Universidad de Ingeniería y Tecnología

Inteligencia artificial
CS2601

Modelo de regresion no lineal para datos COVID-19

Mauricio Nieto
Jeamlee Barreto
Jonathan Hoyos

Profesor: Christian Lopez

16 de mayo del 2021

Contents

1 Introducción

Vamos a utilizar el Dataset obtenido por covid-19-peru-data-augmented. Y en base a la longitud y latitud de cada departamento, su zona geografica y fecha en formato ISO, se va a contabilizar la cantidad de casos confirmados y muertes. Aplicaremos un modelo lineal con su función de perdida y regularización. Finalmente experimentaremos con diferentes optimizadores de gradiente para ver cual se ajusta mejor a los datos.

2 Implementación

2.0.1 Error medio absoluto (MAE)

Es la promedio de todas las diferencias entre los puntos de los datos y la linea de regresión

$$\frac{1}{n}\sum_{i=1}^{n}|\hat{Y_i} - Y_i|$$

Siendo $\hat{Y_i}$ el dato predecido por la regresión e Y_i el dato real.

Tenemos tres parámetros: el data set x, el data set y, y w, que es el arreglo de parámetros. Siendo $error\_mae$ el valor que resulta de la sumatoria anterior y $loss\_function\_mae$ el valor que resulta de la derivada de la sumatoria.

2.0.2 Error cuadrático medio (MSE)

El error cuadrático medio es el promedio de las diferencias al cuadrado entre la linea de regresión y los datos. MSE penaliza mas las distancias mas grandes.

Tenemos tres parámetros: el data set x, el data set y, y w, que es el arreglo de parámetros, que puede ser vector o matriz. Siendo $error\_mse$ el valor que resulta de la sumatoria anterior y $loss\_function\_mse$ el valor que resulta de la derivada de la sumatoria.

2.1 Técnicas de regularización

Tiene sentido tratar de encontrar siempre los coeficientes que minimicen la función de perdida, ya que así nuestra regresión seria mas precisa. La regularización agrega un termino que penalizara la complejidad del modelo, ya que si el el modelo es muy complejo, no siempre todas sus características afectarán a la predicción. Gracias a este termino, minimizamos la complejidad del modelo y el error a la vez. Esto resulta en modelos mas simples que tienden a generalizar mejor, ya que los modelos muy complejos tienden a hacer overfitting.

$$min(funcion\_de\_perida + \lambda L_p)$$

2.1.1 Regularización L1: Lasso

Lasso busca el w, vector o matriz, que minimice la funcion de perdida y el termino regularizador $\lambda||w||_1$

$$L_1 = \underset{w}{\operatorname{argmin}}(funcion\_de\_perida + \lambda||w||_1)$$

Esta regularización es de ayuda cuando se sospecha que varias de las características, o atributos de entrada son irrelevantes. Al usar Lasso estamos haciendo que la solución sea mas simple mediante la selección de atributos de entrada haciendo que algunos de estos valgan 0.

2.1.2 Regularización L2: Ridge

Ridge busca el w, vector o matriz, que minimice la funcion de perdida y el termino regularizador $\lambda||w||_2$

$$L_2 = \underset{w}{\operatorname{argmin}}\, funcion\_de\_perida + \lambda||w||_2$$

Esta regularización es de ayuda cuando se sospecha que varias de las características, o atributos de entrada están correlacionadas. Al usar Ridge estamos haciendo que la solución sea mas simple mediante la minimizacion de los coeficientes. Esto minimiza el efecto de la correlación y así el modelo generaliza mejor.

3 Resultados

3.1 DataSet

Los datos se obtuvieron del repositorio proporcionado. El total de filas fue de 10957. Para este trabajo se tuvo que limpiar la data, eliminando datos nulos, cambiando las zonas de referencia (por números enteros) y la fecha por un valor incremental con respecto al primer día registrado (2020-03-19).

Luego de limpiar los datos el dataset quedó con unos 9787 registros, de los cuales el 70% se emplearon para el entrenamiento del modelo y el 30% para pruebas. Cabe destacar que solo se trabajó sobre los atributos fecha, zona, latitud y longitud como variables dependientes. Las muertes y los datos confirmados serán las independientes. De aquí en adelante solo se trabajará sobre los datos de confirmados.

3.2 Funciones de perdida

3.2.1 Error medio absoluto (MAE) y error cuadrático medio (MSE)

Variando la tasa de aprendizaje podemos ajustar un modelo polinomial de grado 2 para predecir el comportamiento de los datos. A continuación se muestran

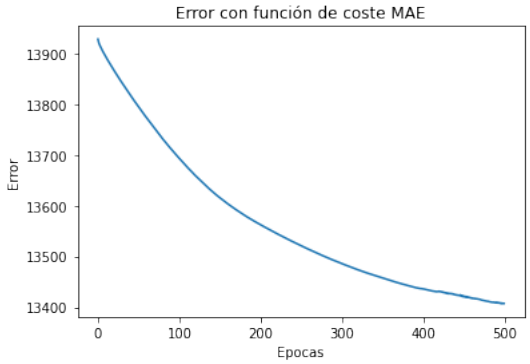

Figure 1: Ajuste de modelo con MAE, $\alpha = 9.9e^{-}7$

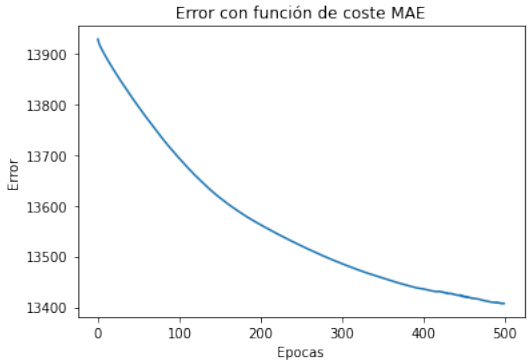

Figure 2: Ajuste de modelo con MSE, $\alpha = 9.9e^{-}9$

Como se puede observar en las gráficas, el error del modelo disminuye conforme se va entrenando entre épocas. Sin embargo, es importante apreciar que para lograr dicho ajuste, la tasa de aprendizaje ha tenido que ser muy pequeña. Esto se observa más cuando utilizamos la función de coste MSE. Esto puede estar asociado con el comportamiento que tiene el modelo, ya que los datos estar compuesto por cuatro variables, agregando además el grado 2, lo que nos da unos 15 coeficientes a calcular. A pesar de ello, el error con los datos de prueba no son tan elevados. Con la función de coste MSE el error ABSOLUTO es de 49871.97, mientras que para el MAE el error es de 36034.31.

3.3 Técnicas de regularización

3.3.1 Regularización L1: Lasso

Aplicando la regularización L1 podemos ver como se ajusta el modelo al manipular el hiperparámetro lambda (λ) para disminuir el error.

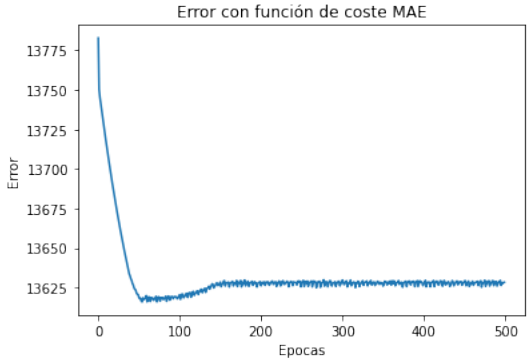

Figure 3: Ajuste de modelo con MAE, $\alpha = 0.000009$, λ=0.005

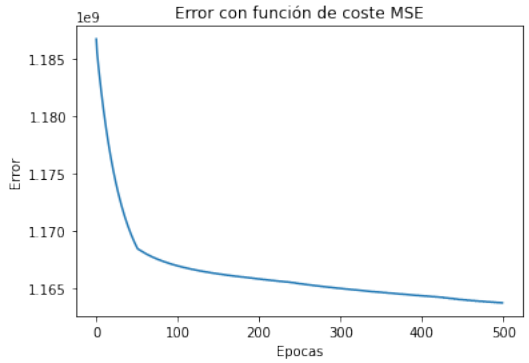

Figure 4: Ajuste de modelo con MSE, $9.9e^{-}9$, λ=0.005

Como se puede observar en las figuras 3 y 4, la regularización L1 ha ayudado a minimizar el error mucho más rápido. Esto se aprecia mucho más en el descenso con la función de coste MAE, donde la convergencia se alcanza en épocas tempranas. Sin embargo, si nos fijamos en el menor error alcanzado por el modelo, con MSE se llegó a un error absoluto respecto a los datos de prueba de 52254.14 y con MAE se alcanzó un error de 37639.26. Comparado con los resultados anteriores (sin L1) se aprecia un error algo mayor. Esto nos podría indicar que si existe alguna correlación entre las variables. Este es porque Lasso hace que algunos coeficientes tiendan a cero, situación que explicaría el relativo aumento en el error.

3.3.2 Regularización L2: Ridge

La regularización L2 es ampliamente conocida. Evaluaremos su desempeño al aplicarlo a nuestro modelo para ajustarse a los datos y disminuir el error.

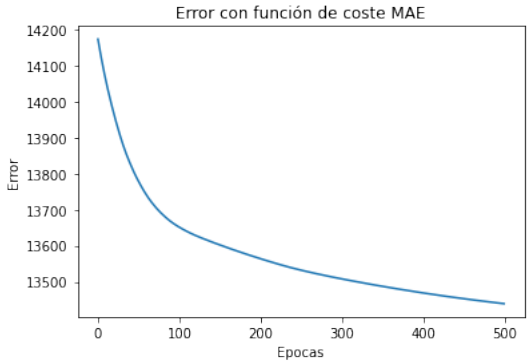

Figure 5: Ajuste de modelo con MAE, α = 0.0000099, λ=0.005

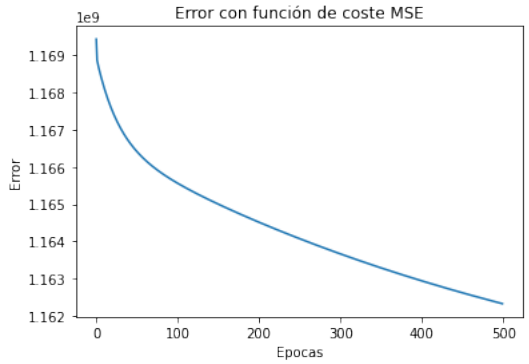

Figure 6: Ajuste de modelo con MSE, $\alpha = 9.9e^{-}9$, λ=0.005

Observando las figuras 5 y 6 se destaca que con Ridge el descenso del error no es tan pronunciado como con Lasso, pero ligeramente más rápido en comparación al descenso sin regularización. Algo a destacar es que con la Ridge el valor de α se elevo ligeramente para el ajuste con la función de coste MAE. En cuanto al error para los datos de prueba, con la función de coste MSE se obtiene un error absoluto de 47992.33, y para el MAE se obtiene 36278.36. Se tiene, por tanto, un error menor al utilizar la regularización L2 en lugar de la L1. Lo que nos sugiere que existe cierta correlación entre las variables involucradas.

3.4 Métodos de optimización de la gradiente

Un método conocido que ayuda a optimizar el descenso de gradiente es el de Momentum. Es un método rápido y mucho más ligero que otros métodos como AdaGrad. Por tanto, observaremos el comportamiento del modelo bajo este método.

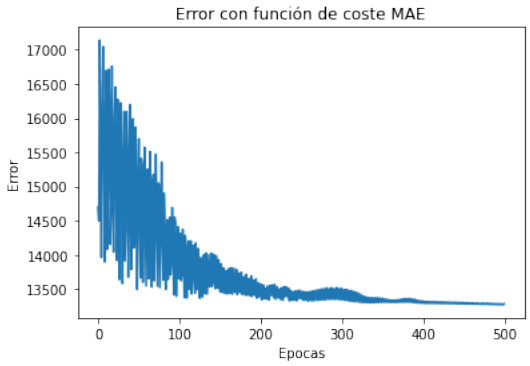

Figure 7: Ajuste MAE con método Momentum, $\alpha = 0.0000099$, γ=0.99

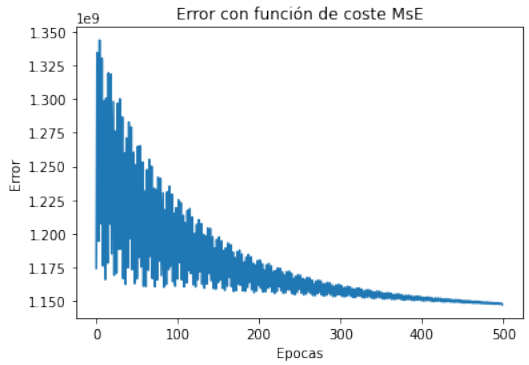

Figure 8: Ajuste MAE con método Momentum, $\alpha = 9.5e^{-}9$, γ=0.99

El comportamiento del error a medida que van pasando las épocas con el método Momentum es muy particular. La teoría indica que este método actúa como una bola que cae, cogiendo momento en ciertas zonas, lo que evita que se quede estancado y con esto se mueve más rápido. Los errores absolutos en las funciones de coste con este método son; para MSE un 35842.32 y para MAE un 33258.83. Es evidente que en ambos casos el error es menor en comparación a las situaciones anteriores. Lo más destacado es que el error con MSE haya bajado de 40000 y se haya situado en un 35842.32. Esto evidencia lo rápido y efectivo que es este método para nuestros datos tan variados.

4 Conclusiones

- La primera conclusión a la que se llega con este trabajo es que el descenso de gradiente puede tener un mejor desempeño con la regularización o con un optimizador de gradiente.
- Tomando en cuenta a las regularizaciones, en base a los resultados se puede afirmar que la regularización L2 provee un menor error de estimación que la L1. Esto indica que los datos si están correlacionados y que en el caso de personas confirmadas con COVID, la zona (y por ende también la latitud y longitud) pueden llegar a contribuir con el contagio.
- Por último, respecto al método de Momentum. Este método de optimización es muy ligero y rápido al momento de realizar el descenso de gradiente. Por lo evaluado en el presente trabajo, se considera que su uso para el análisis de los datos es muy apropiado y, en comparación a los regularizadores sobre la data procesada, brinda mejores resultados.

5 Código fuente del proyecto

Repositorio Colab: Haga clic aquí