

Wilson Rodas

Modelo lógico

1) Formulación del problema

En este proyecto se aplicará el modelo de regresión logística para realizar la predicción de casos nuevos confirmados de COVID-19 en el Ecuador. Para este propósito se utilizarán como variables primordiales el número de casos confirmados y la fecha en formato de días. Además se realizará una comparación con el modelo lineal para ver cual se ajusta mejor a los datos analizados.

2) Definición del sistema

Para realizar la simulación se utilizará un cuaderno de Python (Jupyter Notebook), en el que el usuario podrá ir ejecutando segmentos de código y ver como se desarrolla el entrenamiento y predicción del modelo.

3) Formulación del modelo

Como se mencionó previamente, el modelo a utilizar en este proyecto será el de regresión logística.

4) Colección de datos

Los datos a utilizar son los de los informes nacionales del Servicio Nacional de Gestión de Riesgos y Emergencias del Ecuador (SNGRE), Ministerio de Salud Pública del Ecuador (MSP), y Registro Civil del Ecuador.

5) Implementación

```
In [2]: #Importación de las librerías necesarias  
import pandas as pd  
import numpy as np  
from datetime import datetime  
from sklearn import linear_model  
from scipy.optimize import curve_fit
```

```
import matplotlib.pyplot as plt
%matplotlib inline
```

```
In [3]: #Obtención de los datos de COVID-19 de Ecuador
url_datos = 'DatosCOVID2020.csv'

datos = pd.read_csv(url_datos, sep = ',')
datos
```

```
Out[3]:
```

	muestras	muestras_pcr	muestras_pcr_nuevas	pruebas_rezagadas	muertes_confirmadas	muertes_probables	muertes	muertes_nuevas	positivas
0	129	129	0	106	1	0	1	0	23
1	206	206	77	178	2	0	2	1	28
2	273	273	67	236	2	0	2	0	37
3	354	354	81	296	2	0	2	0	58
4	762	762	408	651	2	0	2	0	111
...
242	597099	576527	559	49598	8525	4324	12849	10	184667
243	600741	580169	3642	49067	8592	4328	12920	71	185586
244	605331	584759	4590	50256	8614	4332	12946	26	186469
245	610265	589693	4934	50604	8642	4335	12977	31	187630
246	614531	593959	4266	50485	8658	4339	12997	20	188583

247 rows × 32 columns

```
In [4]: #Se filtran los datos para obtener solo los casos positivos
datos_infectados = datos.loc[:, ['positivas', 'created_at']]

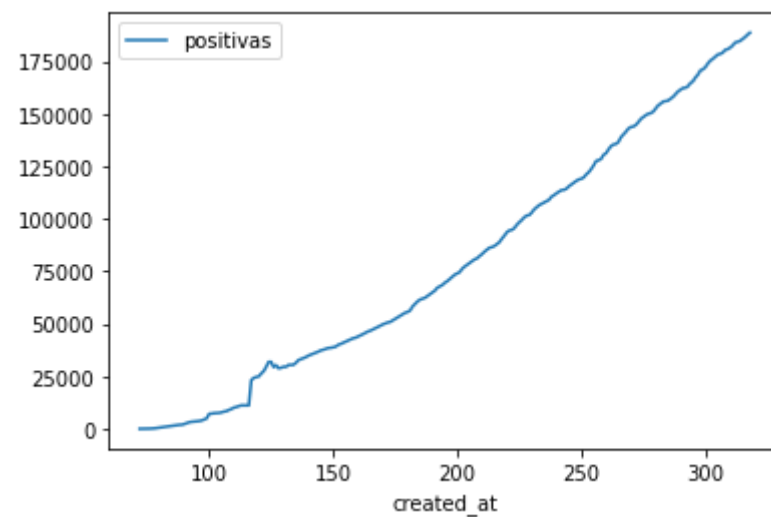
#Expresar los datos de fecha en número de días desde el inicio de año
formato = '%d/%m/%Y'
fecha = datos_infectados['created_at']
datos_infectados['created_at'] = fecha.map(lambda x:
    (datetime.strptime(x, formato) - datetime.strptime('01/01/2020', formato)).days
)

datos_infectados
```

Out[4]:

	positivas	created_at
0	23	72
1	28	73
2	37	74
3	58	75
4	111	76
...
242	184667	314
243	185586	315
244	186469	316
245	187630	317
246	188583	318

247 rows × 2 columns

In [5]: `datos_infectados.plot(x = 'created_at', y = 'positivas')`Out[5]: `<AxesSubplot:xlabel='created_at'>`

Como se puede observar, el número de casos inicia el día 72 (13/03/2020) con 23 casos confirmados. Hasta el día 318 (14/11/2020) se registran 188583 casos confirmados.

```
In [6]: #Definición del modelo logístico
modelo_logistico = lambda x, a, b: a + b * np.log(x)

#Obtención de las variables para el entrenamiento
x = list(datos_infectados.iloc[:, 1]) #Fecha (Número de día)
y = list(datos_infectados.iloc[:, 0]) #Numero de infectados

#Realizamos el ajuste de curva para los datos
ajuste = curve_fit(modelo_logistico, x, y)
```

6) Verificación

La verificación de la implementación del modelo puede ser realizada mediante la ejecución paso a paso del propio cuaderno de python. En este caso no se han detectado errores, por lo que se prosigue al diseño de experimentos

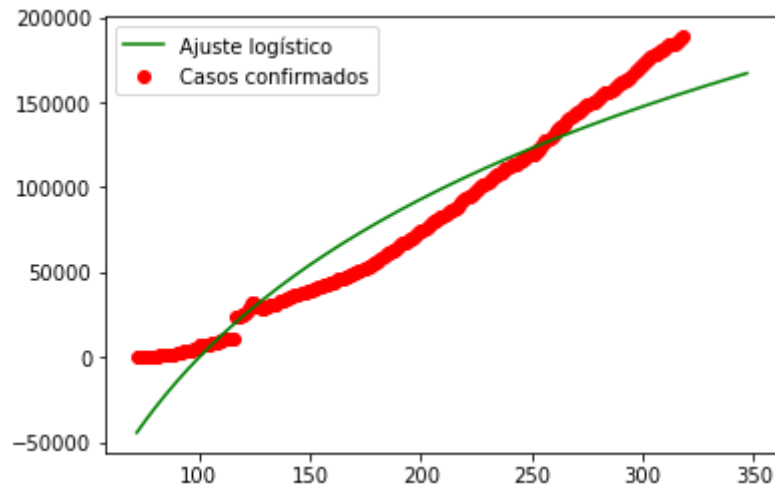
7) Diseño de experimentos

Para probar el modelo se utilizara como premisa un periodo de predicción de 30 días posteriores a la fecha 14/11/2020. Al ser un modelo sencillo unicamente se realizara una simulación.

8) Experimentación

```
In [7]: #Comprobación de la predicción para 30 días despues del 14/11/2020
dias = 30
x_real = list(range(min(x), max(x) + dias)) #Realizamos la predicción para un mes

plt.scatter(x, y, color = 'red')
plt.plot(x_real, [modelo_logistico(i, ajuste[0][0], ajuste[0][1]) for i in x_real], color = 'green')
plt.legend(('Ajuste logístico', 'Casos confirmados'))
plt.show()
```



9) Interpretación

De acuerdo a los parámetros de entrada de casos confirmados y fecha, se puede observar que el modelo no se ajusta correctamente los datos, por lo que sería conveniente buscar y utilizar otro tipo de función para predecir.

10) Implementación

Este proyecto no cuenta con interfaz gráfica de usuario externa, sin embargo, el usuario que desee ver el comportamiento del sistema cambiando el número de días a predecir lo puede hacer modificando la variable 'días' ubicada en ln[25] de este cuaderno (Para el modelo logístico).

11) Uso del modelo lineal para el mismo conjunto de datos

```
In [8]: #Creación y entranamiento del modelo
modelo_lineal = linear_model.LinearRegression()
modelo_lineal.fit(np.array(x).reshape(-1, 1), y)

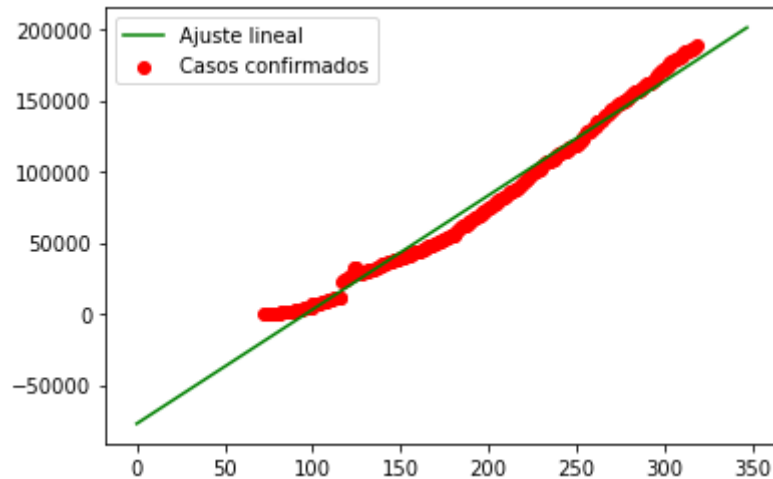
#Comprobación de la predicción para 30 días despues del 14/11/2020
dia = x[-1] + 30 #Se obtiene el dia a predecir sumando la ultima fecha registrada mas un mes
prediccion = modelo_lineal.predict([[dia]])

#Gráfica de los datos y la predicción
x_real = np.array(range(0, 348))

print('Predicción de casos positivos para el día 347: ', int(prediccion))
```

```
plt.scatter(x, y , color = 'red')  
plt.plot(x_real, modelo_lineal.predict(x_real.reshape(-1, 1)), color = 'green')  
plt.legend(('Ajuste lineal', 'Casos confirmados'))  
plt.show()
```

Predicción de casos positivos para el día 347: 201854



En este caso se muestra la reacción del modelo lineal ante el conjunto de datos.

12) Análisis

Los datos fueron obtenidos del siguiente enlace: <https://github.com/andrab/ecuacovid>, el cual es un proyecto que reúne los datos de los informes nacionales. Con respecto al modelo lineal entrenado en este trabajo, se ha utilizado un total de 246 registros, de los cuales los datos principales ha tratar fueron: la fecha ('created_at') y los casos confirmados ('positivas').

El modelo logístico utilizado para la muestra de datos presenta una curva que no se ajusta adecuadamente a esto, por lo que la predicción de casos en el futuro se ve equivocada por un margen de error considerable. Por el contrario al usar el modelo lineal se puede observar que si bien podría mejorar, el ajuste es mejor que el del modelo logístico (Para esta muestra de datos).

13) Conclusiones

Mediante el presente trabajo se establece que de acuerdo a la comparativa realizada entre los modelos logístico y lineal, el lineal es claramente mejor cuando el conjunto de muestras tiene una tendencia lineal y constante, mientras que el modelo logístico se recomienda para datos que presenten un gran crecimiento inicial y luego se vaya estabilizando con crecimiento moderado.

14) Criterio personal (Político, económico y social de la situación)

La emergencia sanitaria por COVID-19 en el Ecuador y en el resto del mundo es una situación realmente dura, siendo así que la gran mayoría de sistemas de salud se han visto saturados con el incremento y llegada de nuevos pacientes. La prioridad de las naciones debería ser apoyar a la investigación de la cura contra el virus, o caso contrario, apoyar a la compra de insumos médicos para respaldar a las casas de salud.