



# REGRESIÓN LOGÍSTICA



## \* CLASIFICADOR BINARIO O MULTICLASES

- \* Modelo de aprendizaje supervisado

- \* Incluido dentro de los Modelos Lineales Generalizados

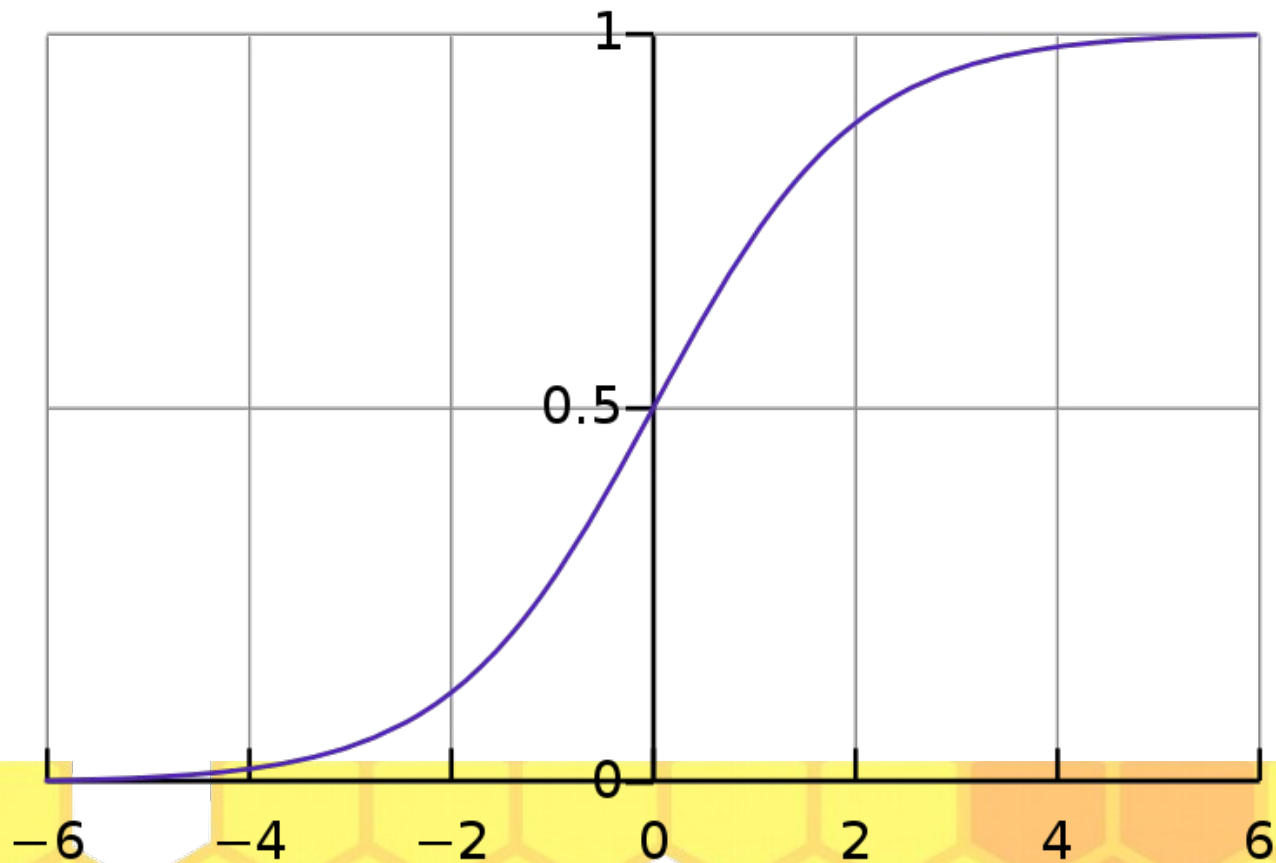
- \* Modela sobre datos con distribución tipo Bernoulli, es decir, el resultado sólo puede Éxito o Fracaso (o sea, la salida es DISCRETA)

- \* La distribución de probabilidades tiene la forma de una función Sigmoide ( $1/(1+\exp(-X))$ ), es decir,

$$\text{logit}(\pi) = \ln(\pi/(1-\pi)) = b_0 + b_1x_1 + b_2x_2 + \dots$$

donde  $x_i$  son las características predictoras (nuestros datos)

# Función Sigmoide



# APLICACIONES

- Decidir si el correo entrante es Spam o no
- Diagnóstico médico: x ej. Decidir si un tumor es benigno o no
- Clase del texto de un artículo: Deporte, Entretenimiento, Política, Ciencia...
- Scoring crediticio a partir del historial previo
- Y muchos casos más...

# USAMOS LA BIBLIOTECA SCIKIT LEARN DE PYTHON

## ACÁ LA DOCUMENTACIÓN OFICIAL

[https://scikit-learn.org/stable/modules/generated/sklearn.linear\\_model.LogisticRegression.html](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html)

### VAMOS A UN EJEMPLO:

Queremos predecir el sistema operativo que usan los usuarios analizados, usando diversos datos de una web que utiliza GOOGLE ANALYTICS:

- \* duración de la visita en segundos
- \* cantidad de páginas visitadas en la sesión
- \* cantidad de acciones del usuario
- \* suma del valor de las acciones (asignada por la analítica web)



Preguntas?

Muchas Gracias!

