

Práctica. Regresión Polinomial

DRA. CONSUELO VARINIA GARCÍA MENDOZA



Especificaciones del programa 1

- 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

Especificaciones del programa 1

V. Salida

- Una gráfica para cada regresión con los siguientes elementos:
 - ✓ distribución de los datos de prueba
 - ✓ $(X_{\text{test}}, y_{\text{pred}})$ // 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	y
-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

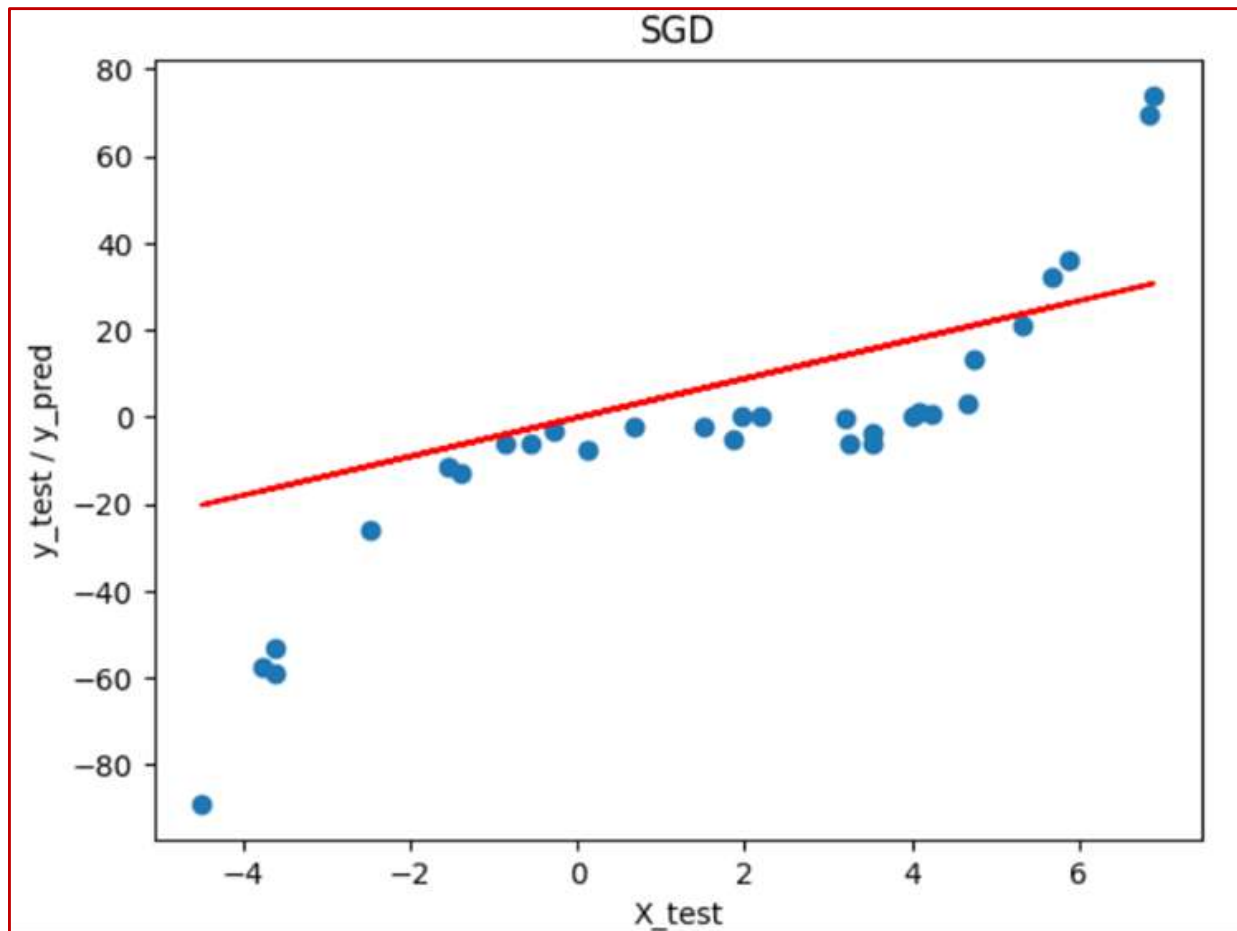


Figura 1. Ejemplo de la gráfica de la regresión con SGD

	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

Especificaciones del programa 2

- 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*

Especificaciones del programa 2

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

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

donde:

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

Regresión con OLS

	mse	r2
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