.799738e+00
Regresión polinomial de grado 3 con OLS	2.570015e+20 -1.970939e+10
Regresión lineal con SGD	8.684145e+12 -6.649851e+02
Regresión polinomial de grado 2 con SGD	2.570015e+20 -1.970939e+10
Regresión polinomial de grado 3 con SGD	3.650732e+10 -1.799738e+00

Figura 2. Ejemplo de resumen de resultados

- I. Para este programa se utilizará el dataset *cal_housing.csv.* Las primeras columnas son las características y la última columna el target
- II. Divide cal_housing.csv en 80% para entrenamiento y 20% para pruebas, con los parámetros shuffle=True y random_state =0
- III. Con las bibliotecas de scikit-learn realiza las siguientes regresiones con OLS:
 - Lineal
 - Polinomial de grado 2
 - Polinomial de grado 2 con escalamiento estándar
 - Polinomial de grado 2 con escalamiento robusto
 - Polinomial de grado 3
 - Polinomial de grado 3 con escalamiento estándar
 - Polinomial de grado 3 con escalamiento robusto
- IV. Entrada
 - Archivo cal_housing

V. Salida

Resumen de los resultados (ver el ejemplo de la Fig. 3)

VI. Bibliotecas de scikit-learn

from sklearn import preprocessing

Escalamiento

Estandarizado

$$\frac{(X - media)}{std}$$

Robusto

$$\frac{(X - rango_intercuat\'ilico)}{std}$$

donde:

- std: desviación estándar
- rango_intercuatílico: diferencia entre el percentil 75 y el percentil 25 de una muestra

Port	tapapeles 🗔	F	uente	IZI	Alineación	[2]	Número ဩ	E	stilos
A1	•	: ×	√ fx lo	ongitude					
1	А	В	C	D	E	F	G	Н	Ť.
1	longitude	latitude	housingMedia	totalRooms	totalBedroon	population	households	medianIncom	medianHouseValu
2	-122.23	37.88	41	880	129	322	126	8.3252	452600
3	-122.22	37.86	21	7099	1106	2401	1138	8.3014	358500
4	-122.24	37.85	52	1467	190	496	177	7.2574	352100
5	-122.25	37.85	52	1274	235	558	219	5.6431	341300
6	-122.25	37.85	52	1627	280	565	259	3.8462	342200
7	-122.25	37.85	52	919	213	413	193	4.0368	269700
8	-122.25	37.84	52	2535	489	1094	514	3.6591	299200
9	-122.25	37.84	52	3104	687	1157	647	3.12	241400
10	-122.26	37.84	42	2555	665	1206	595	2.0804	226700
11	-122.25	37.84	52	3549	707	1551	714	3.6912	261100
12	-122.26	37.85	52	2202	434	910	402	3.2031	281500
13	-122.26	37.85	52	3503	752	1504	734	3.2705	241800
14	-122.26	37.85	52	2491	474	1098	468	3.075	213500
15	-122.26	37.84	52	696	191	345	174	2.6736	191300
16	-122.26	37.85	52	2643	626	1212	620	1.9167	159200
17	-122.26	37.85	50	1120	283	697	264	2.125	140000

```
Regresión con OLS
                                                                            r2
                                                              mse
Lineal
                                                      4.163454e+09 6.807057e-01
Polinomial de grado 2
                                                      3.650732e+10 -1.799738e+00
Polinomial de grado 2 con escalamiento estandar
                                                      2.570015e+20 -1.970939e+10
Polinomial de grado 2 con escalamiento robusto
                                                     8.862040e+09 3.203722e-01
Polinomial de grado 2
                                                     8.684145e+12 -6.649851e+02
Polinomial de grado 2 con escalamiento estandar
                                                     2.570015e+20 -1.970939e+10
Polinomial de grado 2 con escalamiento robusto
                                                      3.650732e+10 -1.799738e+00
```

Figura 3. Ejemplo de resumen de resultados