**Curso de Regresión Logística con Python y Scikit-learn**

La regresión logística es una herramienta que se complementa muchísimo con la regresión lineal. Es otro algoritmo supervisado en el cuál hace parte de la clasificación, no de regresión. ¿Por qué? Porque es un algoritmo que no busca regresar un valor continuo, sino un valor de es 0 o 1, clasificación de verdadero o falso. A qué clase pertenecen los datos según la característica. Sin dudas, la SIGMOIDE es fundamental. Qué probabilidad tiene de estar en una clase o en otra.

Diagrama

Descripción generada automáticamente

Diagrama

Descripción generada automáticamente

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

**¿CUÁNDO UTILIZAR REGRESIÓN LOGÍSTICA?**

VENTAJAS:

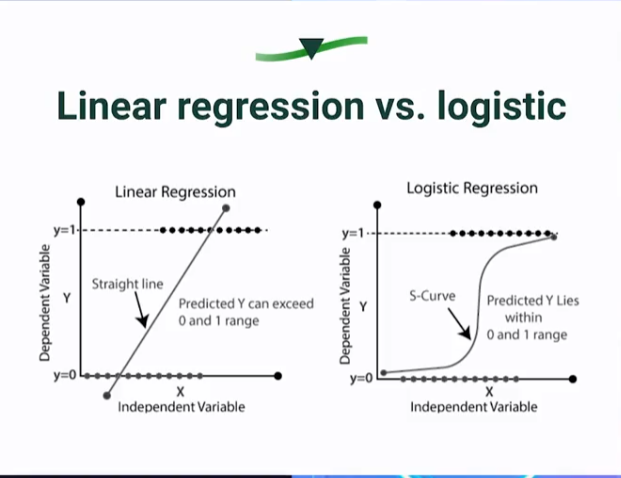
* Sencillo y fácil de implementar
* Se comporta muy bien en entornos donde las variables son linealmente separables
* Coeficientes que son interpretables
* Hay inferencia de la importancia de cada variable
* Clasifica en porcentajes, de ahí que se basa en la sigmoidal

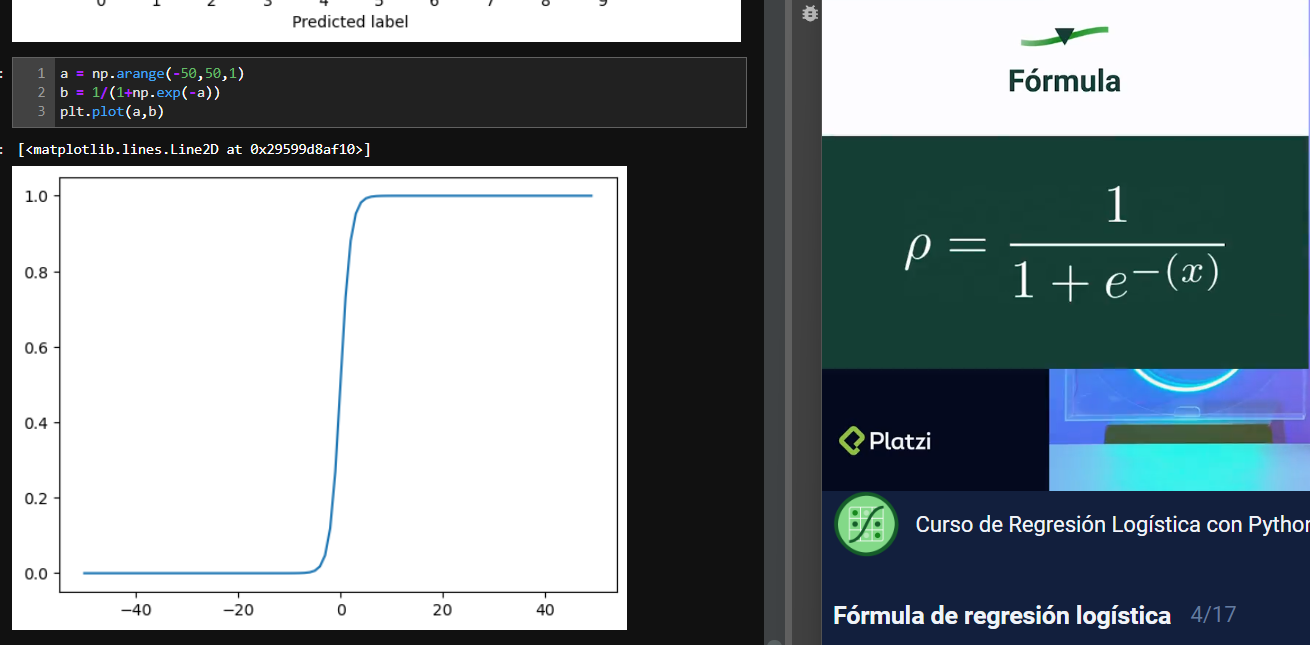
DESVENTAJAS:

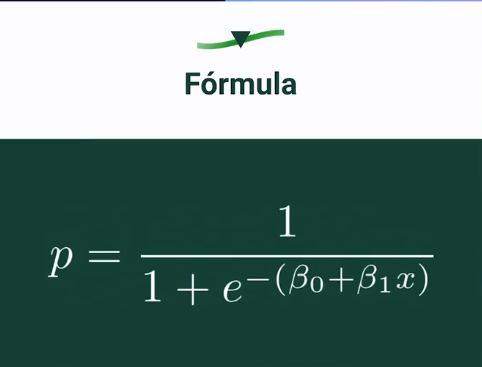
* Es susceptible al overfitting con datasets de alta dimensionalidad
* Asume linealidad entre las variables dependientes. Pero, cuando no es así, va a fallar y es mejor usar otro algoritmo
* No funciona bien con pocos datos, es mucho mejor con grandes datasets
* Le afecta la multicolinealidad de las variables

¿CUÁNDO ENTONCES?

* Si se buscan las probabilidades de ocurrencia de un evento categórico
* Dataset linealmente separable
* Dataset grande
* Dataset balanceado en lo que me interesa clasificar (50% una cosa y 50% la negación
* Algo fácil y rápido







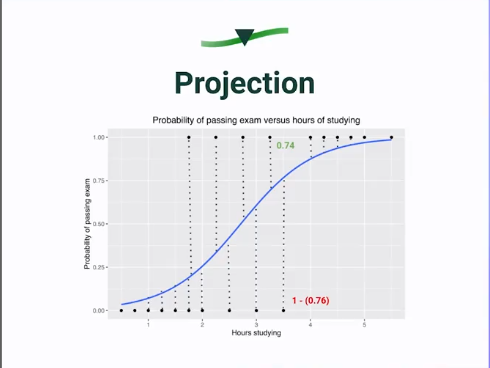
Por eso es que la regresión lineal va a resolver el problema en la regresión logística

Gráfico, Gráfico de líneas

Descripción generada automáticamente

**EVALUACIÓN DEL MODELAMIENTO (MLE):**

¿Cómo es que obtiene ese valor, cómo itera el resultado?



La idea central yace en que la regresión logística dibuja una línea sigmoidal apoyada por la función lineal simple en la cuál hay puntos que se ubican en el eje y a lo largo de los valores 0 y 1 como se muestra en la gráfica. Cuando el punto está en el 1 la probabilidad es 1 menos la distancia que existe a la curva de forma positiva y cuando es negativa es 0 menos la distancia a la curva.

Hay un algoritmo que se llama el Maximum Likelihood Estimator o lo que en español es conocido como la Máxima Verosimilitud.

Interfaz de usuario gráfica, Texto, Aplicación, Chat o mensaje de texto

Descripción generada automáticamente

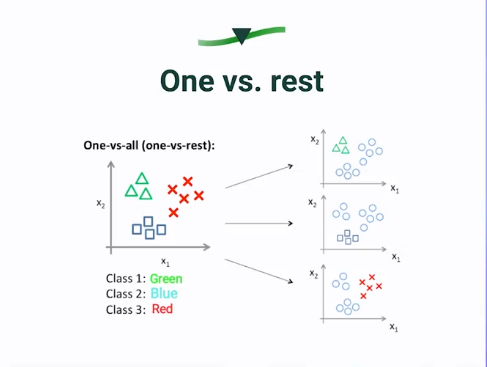
Se hace una suma ponderada de todas las probabilidades y a éstas sumas se les hace un log. Cuando más alto sea el número del MLE es mucho mejor la optimización, es buscar el máximo.

EL otro método es el Gradiente Descendente, que disminuye la función de costo con respecto a la sigmoidal con derivadas parciales hasta encontrar el punto más bajo.

**REGURALIZADORES**

Debido al overfitting se busca regularizar dicho problema, reducir la complejidad de los datos disminuyendo el peso de los W en los parámetros de la regresión logística. Se podría entender como un costo extra que se agrega a la de costo normalmente entendida

**REGRESIÓN LOGÍSTICA MULTICLASE**

* One vs Rest: Hay clases que se disputan entre ellos, evaluada con respecto al resto. Es decir: binomial. O es triángulos o no lo es. Es cuadrado o no. Es cruz o no. Etc…
* 

Si es un problema múltiple pero lo convierto en binomial es posible que se logre con una logística.

* Multinomial logistic classifier: Linear model, Softmax y Cross Entropy:

No hay tres algoritmos sino todos contra todos se van a evaluar, en la salida lo que tengo son tantas salidas según tenga clases

