Regresión logística.

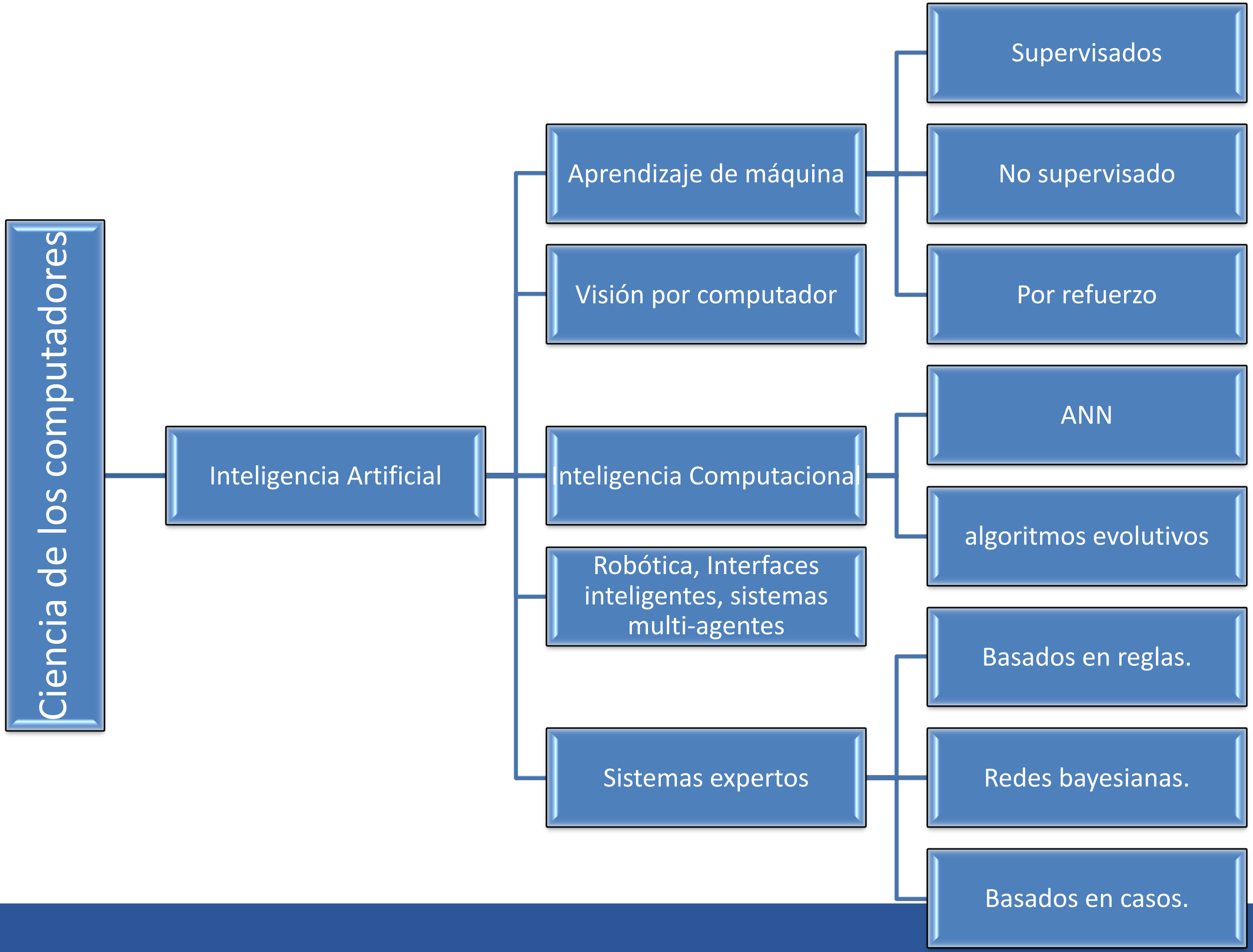
Pontificia Universidad Javeriana
Francisco Carlos Calderon Ph.D
2020

Objetivos

Definir un problema de clasificación.

Usar la regresión logística para solucionar un problema de clasificación

Reconocer que a pesar de que se llama regresión logística, no es un modelo de regresión es un modelo de clasificación!


Ciencia de los computadores
Inteligencia Artificial
Aprendizaje de máquina
Visión por computador
Inteligencia Computacional
Robótica, Interfaces inteligentes, sistemas multi-agentes
Sistemas expertos
Supervisados
No supervisado
Por refuerzo
ANN
algoritmos evolutivos
Basados en reglas.
Redes bayesianas.
Basados en casos.

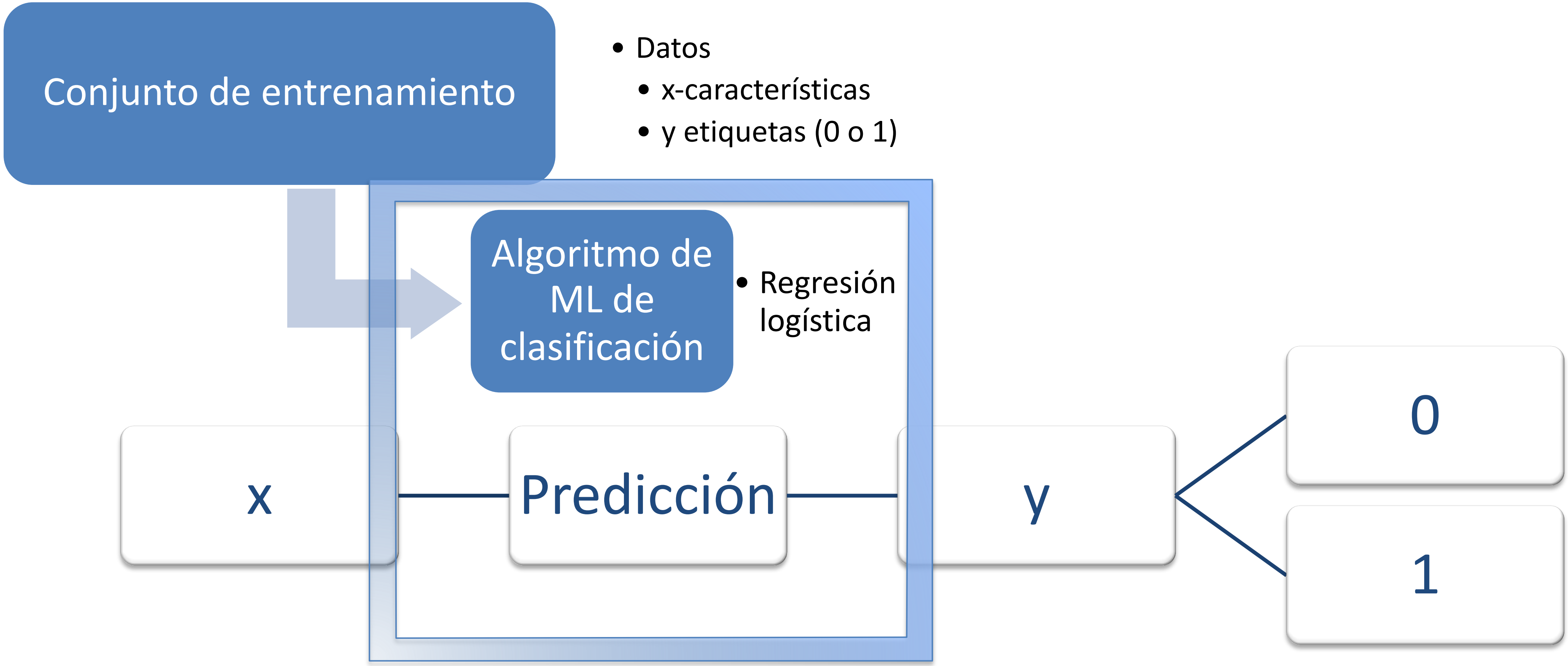
Aprendizaje de máquina, clasificación general

- Supervisados
  - Crean un modelo matemático que busca explicar unas “**etiquetas**” de entrada/salida a partir de un conjunto de “**características**” de entrada.
  - Se pueden dividir principalmente en:
    - Clasificación
    - Regresión
  - Existen otros sub-métodos como:
    - Aprendizaje activo.
    - “Similarity learning”
    - Recommender systems

No Supervisados

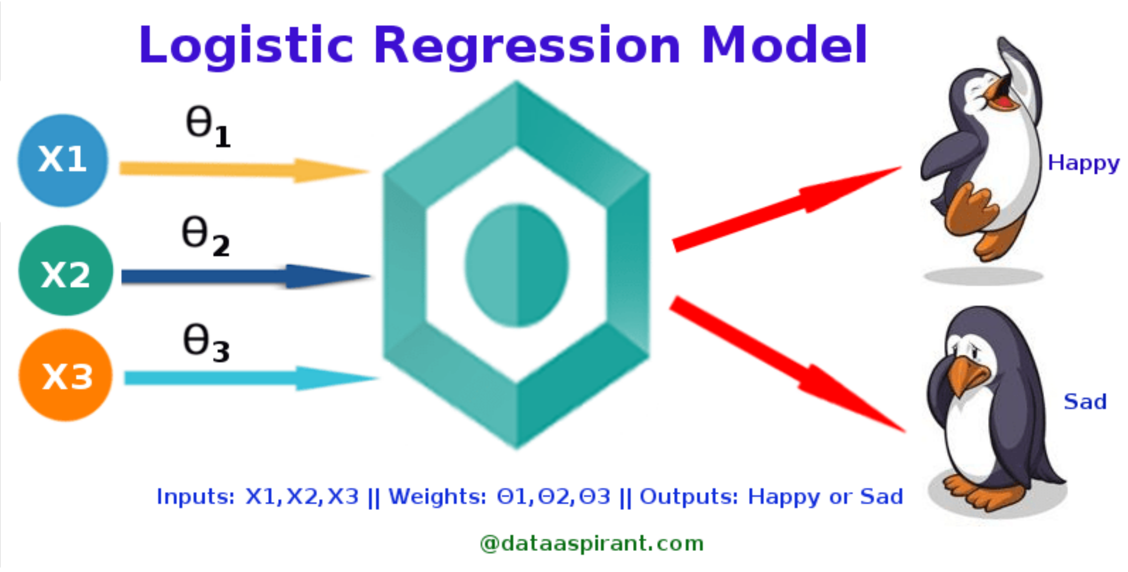
- Crean un modelo que busca explicar las **características** de entrada sin contar con etiquetas.
- Se pueden dividir en
  - Agrupamiento. “clustering”
  - Estimación de densidad (pdf).
  - Reducción dimensional

Idea de Clasificación



Clasificación

Es un proceso estadístico para estimar estimar una etiqueta de las características de entrada



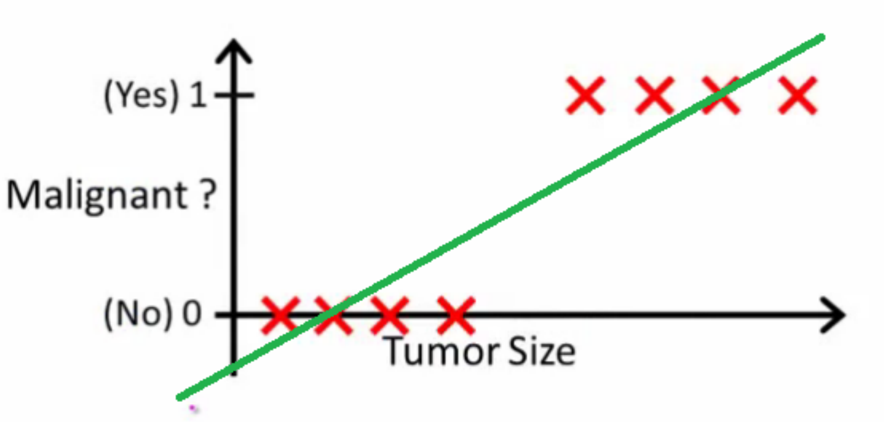
Fuente: https://dataaspirant.com/how-logistic-regression-model-works/

Podemos hacer clasificación con regresión

Sabemos llegar a una función hipótesis que explique nuestros datos. De salida “y”

Si nuestra salida solo incluye dos estados

$$y \in \{0,1\}$$



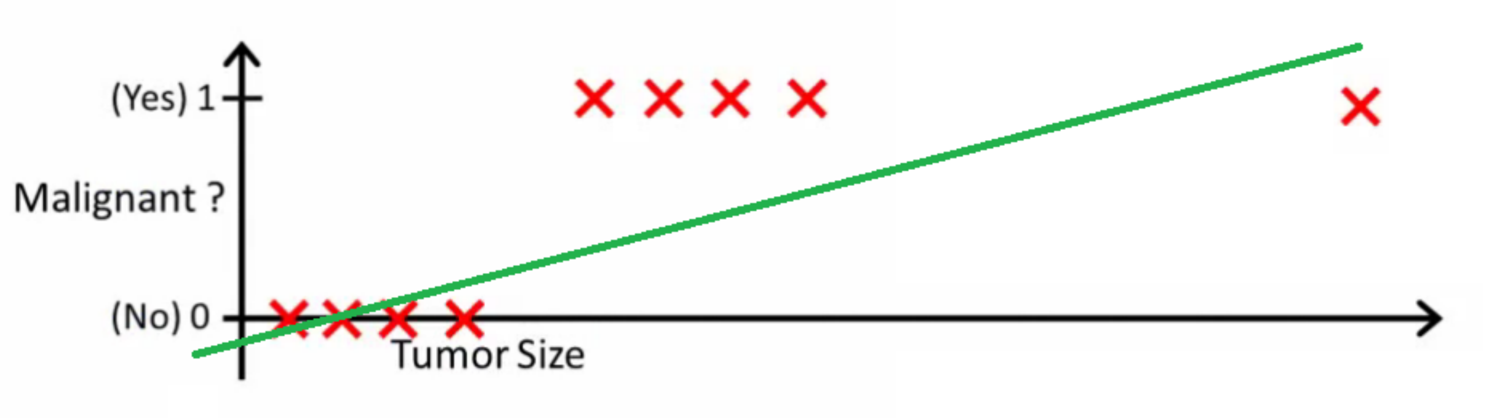


Imagen de https://www.coursera.org/learn/machine-learning

Podemos hacer clasificación con regresión

Nuestra salida solo incluye dos estados

$$y \in \{0,1\}$$

Pero en nuestra regresión hay valores menores y mayores a cero

Solución:

$$0 \leq h(x) \leq 1$$

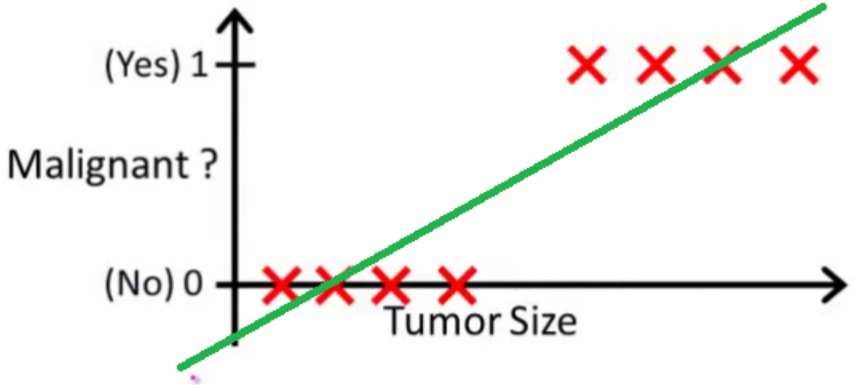


Imagen de https://www.coursera.org/learn/machine-learning

Selección de una hipótesis adecuada

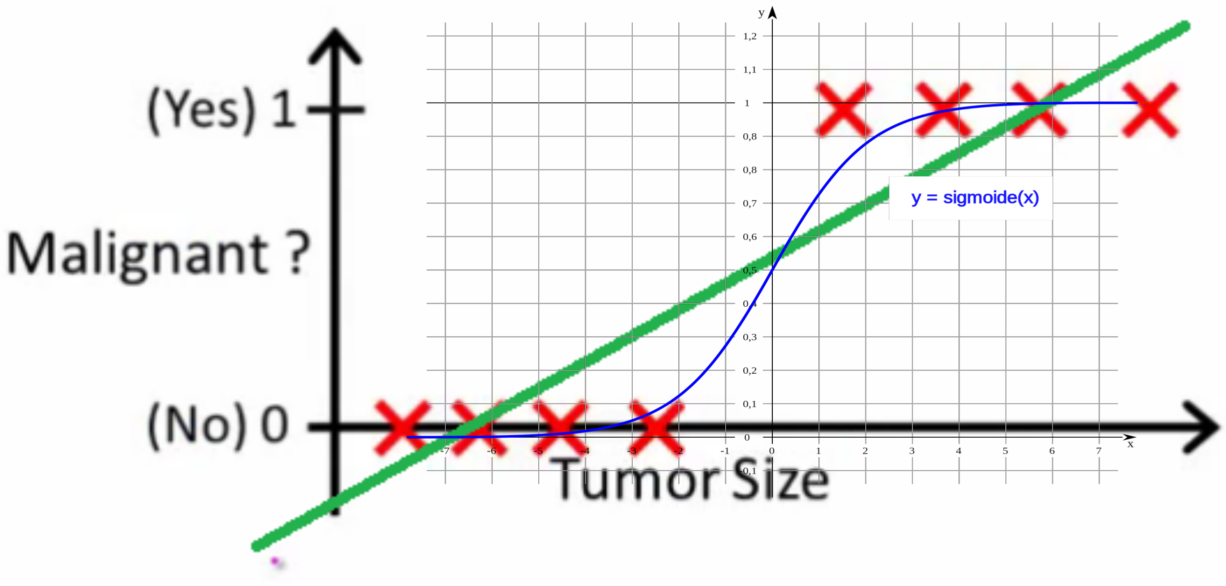
Función sigmoide o logística:

$$y = \frac{1}{1+e^{-x}}$$

$\theta^T x$

$$\lim_{x\to-\infty} \frac{1}{1+e^{-x}} = 0$$

$$\lim_{x\to+\infty} \frac{1}{1+e^{-x}} = 1.$$



Fuente https://es.wikipedia.org/wiki/Funci%C3%B3n_sigmoide

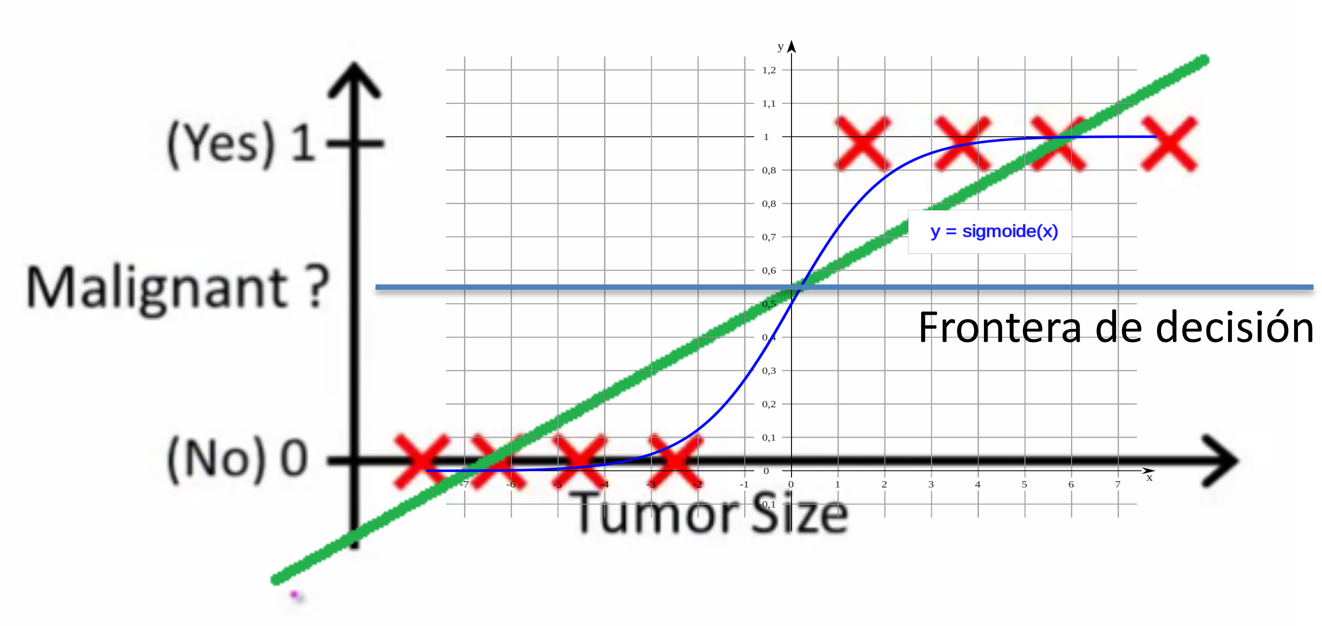
Interpretación del modelo de clasificación

h(x) será ahora la probabilidad estimada que y=1 a una entrada x

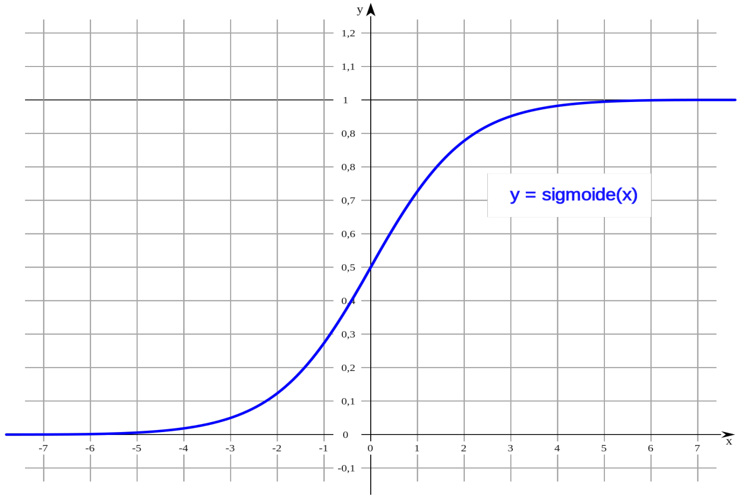
$$h_\theta(x) = P(y = 1|x;\theta)$$

OJO: $h_\theta(x)$ no es la estimación de y

$$P(y = 0|x;\theta) + P(y = 1|x;\theta) = 1$$
$$P(y = 0|x;\theta) - 1 = P(y = 1|x;\theta)$$



Notación:



Definimos la hipótesis:

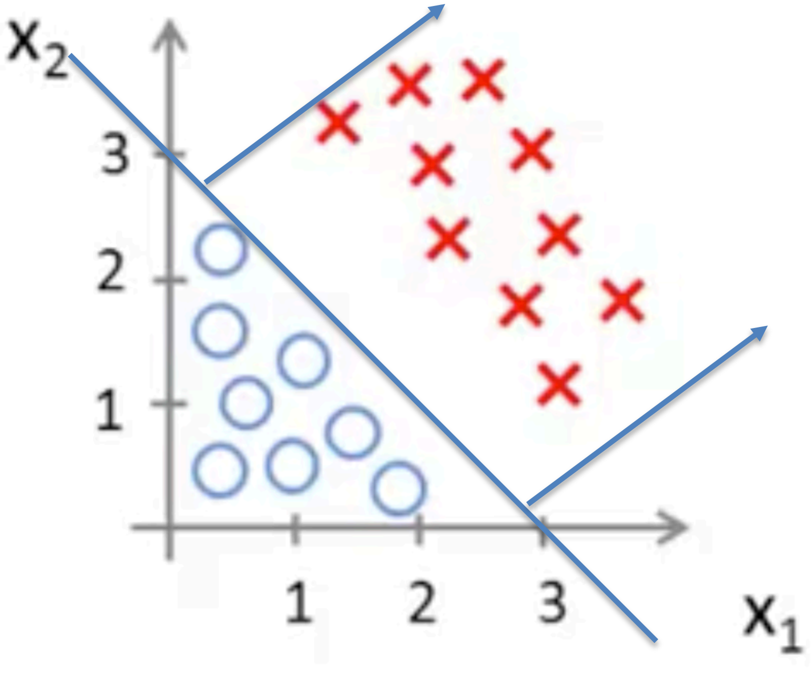
$$h_\theta(x) = g(\theta^T x) = P(y = 1|x;\theta)$$

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

Predecimos "$\ddot{y} = 1$" si $h_\theta(x) \geq 0.5$ o $\theta^T x \geq 0$

Predecimos "$\ddot{y} = 0$" si $h_\theta(x) < 0.5$ o $\theta^T x < 0$

Frontera de decisión



Ejemplo multivariado

$$h_\theta(x) = g(\theta^T x) = P(y = 1 | x; \theta)$$

$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$$

Predecimos "$\ddot{y} = 1$" si $h_\theta(\ddot{x}) \geq 0.5$ o $\theta^T \ddot{x} \geq 0$

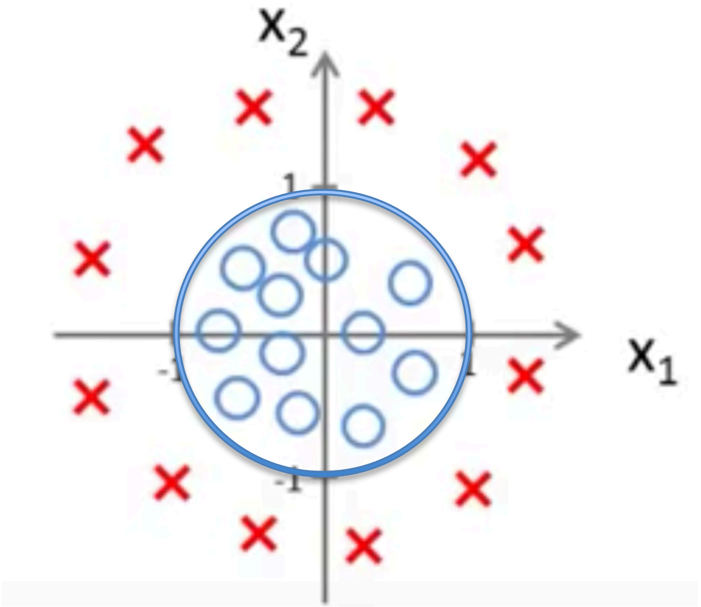
si $\theta_0 = -3, \theta_1 = 1, \theta_2 = 1$: $x_1 + x_2 \geq 3$

Predecimos "$\ddot{y} = 0$" si $h_\theta(x) < 0.5$ o $\theta^T x < 0$

Ejemplo de Andrew NG

Frontera de decisión

Ejemplo multivariado polinómico



$$h_\theta(x) = g(\theta^T x) = P(y = 1 | x; \theta)$$

$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2)$$

Predecimos "$\ddot{y} = 1$" si $h_\theta(\ddot{x}) \geq 0.5$ o $\theta^T \ddot{x} \geq 0$

si $\theta_0 = -1, \theta_1 = 0, \theta_2 = 0 , \theta_3 = 1 , \theta_4 = 0$

: $x_1^2 + x_2^2 \geq 1$

Predecimos "$\ddot{y} = 0$" si $h_\theta(x) < 0.5$ o $\theta^T x < 0$

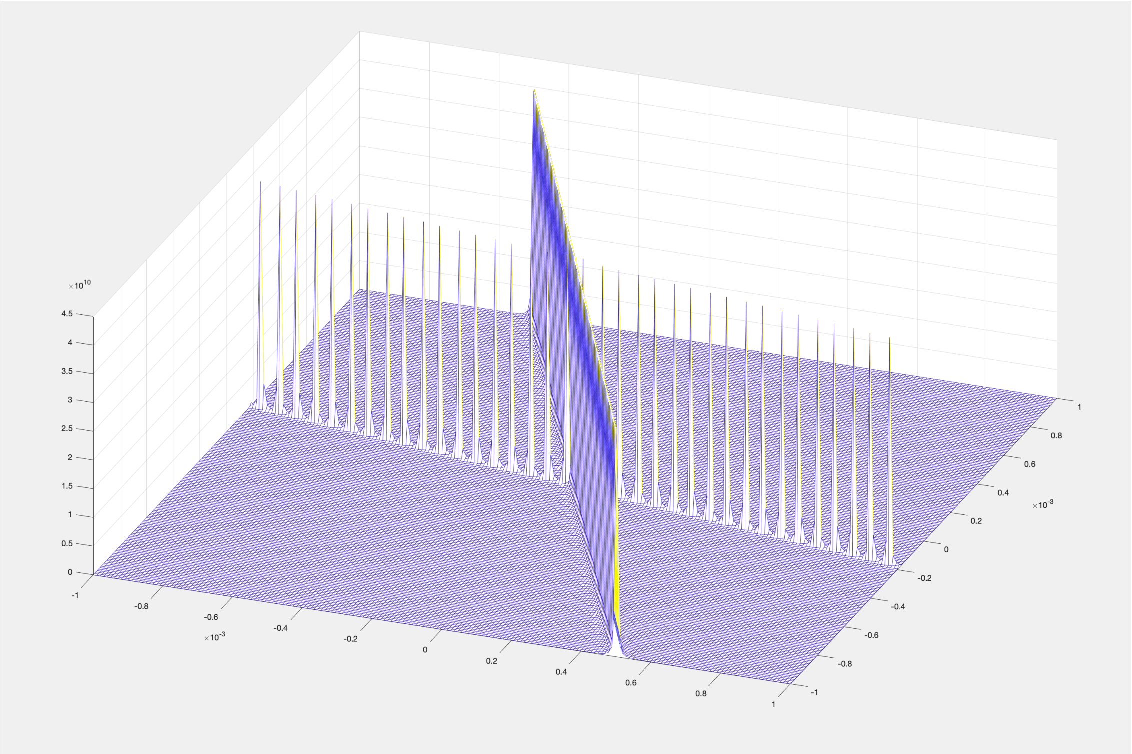
Ejemplo de Andrew NG

Función de costo

Problemas ☹

Si minimizamos nuestra función de costo SSD esta no va a ser convexa

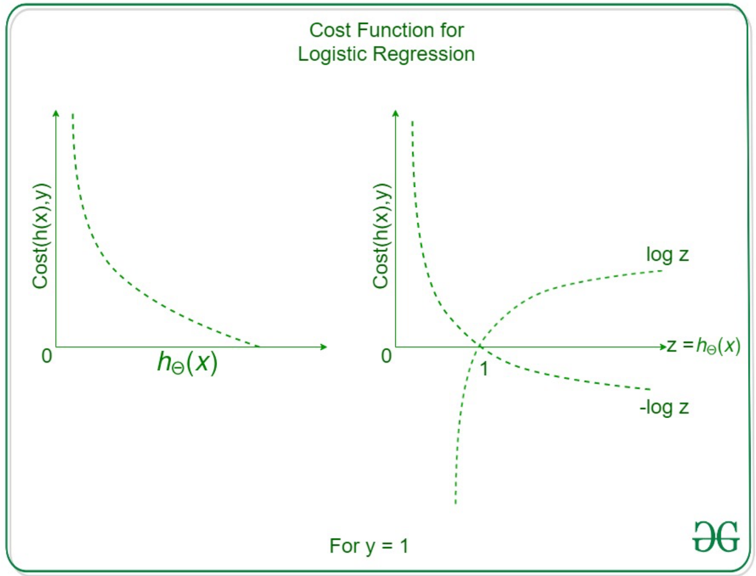
$$\min_{\theta_0,\theta_1,\ldots,\theta_m} \frac{1}{2n}\sum_{i=1}^{n}\left(h_{\boldsymbol{\theta}}\left(\boldsymbol{x}^{(i)}\right)-y^{(i)}\right)^2$$

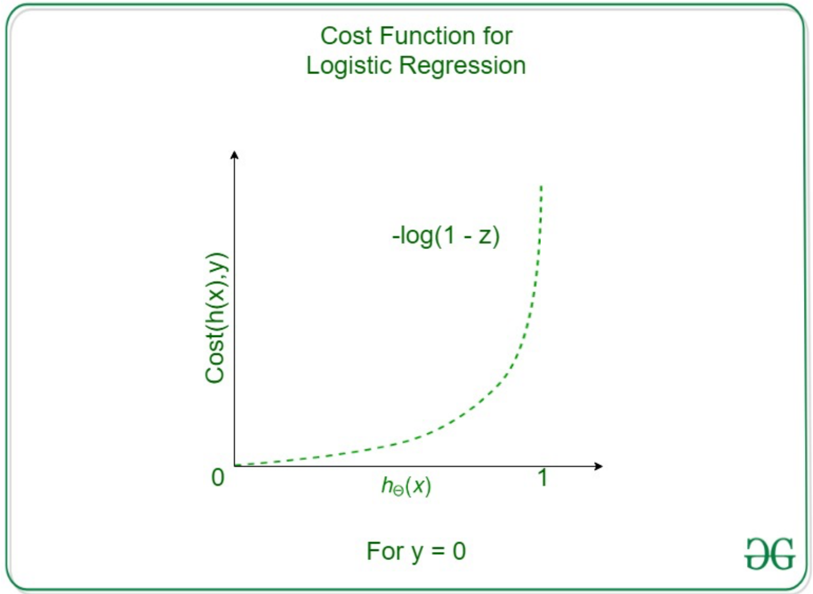


Fuente propia función de costo ejemplo en wikipedia
https://en.wikipedia.org/wiki/Logistic_regression

Y entonces cual costo uso?

$$cost\left(h_{\boldsymbol{\theta}}\left(\boldsymbol{x}^{(i)}\right), \boldsymbol{y}^{(i)}\right) = \begin{cases} -\ln\left(h_{\boldsymbol{\theta}}\left(\boldsymbol{x}^{(i)}\right)\right) & \boldsymbol{y}^{(i)} = 1 \\ -\ln\left(1 - h_{\boldsymbol{\theta}}\left(\boldsymbol{x}^{(i)}\right)\right) & \boldsymbol{y}^{(i)} = 0 \end{cases}$$





$$J(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} cost\left(h_{\boldsymbol{\theta}}\left(\boldsymbol{x}^{(i)}\right), y^{(i)}\right)$$

$$\min_{\theta_0, \theta_1, \ldots, \theta_m} \frac{1}{n}\sum_{i=1}^{n} cost\left(h_{\boldsymbol{\theta}}\left(\boldsymbol{x}^{(i)}\right), y^{(i)}\right)$$

Tomado de: https://www.geeksforgeeks.org/ml-cost-function-in-logistic-regression/

Re-escribiendo en una forma más trabajable…

Simplificando

$$cost(h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)}), \boldsymbol{y}^{(i)}) = -\boldsymbol{y}^{(i)}\ln(h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)}) - (\mathbf{1} - \boldsymbol{y}^{(i)})\ln(1 - h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)})$$

$$J(\boldsymbol{\theta}) = -\frac{1}{n}\left[\sum_{i=1}^{n} \boldsymbol{y}^{(i)} ln(h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)})) + (\mathbf{1} - \boldsymbol{y}^{(i)})\ln(1 - h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)})\right]$$

Esta función si es convexa!☺.

Si minimizamos usando descenso del gradiente:

$$\theta_j^{(k+1)} = \theta_j^{(k)} - \alpha\frac{1}{n}\sum_{i=1}^{n}(h_{\boldsymbol{\theta}^k}(x^{(i)}) - y^{(i)})\ x_j^{(i)}$$

$$h_\theta(x) = g(\theta^T x)$$

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

Algunos autores omiten el 1/n ☹
https://ml-cheatsheet.readthedocs.io/en/latest/logistic_regression.html#id13
https://math.stackexchange.com/questions/78575/derivative-of-sigmoid-function-sigma-x-frac11e-x

Notas finales de la regresión logística

1. Existen mejores estrategias para hallar ese mínimo.
    1. No es necesario seleccionar un alpha.
    2. Pueden decrecer más rápido que el descenso del gradiente.
2. Algunos de los métodos comunes son:
    - Gradiente conjugado “Newton-CG”
    - BFGS
    - L-BFGS
    - SAG
    - SAGA
3. Y si tengo múltiples clases? No solo 0 y 1?

https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

Tarea

- Implementar (adecuar) los dos métodos descritos en:
- https://ml-cheatsheet.readthedocs.io/en/latest/logistic_regression.html#id13
- Con los datos en el csv en teams.
- Puede descargar el código en:
- https://github.com/bfortuner/ml-glossary
- correr el ejemplo multiclase al final de:
- https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html
- Comentar línea por línea los que usen sklearn
- Encontrar la derivada de J para un theta de la Regresión Logística