Regresión Logística

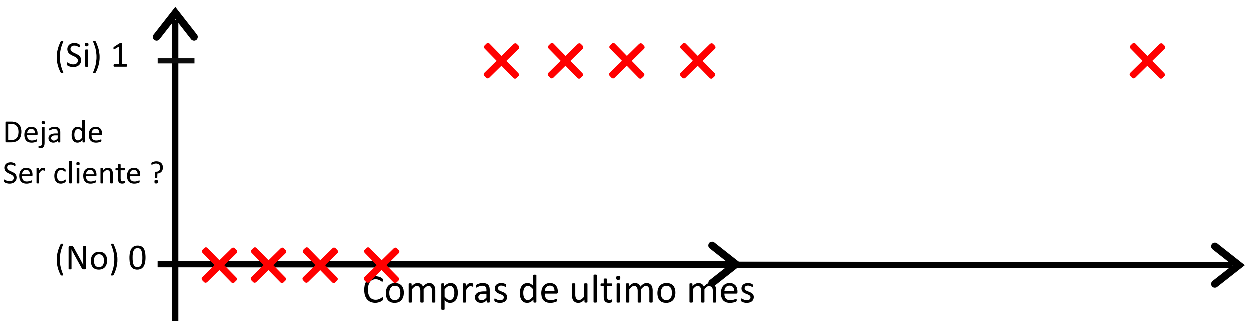
Clasificación

Clasificación

Email: *¿Spam / Not Spam?*
Transacciones online: *Fraudulentas ¿Si / No?*
Tumor: *¿Maligno / Benigno ?*
Cliente: *Confiable de crédito ¿Si / No?*

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

0: "Clase negativa" (e.g. tumor benigno)
1: "Clase positiva" (e.g. tumor maligno)



¿REGRESOR LINEAL?

Threshold de clasificador $h_\theta(x)$ en 0.5:

Si $h_\theta(x) \geq 0.5$, predice "y = 1"

Si $h_\theta(x) < 0.5$, predice "y = 0"

Classificación: y = 0 o 1

$h_\theta(x)$ Puede ser > 1 o < 0

Regresión logistica: $0 \leq h_\theta(x) \leq 1$

Regresión logistica
Representación de hipotesis

Modelo de regresión logistica

$$0 \le h_\theta(x) \le 1$$

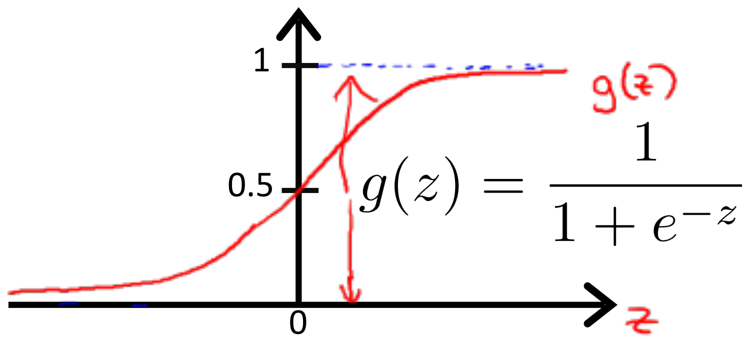
$$h_\theta(x) = g(\theta^T x)$$

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

$\theta^T x$

$$\boxed{h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}}$$

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



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

Función sigmoide
(función logistica)

Interpretación de salida de hipotesis

$h_\theta(x)$ = probabilidad estimada de y = 1 dada entrada x

Ejemplo: Si $x = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} 1 \\ \text{tumorSize} \end{bmatrix}$

$$h_\theta(x) = 0.7$$

Esto indica que hay 70% de chance de tumor maligno.

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

"probabilidad que y = 1, dado x, parametrizado por θ"

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

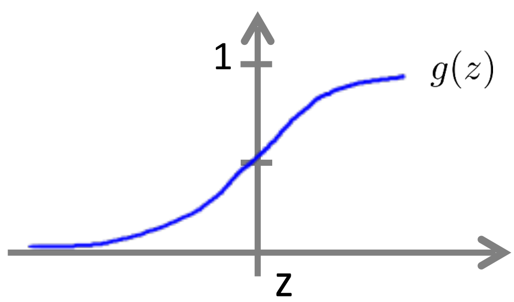
Una de las siguientes afirmaciones sobre la regresión logística (vista hasta el momento) es verdadera. Indique cuál es.

- ○ La función de hipótesis es no lineal.
- ○ La salida corresponde a un número real mayor a 0.
- ○ Es adecuada para el problema de predicción de precio de limones para el día de mañana.
- ○ Ninguna de las anteriores.

Regresión Logistica

Bordes de decisión

Regresión logistica



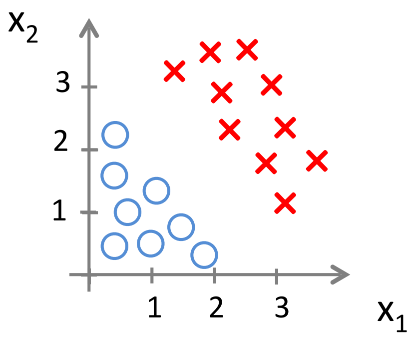
$$h_\theta(x) = g(\theta^T x)$$

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

Suponga predecir "$y = 1$" if $h_\theta(x) \geq 0.5$

predecir "$y = 0$" if $h_\theta(x) < 0.5$

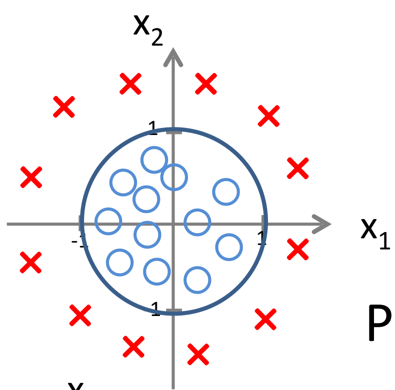
Borde de decisión (lineal)



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

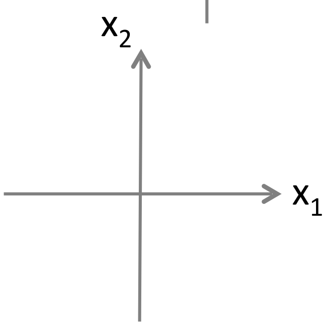
Predecir "$y = 1$" si $-3 + x_1 + x_2 \geq 0$

Borde de decisión no-lineal



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

Predecir "$y = 1$" if $-1 + x_1^2 + x_2^2 \geq 0$



$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_1^2 x_2 + \theta_5 x_1^2 x_2^2 + \theta_6 x_1^3 x_2 + \dots)$$

Regresión logistica

Función de costo

Training set: $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \cdots, (x^{(m)}, y^{(m)})\}$

m ejemplos $x \in \begin{bmatrix} x_0 \\ x_1 \\ \dots \\ x_n \end{bmatrix}$ $x_0 = 1, y \in \{0, 1\}$

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}}$$

¿Como escoger parametros θ?

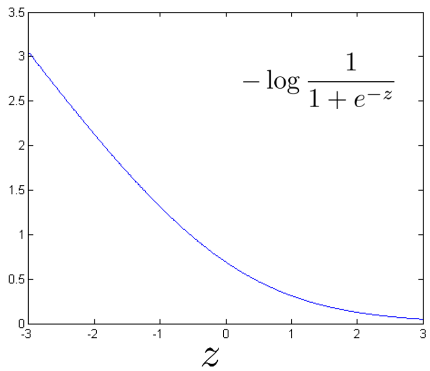
Función de costo de regresión logistica

Cost=0; si y=1, h(x)=1
Pero si h(x) -> 0 luego Cost -> inf.

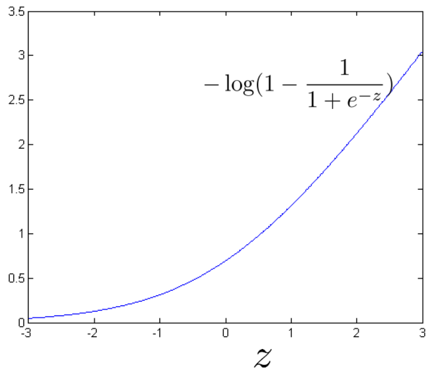
Esto captura intuición que si h(x)=0 y dado que y=1, el costo se hace muy grande, lo que es coherente.

$$\mathrm{Cost}(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & \text{if } y = 1 \\ -\log(1 - h_\theta(x)) & \text{if } y = 0 \end{cases}$$

Si $y = 1$ (osea $\theta^T x \gg 0$):



Si $y = 0$ (osea $\theta^T x \ll 0$):



Regresión Logistica

Función de costo simplificada
y descenso de gradiente

Función de costo de regresión logistica

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} \mathrm{Cost}(h_\theta(x^{(i)}), y^{(i)})$$

$$\mathrm{Cost}(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & \text{if } y = 1 \\ -\log(1 - h_\theta(x)) & \text{if } y = 0 \end{cases}$$

Nota: y =0 o 1; no otro valor!.

Función de costo de regresión logistica

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} \mathrm{Cost}(h_\theta(x^{(i)}), y^{(i)})$$

$$= -\frac{1}{m}[\sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)}))]$$

Para ajustar parametro : θ

$$\min_\theta J(\theta)$$

Para hacer predicción dado nuevo data x:

Salida $h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$

Descenso de gradiente

$$J(\theta) = -\frac{1}{m}[\sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1-y^{(i)}) \log (1-h_\theta(x^{(i)}))]$$

Se requiere $\min_\theta J(\theta)$:

Repetir $\{$

$$\theta_j := \theta_j - \alpha \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

$\frac{\partial}{\partial \theta_j} J(\theta)$

Operación para hallar mínimo!

(simultaneamente actualizar todo θ_j)

$\}$

Algoritmo es casi idéntico a regresión lineal!

BONUS: SVM

Regresor logístico

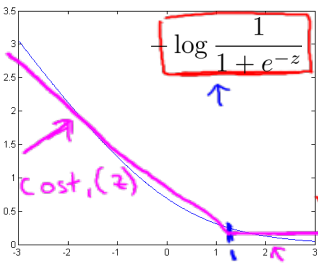
$$J(\theta) = -\frac{1}{m}[\sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1-y^{(i)}) \log (1-h_\theta(x^{(i)}))] + \frac{1}{2}\sum_{i=1}^{n} \theta_j^2$$

Máquina de soporte vectorial (SVM)

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} \Big[y^{(i)} cost_1(\theta^T x^{(i)}) + (1-y^{(i)}) cost_0(\theta^T x^{(i)})\Big]_{-\log(1-\frac{1}{1+e^{-z}})} + \frac{1}{2}\sum_{i=1}^{n} \theta_j^2$$

$cost_1(\theta^T x^{(i)}) = max(0, 1 - \theta^T x)$

$cost_0(\theta^T x^{(i)}) = max(0, 1 + \theta^T x)$



z

$y = 1$

$\theta^T x \gg 0$

Tiene óptimo global!!: usando programación cuadrática.

z