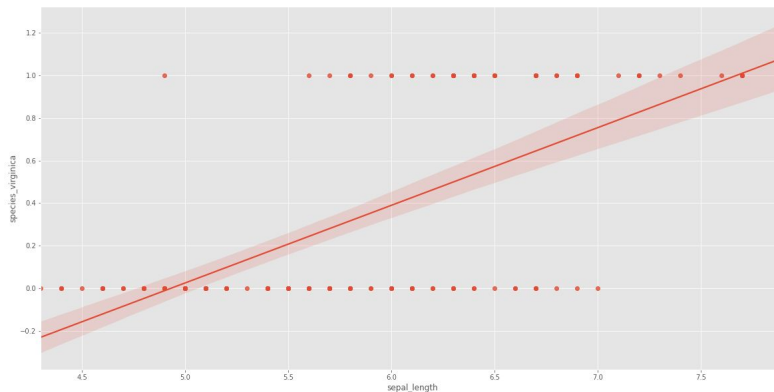


Regresión Logística

Regresión Logística

La regresión logística es uno de los algoritmos de aprendizaje supervisado más utilizados. Es un modelo de clasificación lineal, que se basa en los mismos principios que la regresión lineal.

Sin embargo es importante mencionar que si bien la regresión lineal posee un potencial clasificativo su desempeño suele no ser el ideal en comparación a sus capacidades regresivas como lo vimos anteriormente.



Regresión Logística - Explicación

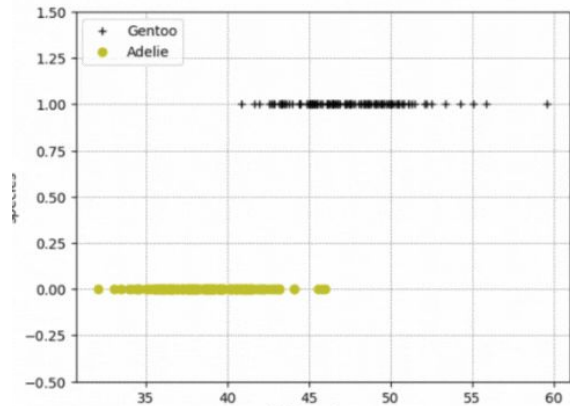
En una regresión lineal la variable dependiente es una variable métrica, por ejemplo, el salario o las notas obtenidas por X horas de estudio.

En la regresión logística la variable dependiente es una variable dicotómica, si o no, por ejemplo es una enfermedad está presente donde 0 podría significar enfermo y 1 no enfermo.

Así, nuestro **conjunto de datos** podría tener este aspecto:

Aquí tenemos las **variables independientes** y aquí la **variable dependiente** con 0 y 1.

Edad	Sexo	Hábito de fumar	Enfermedad
22	mujer	No fumadora	1
25	mujer	Fumadora	1
18	hombre	Fumador	0
45	hombre	No fumador	0
12	mujer	Fumadora	0
43	hombre	Fumador	1
23	hombre	Fumador	0
33	hombre	Fumador	1



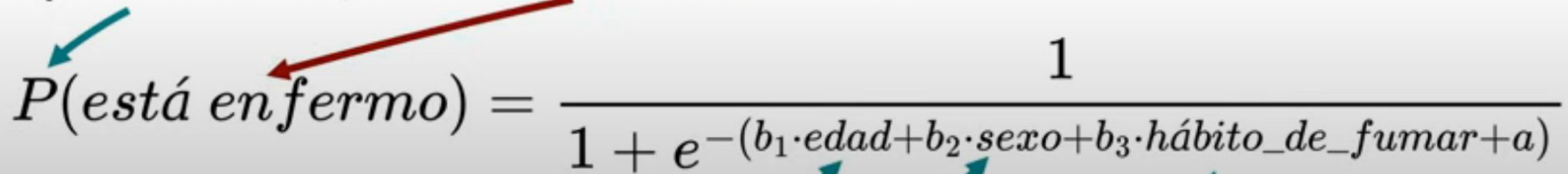
Regresión Logística - Explicación

En resumen la función logística es una función que para el eje x sólo toma valores entre 0 y 1 por lo tanto es ideal para clasificación binaria.

$$f(x) = \frac{1}{1 + e^{-x}}$$

En nuestro ejemplo,

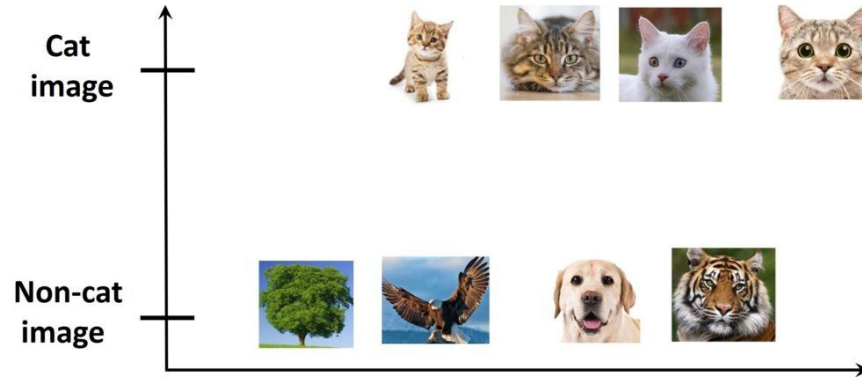
la **probabilidad** de padecer una **cierta enfermedad**


$$P(\text{está enfermo}) = \frac{1}{1 + e^{-(b_1 \cdot \text{edad} + b_2 \cdot \text{sexo} + b_3 \cdot \text{hábito_de_fumar} + a)}}$$

es función de la **edad**, el **sexo** y el **hábito de fumar**.

Regresión Logística - implementación

Utilizaremos algoritmos de Machine Learning en Python para resolver un problema de Regresión Logística. A partir de un conjunto de datos de entrada (características), nuestra salida será discreta (y no continua) por eso utilizamos Regresión Logística (y no Regresión Lineal). La Regresión Logística es un Algoritmo Supervisado y se utiliza para clasificación.



Regresión Logística

Vamos a clasificar problemas con dos posibles estados “SI/NO”: binario o un número finito de “etiquetas” o “clases”: múltiple. Algunos Ejemplos de Regresión Logística son:

- Clasificar si el correo que llega es Spam o No es Spam
- Dados unos resultados clínicos de un tumor clasificar en “Benigno” o “Maligno”.
- El texto de un artículo a analizar es: Entretenimiento, Deportes, Política ó Ciencia
- A partir de historial bancario conceder un crédito o no

Utilizaremos nuevamente la implementación del paquete sklearn en Python para ponerlo en práctica.

Regresión Logística

<https://colab.research.google.com/drive/1rk6GM9DeNHkweOSBrEU3cFdBx0ERJBSd?usp=sharing>