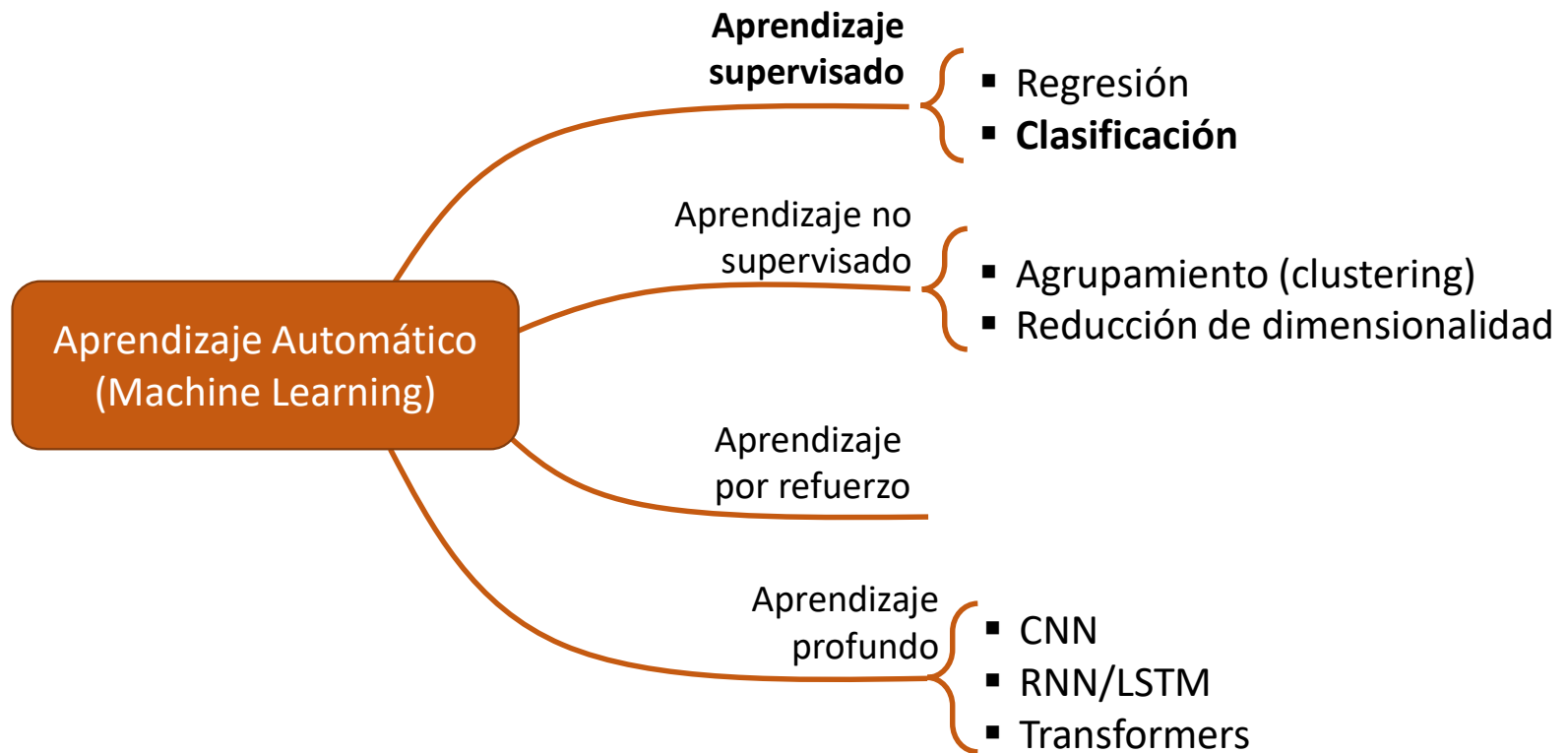


Maestría en Inteligencia Artificial Aplicada

Regresión Logística

Inteligencia Artificial y Aprendizaje Automático



Tipos de variable dependiente (o de salida)
para diferentes tipos de problemas:

- Reales (continuas):
 - Costo de una casa
 - Monto de crédito autorizado
 - Presión arterial
- Discretas/Categóricas:
 - Binarias:
 - Crédito aprobado: Sí/No
 - Comentario positivo: Sí/No
 - Calidad de un producto: Aprobado/No_Aprobado
 - Múltiples:
 - Productos más vendidos de un empresa en línea
 - Tipo de anuncio que más impacto a generado en redes sociales
 - Análisis de sentimiento de un servicio en 5 niveles

En esta sección estudiaremos el caso de la variable de salida binaria con distribución de probabilidad Bernoulli.

Distribución de Probabilidad

Al conjunto de parejas ordenadas $(x, f(x))$ le llamaremos Función de Probabilidad, Función Masa de Probabilidad o Distribución de Probabilidad de la v.a. discreta X , si para cada resultado posible $X = x$, f es una función que satisface las siguientes condiciones:

$$1) \quad f(x) \geq 0$$

Condición de No Negatividad.

$$2) \quad \sum_{x \in S} f(x) = 1$$

Esta condición nos asegura que algún elemento del Espacio Muestral deberá de ocurrir.

$$3) \quad P(X = x) = f(x)$$

Esta condición nos dice que la función f deberá tener un significado probabilístico.

Estas dos primeras condiciones implican además que

$$0 \leq f \leq 1,$$

para el caso de una Distribución Discreta.

Función Logística

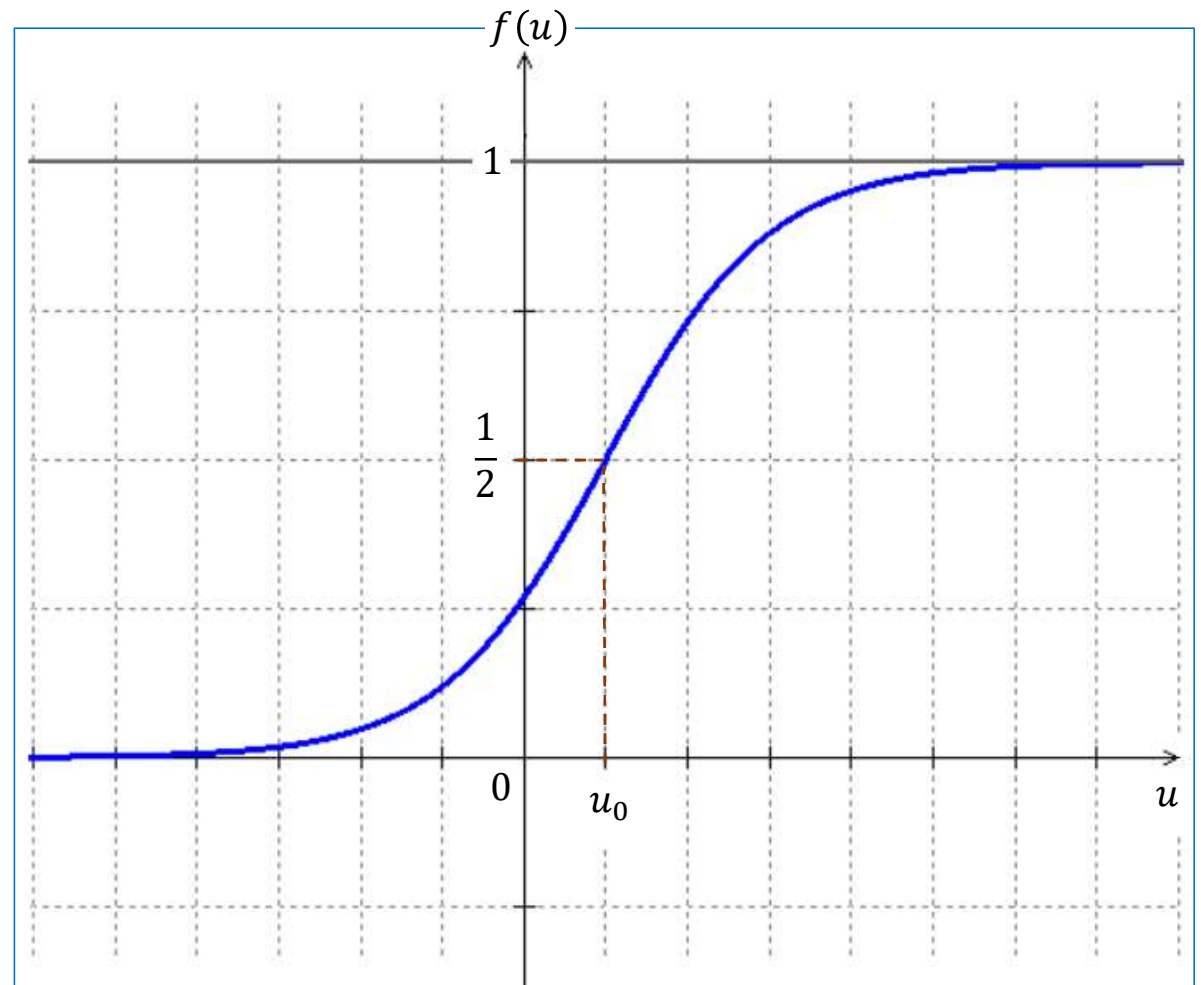
$$f(u) = \frac{1}{1 + e^{-\frac{u-u_0}{\delta}}}$$

donde:

u_0 : media de la v.a. u

δ : proporcional a la desviación estándar

Llamada en estadística Distribución Logística ya que cumple con las característica de una función de distribución de probabilidad acumulada.



Dominio: $(-\infty, +\infty)$

Rango: $(0, +1)$

Función Sigmoide

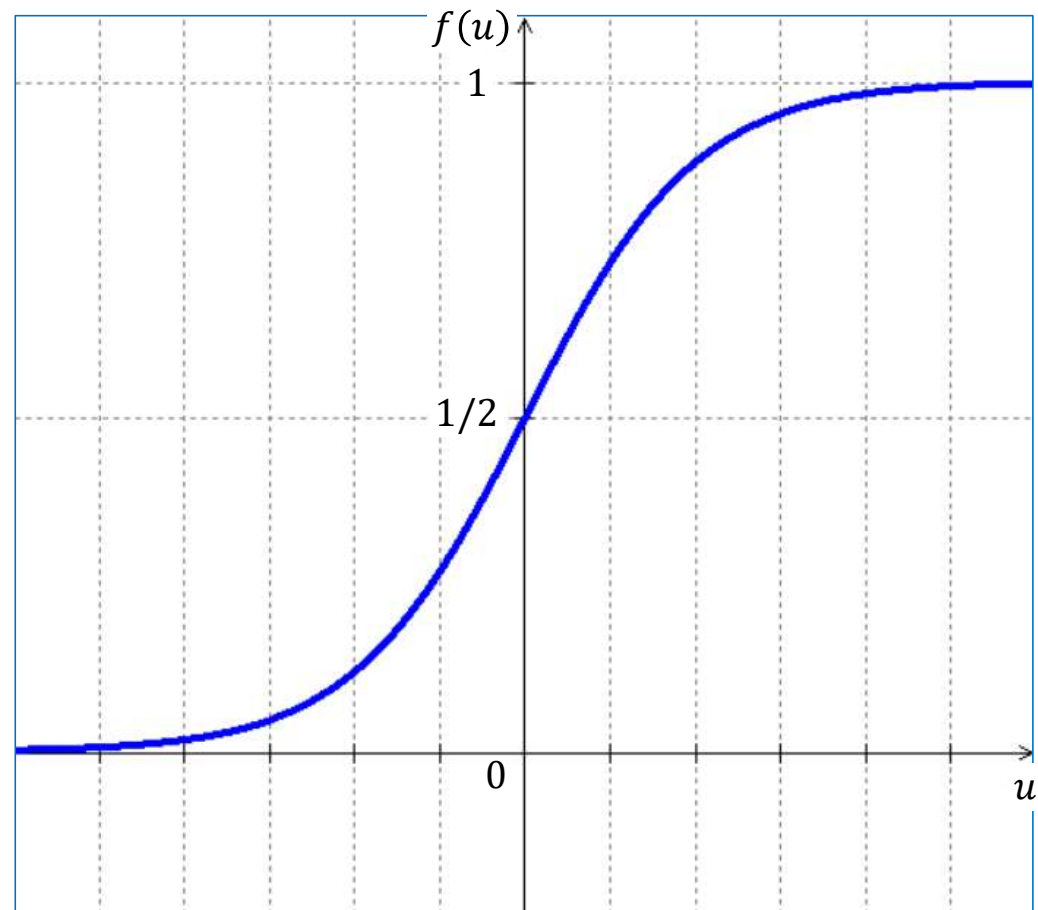
$$f(u) = \frac{1}{1 + e^{-u}}$$

O bien,

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

También llamada **función sigmoide exponencial**.

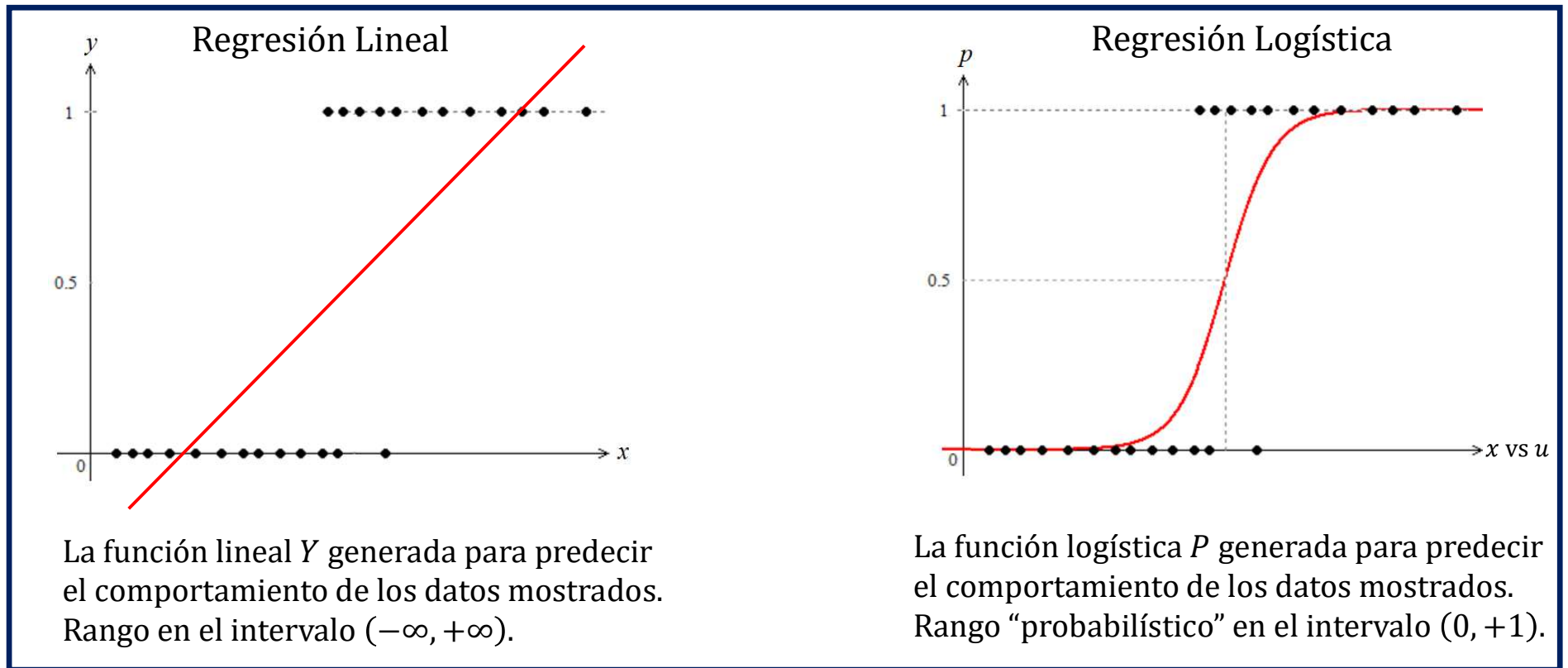
Observamos que la función sigmoide es un caso particular de la función logística, aunque muchos autores a esta función también la llaman función logística.



Dominio: $(-\infty, +\infty)$

Rango: $(0, +1)$

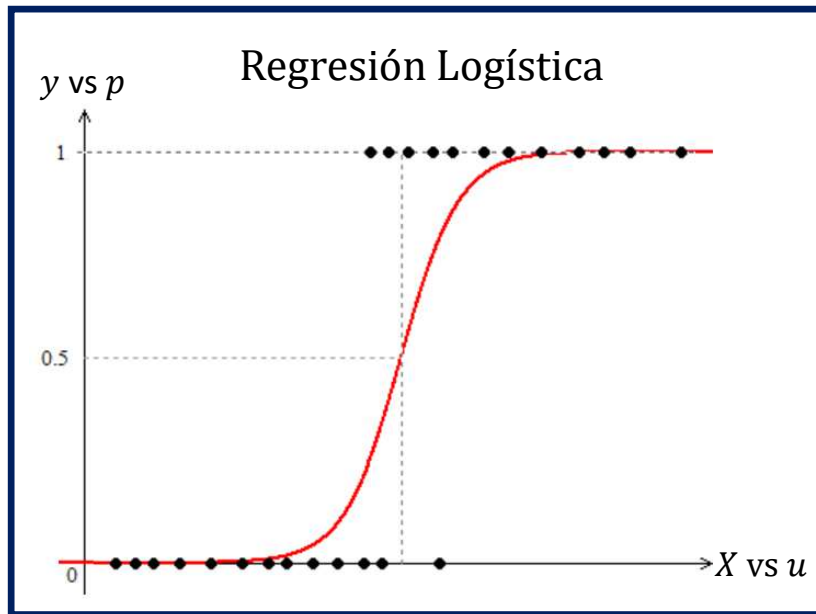
La función logística como un aproximador de la función escalón



Aún cuando matemáticamente se puede utilizar el modelo de regresión lineal para encontrar la ecuación de la recta que aproxime la salida de un problema de clasificación, los resultados no serían muy satisfactorios.

La función logística permitirá aproximar con un sentido probabilístico los valores de predicción en un problema de clasificación.

Método de Optimización en un problema de Clasificación mediante una Función de Costo



En el caso del modelo de regresión logística para clasificación, se busca la función logística $p = f(X)$ que mejor aproxime los puntos dados (x_k, y_k) .

Consideremos el conjunto de **datos de entrada**:

$$\{(x_{k1}, x_{k2}, \dots, x_{km}, y_k)\}_{k=1}^N$$

de un problema de clasificación binario (logístico), donde las etiquetas $y_k \in \{0, 1\}$. El total de factores es m y la cantidad de registros es N . Entonces, se desea encontrar los parámetros $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_m)$ de un modelo que denotamos

$$\hat{y} = h_{\beta}(X)$$

donde $X_{N \times (m+1)} = [1, x_1, x_2, \dots, x_m]$ y que minimice el valor de una **función de costo** J .

Simbólicamente lo podemos expresar este problema de optimización como:

$$\min_{\beta} J(\beta)$$

Modelo de Regresión Lineal Generalizado

Generalized Linear Model (GLM):

Supuestos sobre la variable dependiente de salida:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m + \text{error}$$

```
graph LR; Eq["y = beta_0 + beta_1 x_1 + ... + beta_m x_m + error"]; subgraph Assumptions; A1["Distribución gaussiana de los errores:  
Modelo de Regresión Lineal Estándar"]; A2["Distribución de los errores diferente a la gaussiana:  
Modelo de Regresión Lineal Generalizado"]; A3["En particular al usar la función Logística:  
Modelo de Regresión Logística"]; end; Eq -- "y" --> A1; Eq -- "y" --> A2; Eq -- "error" --> A1; Eq -- "error" --> A2; A2 --> A3;
```

Distribución gaussiana de los errores:
Modelo de Regresión Lineal Estándar

Distribución de los errores diferente a la gaussiana:
Modelo de Regresión Lineal Generalizado



En particular al usar la función Logística:
Modelo de Regresión Logística

Función sigmoide: $f(u) = \frac{1}{1 + e^{-u}}$ $\left\{ \begin{array}{l} \text{Dominio: } (-\infty, +\infty) \\ \text{Rango: } (0, 1) \end{array} \right.$

Mediante un poco de manipulación algebraica podemos obtener la inversa de la función sigmoide:

$$p = \frac{1}{1 + e^{-u}}$$

$$1 + e^{-u} = \frac{1}{p}$$

$$e^{-u} = \frac{1}{p} - 1$$

$$e^{-u} = \frac{1 - p}{p}$$

$$\ln(e^{-u}) = \ln\left(\frac{1 - p}{p}\right)$$

$$-u = \ln\left(\frac{1 - p}{p}\right)$$

$$u = -\ln\left(\frac{1 - p}{p}\right)$$

$$u = \ln\left(\frac{p}{1 - p}\right)$$

Función Logit:
función inversa
de la sigmoide.
Dominio: $(0, 1)$
Rango: $(-\infty, +\infty)$

Función logit

$$\text{logit}(p) \equiv \ln\left(\frac{p}{1-p}\right)$$

donde $0 < p < 1$

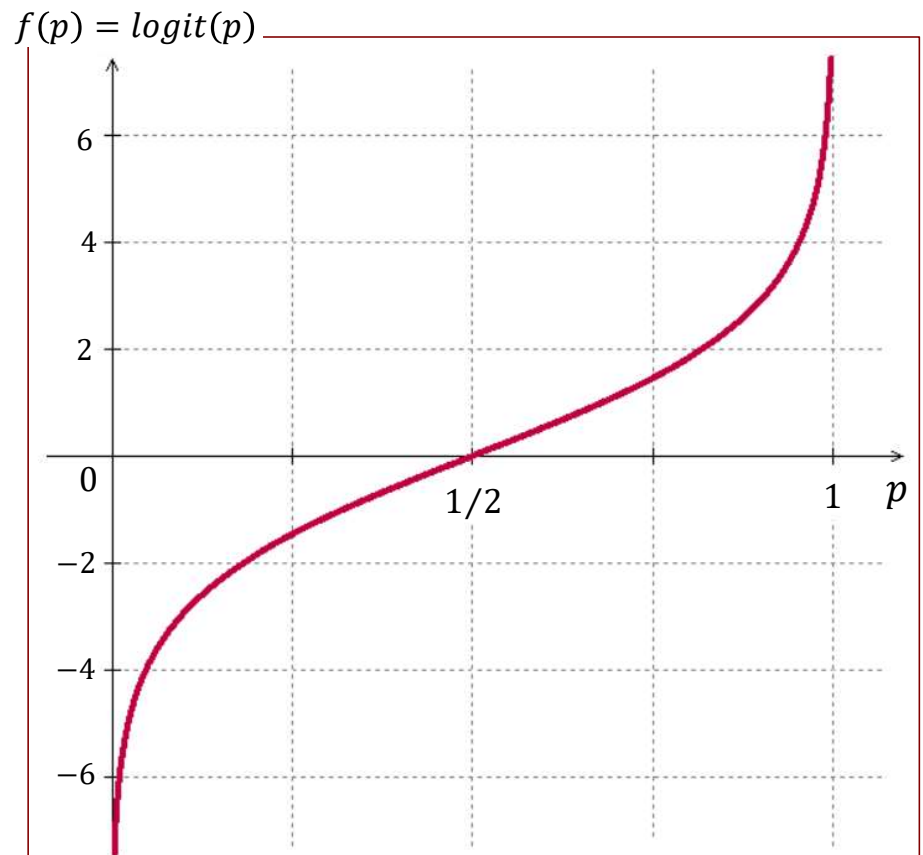
En la literatura Estadística la variable p está asociada a una probabilidad.

Al cociente

$$\frac{p}{1-p}$$

se le llama razón de oportunidades (*odds ratio*).

Observamos que la función *logit* es la función inversa de la función sigmoide.



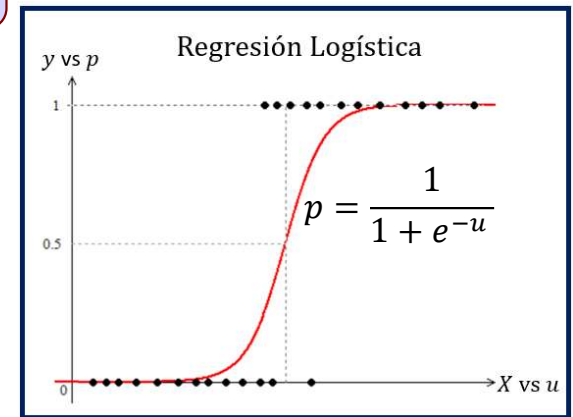
Modelo de regresión logístico multivariable para la variable biclase Y :

$$\text{logit}(p) \equiv \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m$$

Se considera a la variable dependiente Y con distribución Bernoulli, de manera que:

$$\text{Prob}[Y = y \mid x_1, x_2, \dots, x_m] = \begin{cases} p & \text{si } y = 1 \\ 1 - p & \text{si } y = 0 \end{cases}$$

Formalmente: $\text{logit}(\mathbb{E}[Y \mid x_1, x_2, \dots, x_m]) = \text{logit}(p) = \ln(p/(1-p))$.



Al usar el logaritmo estamos extendiendo el Rango del intervalo $(0, 1)$ al intervalo $(-\infty, +\infty)$.

La manera de encontrar los coeficientes de regresión β_k es mediante alguno de los algoritmos numéricos, usualmente de la familia quasi-Newton.

Para variables cuantitativas, al aumentar una unidad en el factor x_j cuando los demás factores permanecen constantes, aumenta el logaritmo del cociente (odds) $\frac{p}{1-p}$ en β_j unidades.

Distribución de Bernoulli y Binomial

Una variable aleatoria discreta X se dice que sigue un modelo dicotómico o de Bernoulli, si al realizar un único experimento aleatorio sus posibles resultados son 1 o 0, con probabilidades p y $1 - p$, respectivamente.

La función probabilidad de Bernoulli está dada como:

$$f(x) = \begin{cases} p & \text{si } x = 1 \\ 1 - p & \text{si } x = 0 \\ 0 & \text{en otro caso} \end{cases}$$

La distribución Binomial es el caso en el cual se tienen n experimentos Bernoulli independientes entre sí.

Nota: Bernoulli es el caso particular de la distribución Binomial con $n = 1$.