Regresión Logística

Clasificación Binaria

mean radius	target
17,99	1
20,57	1
16,13	1
19,81	1
13,54	0
13,08	0
15,34	1
21,16	1
17,57	1
11,84	1
17,02	1
12,05	0
13,49	0
11,76	0
13,64	0
11,94	0
18,22	1

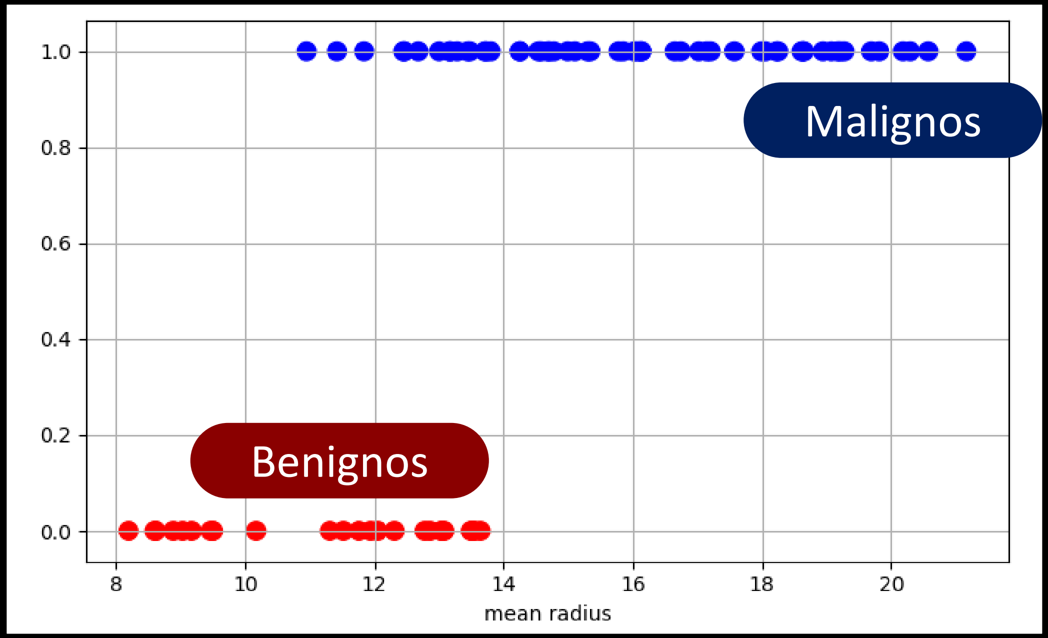
Objetivo:

Clasificación Binaria

Clasificar tumores Benignos (0) y Malignos (1) en base a su tamaño.

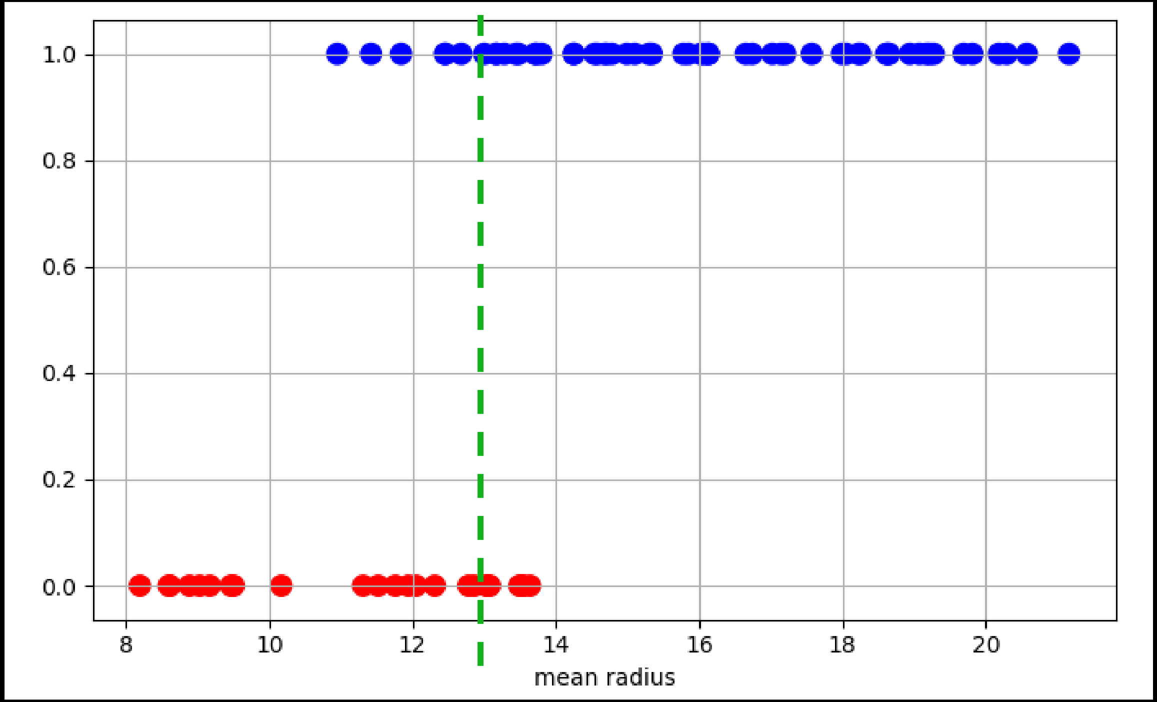
Clasificación Binaria

mean radius	target
17,99	1
20,57	1
16,13	1
19,81	1
13,54	0
13,08	0
15,34	1
21,16	1
17,57	1
11,84	1
17,02	1
12,05	0
13,49	0
11,76	0
13,64	0
11,94	0
18,22	1



Clasificación Binaria

¿Cómo podemos separar las clases?



Ej: Clasificador ad hoc

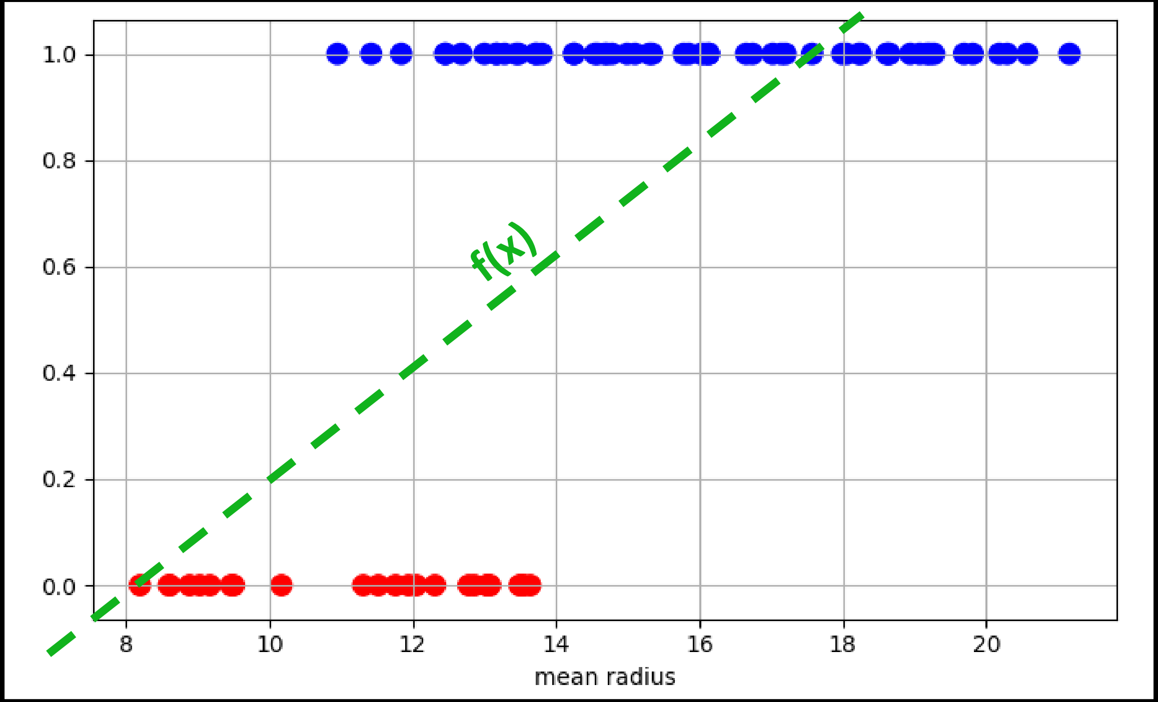
$$f(x) = \begin{cases} 1 & si \;\; x > 13 \\ 0 & si \;\; x \leq 13 \end{cases}$$

Problemas:

- No es derivable
- Muy restrictivo

Clasificación Binaria

¿Cómo podemos separar las clases?

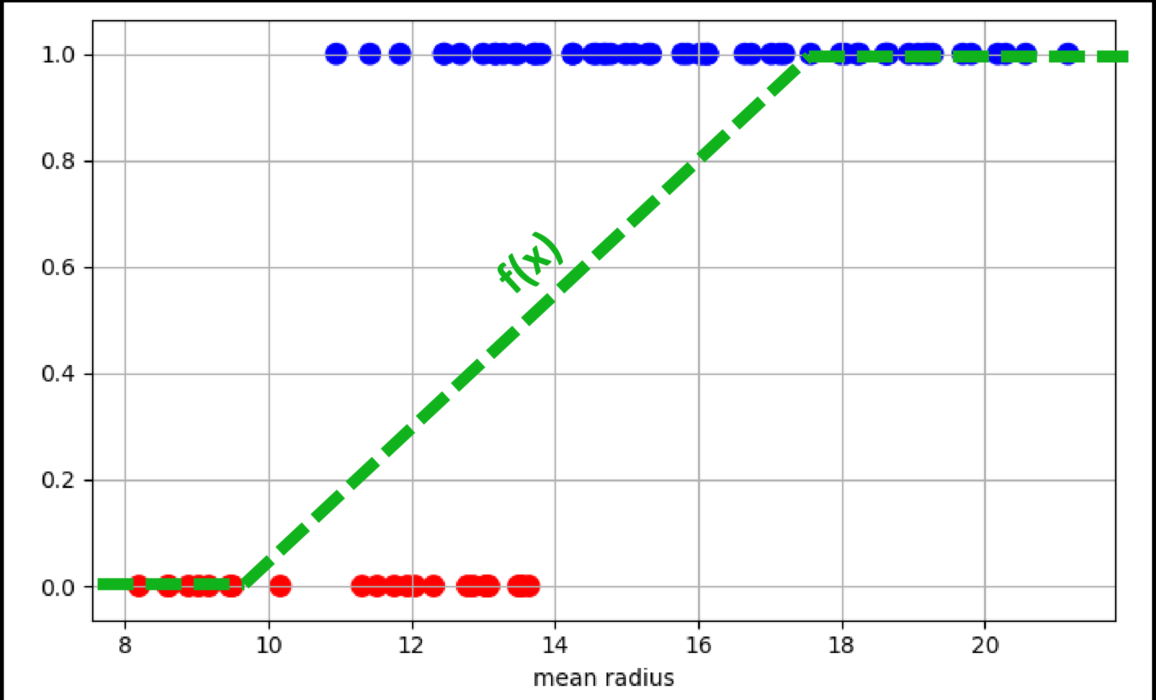


Regresión lineal:

- Valores fuera de rango
- Solución: acotar

Clasificación Binaria

¿Cómo podemos separar las clases?



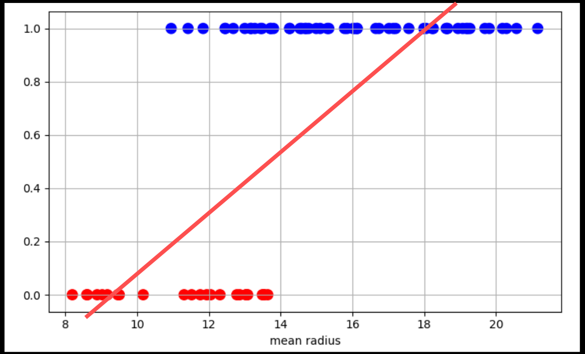
Regresión lineal acotada:

$$g(x) = \begin{cases} 1 & si \ \ x \geq 17 \\ f(x) & si \ 17 > x > 9 \\ 0 & si \ x \leq 9 \end{cases}$$

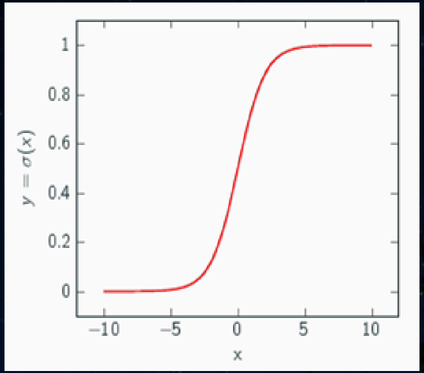
- Difícil de optimizar

Clasificación Binaria

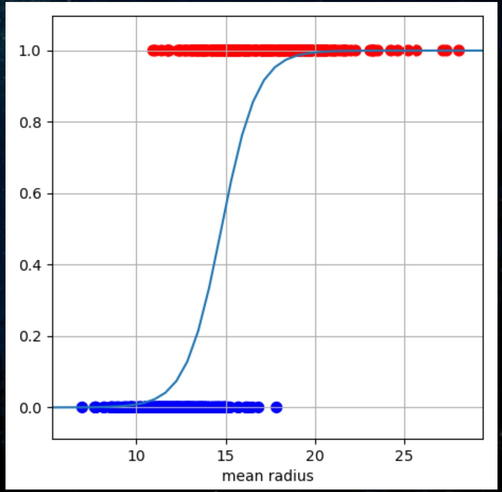
Regresión Logística = Regresión Lineal + Función Logística



+



=



Función Logística o Sigmoidea

$$\boldsymbol{\sigma}(z) = \frac{1}{1 + e^{-z}}$$

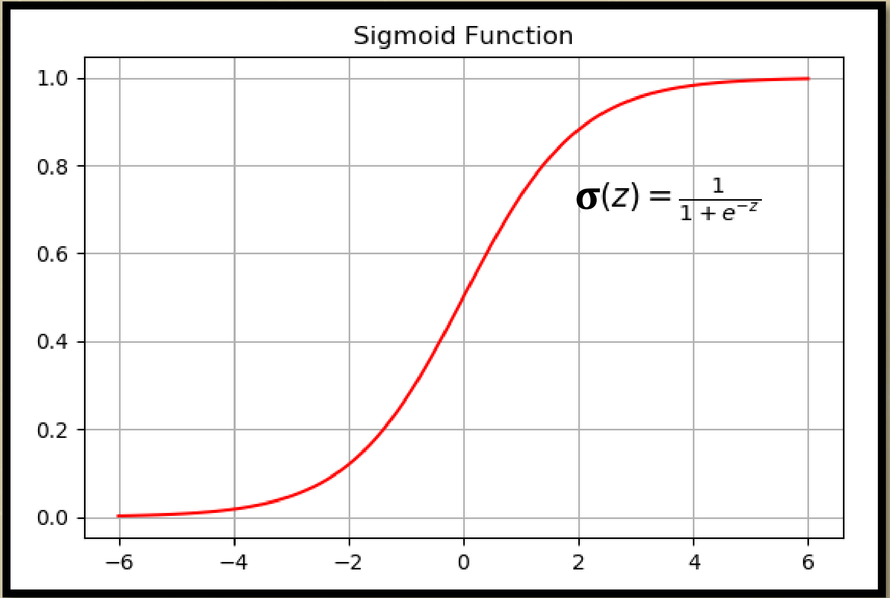
$$\boldsymbol{\sigma}(f(x)) = \frac{1}{1 + e^{-wx-b}}$$

- Tiene un rango de salida entre 0 y 1.
- La función de Error es diferenciable.

  Derivada= $=\boldsymbol{\sigma}(z) * (1 - \boldsymbol{\sigma}(\mathrm{z}))$

- $\boldsymbol{\sigma}$(z) tiene dos asíntotas horizontales:

  Tiende a 1 cuando x tiende a $+\infty$
  Tiende a 0 cuando x tiende a $-\infty$



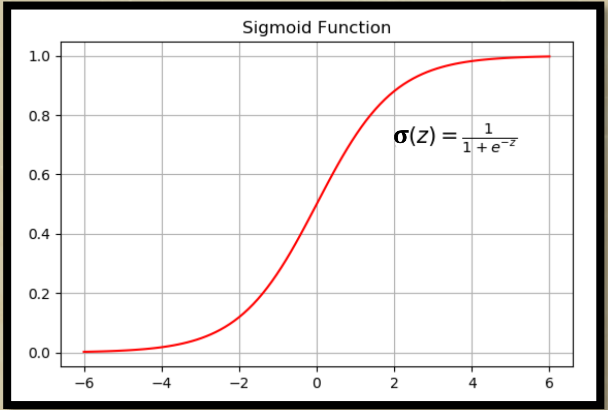
Sigmoidea como Probabilidades

Otra ventaja es que puedo interpretar la función la “probabilidad de que y=1”.

En el ejemplo de los tumores. Si para un ejemplo particular (supongamos x=14), la hipótesis me retorna 0.7, significa que hay un 70% de probabilidad de que sea un tumor maligno.

Formalmente: $\boldsymbol{\sigma}(x) = P(y=1 \mid x; w)$

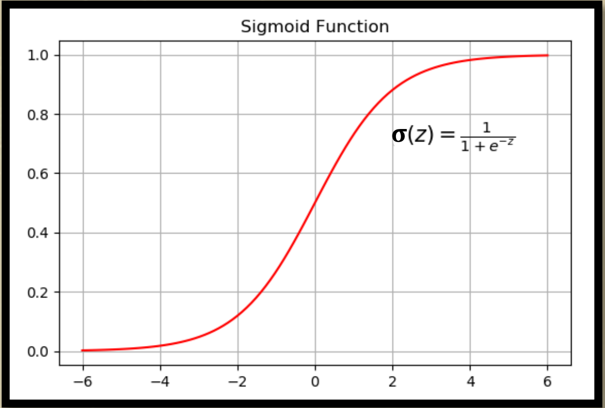
Por otro lado, $P(y=0 \mid x; w) = 1 - P(y=1 \mid x; w)$



Sigmoidea como Probabilidades

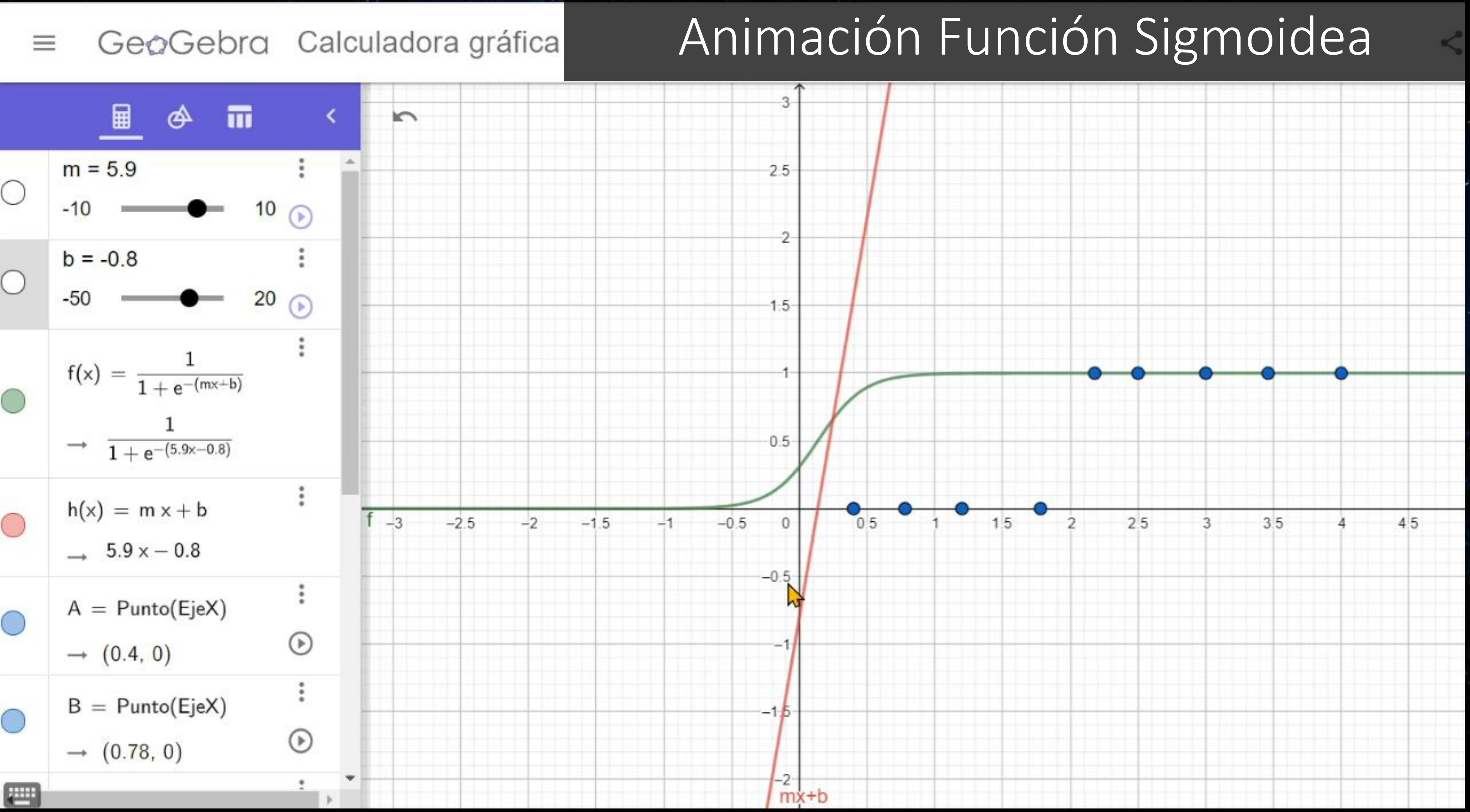
Finalmente, ponemos un umbral para clasificar

$$y = \begin{cases} 1 \quad si \;\; \boldsymbol{\sigma}(x) > 0{,}5 & \rightarrow \quad wx + b > 0 \\ 0 \quad si \;\; \boldsymbol{\sigma}(x) \leq 0{,}5 & \rightarrow \quad wx + b \leq 0 \end{cases}$$



mean radius	σ(x)	Y' (x)
17,99	0,71	1
20,57	0,88	1
19,69	0,76	1
13,54	0,42	0
9,52	0,25	0

Animación Función Sigmoidea



Frontera de decisión 1D

Supongamos, para nuestro problema de tumores, que

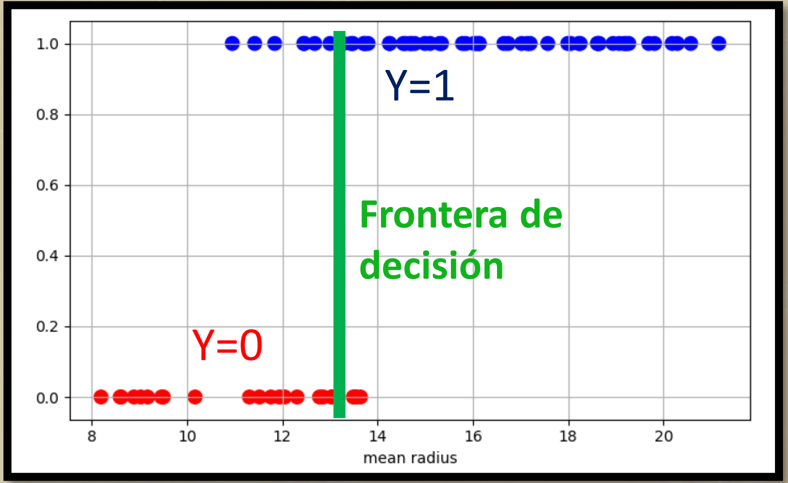
w= 0,3 y **b**= -4, umbral =0.5

Entonces, $\mathbf{y} = \mathbf{1}$ si $f(x) > 0$

si $0{,}3x - 4 > 0$

$$x > \frac{4}{0{,}3}$$

$$\mathbf{y} = \mathbf{1} \quad \text{si} \quad \mathbf{x} > \mathbf{13{,}33}$$



Frontera de Decisión x=13,33

Frontera de decisión 2D

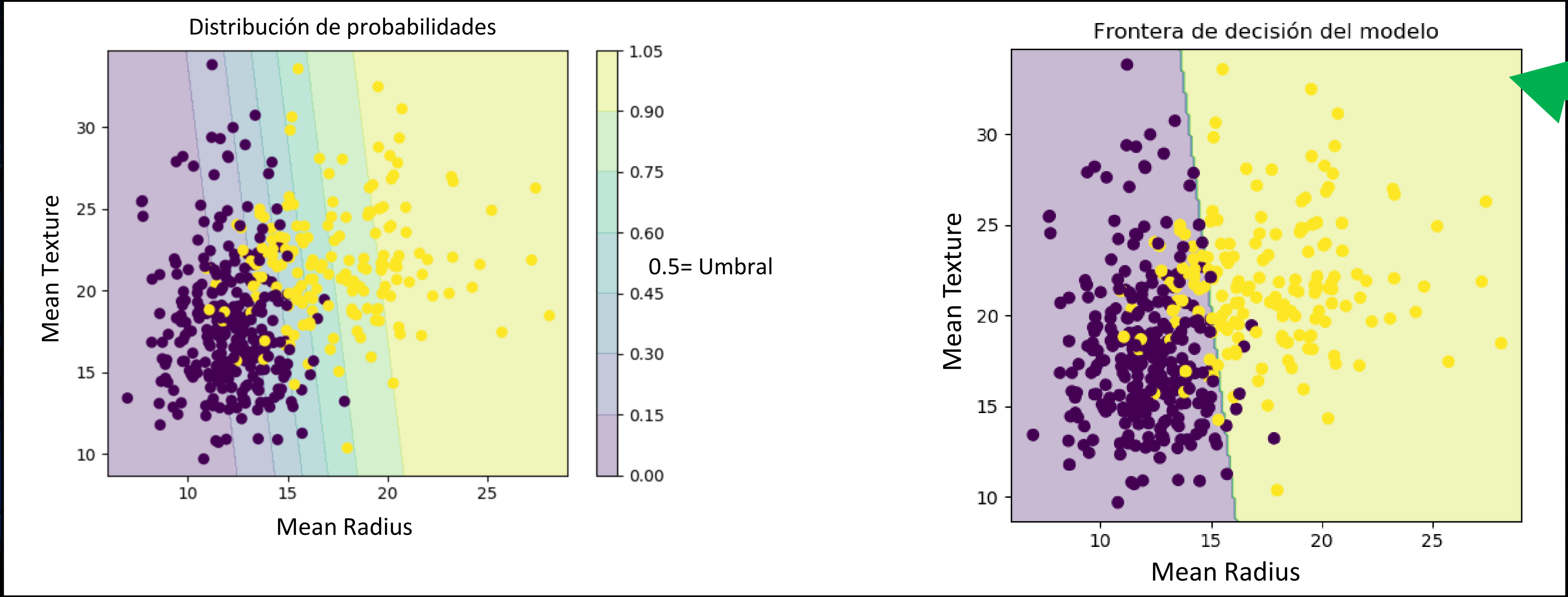
- 2 variables para cada registro
- **σ**(x1,x2)= **σ**(w1x1+w2x2+b)
- w =[0.47,0.05] , b = -8

$$y = 1 \quad si \quad si\ \boldsymbol{\sigma}(x) > 0{,}5 \quad \rightarrow \quad \boldsymbol{w1x1 + w2x2 + b > 0}$$

$$\boldsymbol{0,47x1 + 0,05x2 - 8 > 0}$$

$$\boldsymbol{0,47x1 + 0,05x2 > 8}$$

$$\boldsymbol{x2 = 160 - 9,4x1}$$

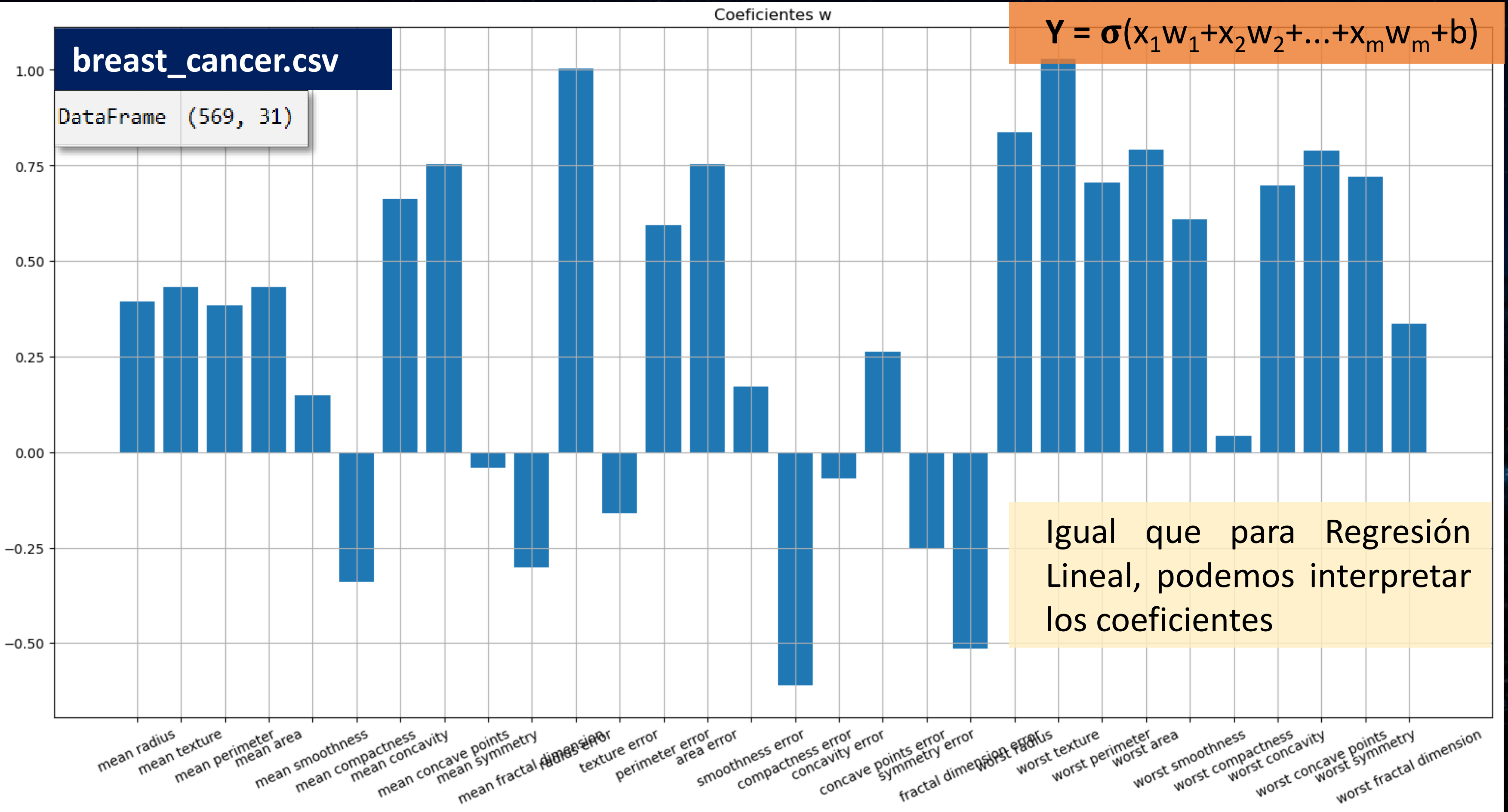


Modelo ND

breast_cancer.csv

`DataFrame (569, 31)`

$$Y = \boldsymbol{\sigma}(x_1w_1+x_2w_2+\ldots+x_mw_m+b)$$



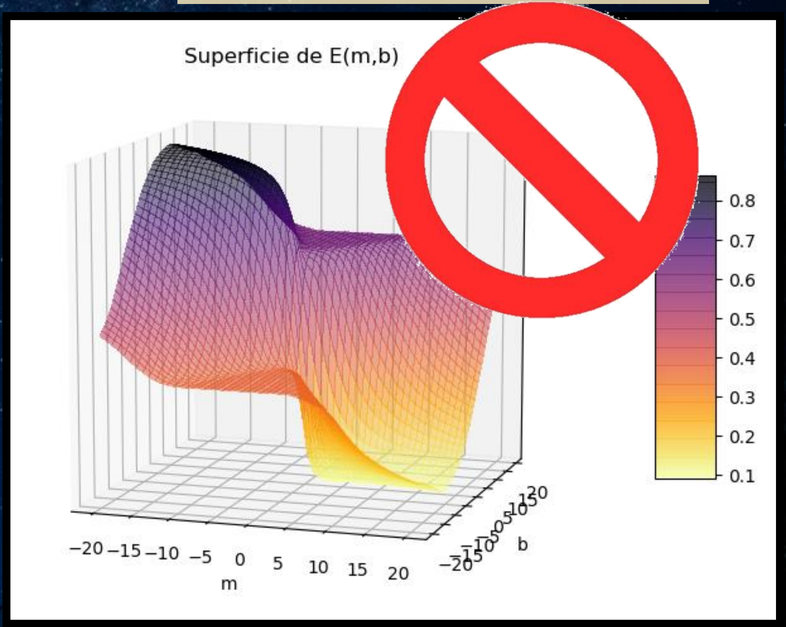
Igual que para Regresión Lineal, podemos interpretar los coeficientes

Entropía cruzada
(Error para Regresión Logística)

Error Cuadrático Medio

- $E = 1/n \, \Sigma_i^n \, Ei$
- $Ei = (\sigma(mx_i+b)-y_i)^2$
- Funciona, pero
  - ECM de σ(mx+b) no es apropiado
  - E no es convexa
- **Entropía Cruzada**
  - Convexa para $\sigma(mx_i+b)$
  - Mide distancia entre distribuciones de probabilidades.

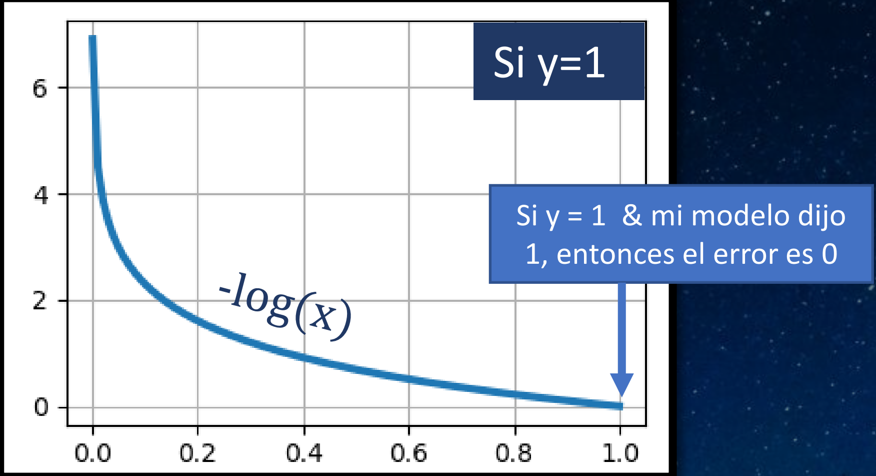
ECM de σ(mx+b)



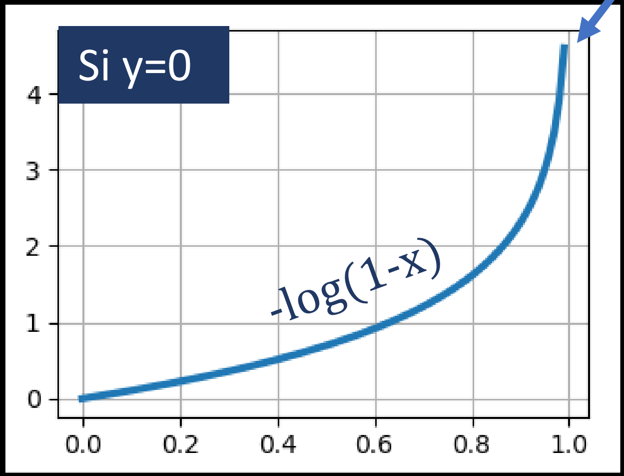
Entropía Cruzada

$$\mathbf{E} = \frac{\mathbf{1}}{\mathbf{n}} \sum_{\mathbf{i}}^{\mathbf{n}} \mathbf{E_i}$$

$$\mathbf{E_i} = \begin{cases} -\log(\boldsymbol{\sigma}(x_i)) & si\ y = 1 \\ -\log(1 - \boldsymbol{\sigma}(x_i)) & si\ y = 0 \end{cases}$$



Si y = 0 & mi modelo dijo 1, entonces el error es muy alto



Entropía Cruzada

$$\mathbf{E} = \frac{\mathbf{1}}{\mathbf{n}} \sum_{\mathbf{i}}^{\mathbf{n}} \mathbf{E_i}$$

$$\mathbf{E_i} = \begin{cases} -\log(\boldsymbol{\sigma}(x_i)) & si\ y = 1 \\ -\log(1 - \boldsymbol{\sigma}(x_i)) & si\ y = 0 \end{cases}$$

Podemos escribir la función partida en una sola línea:

$$\mathbf{E_i} = -y\ \log(\boldsymbol{\sigma}(x_i)) - \left((1 - y) \log(1 - \boldsymbol{\sigma}(x_i))\right)$$

Los coeficientes "y", "$(1 - y)$", actúan como "if-else"

Para cada $\boldsymbol{x_i}$ solo se computa el ***–log*** de la probabilidad de pertenencia a su clase.

Entropía Cruzada

Entr. Cruzada como Distancia entre distribuciones

- **f(x) = 0.3**
  - **P**(x es de clase = 1) = **0.3**
  - **P**(x es de clase = 0) = 1 - P(x es de clase = 1) = 1 - 0.3 = **0.7**
  - P(x es de clase = 1) y P(x es de clase = 0) forman una distribución de probabilidad
- **ECM**
  - Mide distancia entre puntos. Distancia euclídea al cuadrado
- **Entropía Cruzada**
  - Mide distancia entre distribuciones de probabilidad
  - Distancia Kullback-Leibler

Derivadas Entropía Cruzada

- $\frac{\delta E}{\delta b} = \frac{1}{n} \sum_{i}^{n} y_i - f(x_i)$

- $\frac{\delta E}{\delta m} = \frac{1}{n} \sum_{i}^{n} (y_i - f(x_i)) x_i$

- Son iguales que las de Regresión Lineal (*f(x)=σ(mx+b)*)
- *Mismas ecuaciones de Descenso de Gradiente*

  b = b - α δE(m,b)/δb
  m = m - α δE(m,b)/δm

Métrica Accuracy

- Valores de entropía cruzada: **0 a +∞**

  Difíciles de interpretar

Y	Y'	Acierto
1	1	1
1	0	0
0	0	1
0	1	0
0	1	0

Accuracy = % de ejemplos que clasificó correctamente.

Métrica para Clasificación.

Ejemplo: 2 / 5 -> **Acc = 0,4** (o 40%)

Resumen

Regresión Logística

- Permite modelar probabilidades
- Modelo $Y= f(x_1,x_2,\ldots,x_m) = \boldsymbol{\sigma}(x_1w_1+x_2w_2+\ldots+x_mw_m+b)$,

  $\boldsymbol{\sigma}(x) = 1/(1+e^{-x})$

  Interpretación de la salida del modelo como una probabilidad.
- Umbral -> Convierte probabilidades en clases.

  -> Frontera de decisión.
- Entropía cruzada -> Métrica de error. Mide distancia entre distrib. probabilidades.