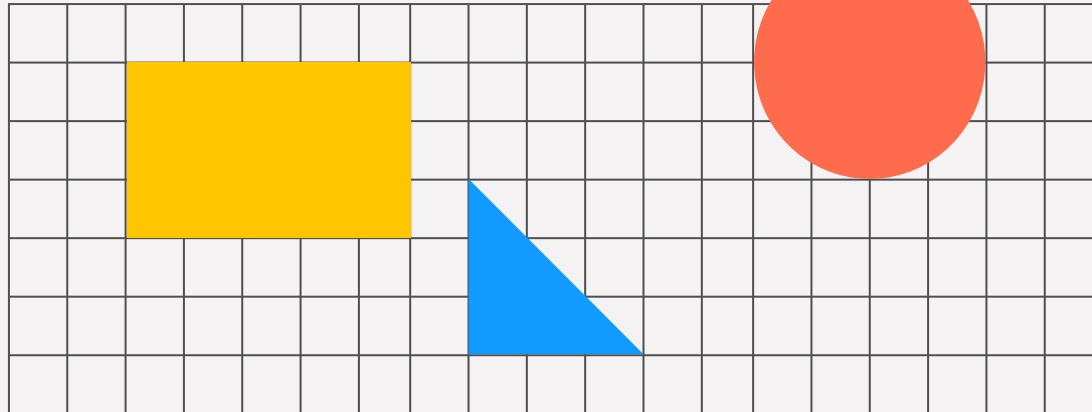
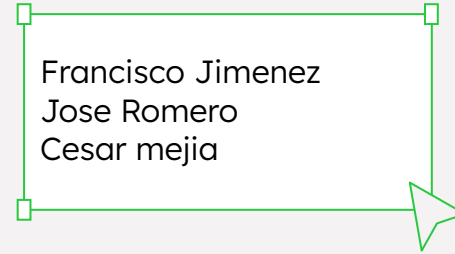


Regresión Logística



Regresión Logística para Detección de Spam

La regresión logística es un algoritmo de aprendizaje supervisado para problemas de clasificación binaria.

Aunque tiene "regresión" en el nombre, en realidad es un clasificador

Aplicaciones comunes:

- Detección de spam
- Diagnóstico médico
- Análisis de riesgo crediticio
- Sistemas de recomendación (like/dislike)

Predecir si un correo es spam (1) o no spam (0) basado en sus características

Objetivo

Fundamentos de Regresión Logística

¿Cómo Funciona la Regresión Logística?

Función sigmoide: Transforma señales en probabilidad:

Si el correo tiene palabras como "oferta", "urgente", o muchos signos de exclamación, la regresión logística combina estas pistas y las convierte en un % de probabilidad.

Formula aplicada al spam

$$P(y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)}}$$

- $\beta_0 + \beta_1 x_1 + \dots$: Es una suma ponderada (como en regresión lineal).
- $e^{-(\dots)}$: Ajusta la curva para que siempre dé probabilidades.

Predecir la probabilidad de que un correo sea spam (ej: "¡Gana un millón ahora!") basado en sus características.

Objetivo

Requisitos de los datos

- **Etiquetado previo:** Necesitas correos ya clasificados como "spam" o "no spam".
- **Variables numéricas:** Las palabras se convierten en números (ej: "1" si "oferta" aparece, "0" si no).

Ejemplo concreto:

Imagina un correo con:

- **3 apariciones de "oferta"** ($\beta_1 = 2.0$).
- **1 aparición de "urgente"** ($\beta_2 = 3.0$).
- **Término base** ($\beta_0 = -5.0$).

Cálculo:

$$P(\text{spam}) = \frac{1}{1 + e^{-(-5.0 + 2.0 \cdot 3 + 3.0 \cdot 1)}} = \frac{1}{1 + e^{-4}} \approx 98\% (\text{¡Spam!})$$



Función sigmoide

Vamos a crear una gráfica de una **función sigmoide** para explicar cómo se usa en la **validación de correos spam** mediante un modelo de clasificación binaria (por ejemplo, regresión logística)

La función **sigmoide** transforma un valor (por ejemplo, la salida de un modelo) en una probabilidad entre 0 y 1.

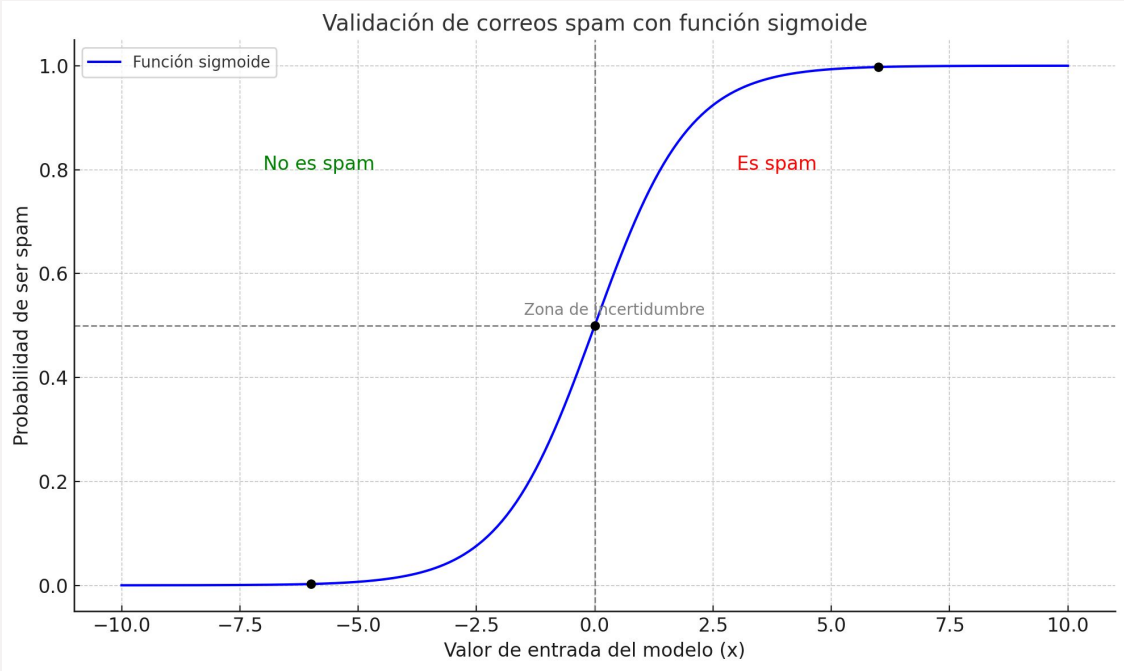
Esta probabilidad se interpreta como:

- Cerca de **1** → Alta probabilidad de que **sea spam**
- Cerca de **0** → Alta probabilidad de que **no sea spam**
- Cerca de **0.5** → Zona de incertidumbre

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

Formula

Gráfica de la función sigmoide aplicada a la detección de correos spam



A la izquierda (valores negativos), la probabilidad es cercana a **0** → el modelo considera que **no es spam**.

- A la derecha (valores positivos), la probabilidad es cercana a **1** → el modelo considera que **sí es spam**.
- En el centro (cerca de 0.5), hay **incertidumbre** y puede requerir más criterios para decidir.

Características Clave



- **Salida clara:**

Si $P(\text{spam}) \geq 0.5 \rightarrow$ **"Spam"** (ej: 75% de probabilidad).

Si $P(\text{spam}) < 0.5 \rightarrow$ **"No spam"** (ej: 10% de probabilidad).

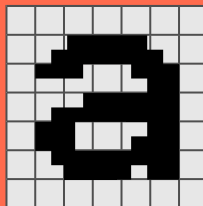
- **Relación lineal:**

- Funciona si palabras como "gratis" aparecen **proporcionalmente** más es spam.
- *No sirve* para detectar patrones complejos como el estilo de escritura.



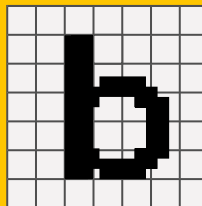
Visualización

Nota: El modelo "aprende" los pesos (β) durante el entrenamiento, ajustándose para **minimizar errores** en los datos históricos.



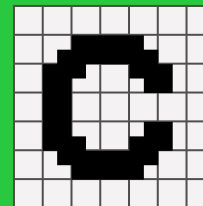
Izquierda
(valores negativos)

Correos normales
("Hola, ¿cómo estás?")



Cetro (0)

Límite de decisión
(ej: "Reunión
mañana").



Derecha
(valores positivos)

Spam claro ("¡Gana
dinero ya!").

Aplicaciones de la Regresión Logística

Salud



Diagnóstico de enfermedades: predecir si un paciente tiene una enfermedad o no, en base a síntomas o análisis (por ejemplo, probabilidad de tener diabetes).

Eficacia de tratamientos: estimar si un tratamiento tendrá éxito (1) o fracaso (0) en un paciente

Finanzas



Aprobación de créditos: determinar si un cliente es apto (1) o no apto (0) para recibir un préstamo.

Probabilidad de impago (default): evaluar si un cliente probablemente no pagará su deuda.

Marketing



Probabilidad de compra: predecir si un cliente hará clic en un anuncio o comprará un producto.

Segmentación de clientes: clasificar clientes como “potenciales compradores” o no.

Seguridad y criminología



Predicción de reincidencia: estimar si una persona con antecedentes volverá a delinquir.

Clasificación de correos electrónicos: detectar si un email es spam (1) o no (0).

Inteligencia artificial y machine learning



Modelos de clasificación binaria: por ejemplo, en reconocimiento facial, detectar si en una imagen hay una cara (1) o no (0).

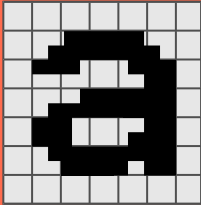
Detección de fraudes: identificar transacciones sospechosas.

Educación

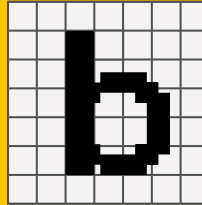


Evaluación de rendimiento estudiantil: predecir si un alumno aprobará (1) o no (0) un curso, basándose en sus características académicas o hábitos de estudio.

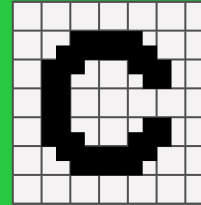
Demostraciones



Sklearn



Pandas



Numpy