



Universidad de San Carlos de Guatemala

Facultad de Ingeniería

Escuela de Ingeniería en Ciencias y Sistemas

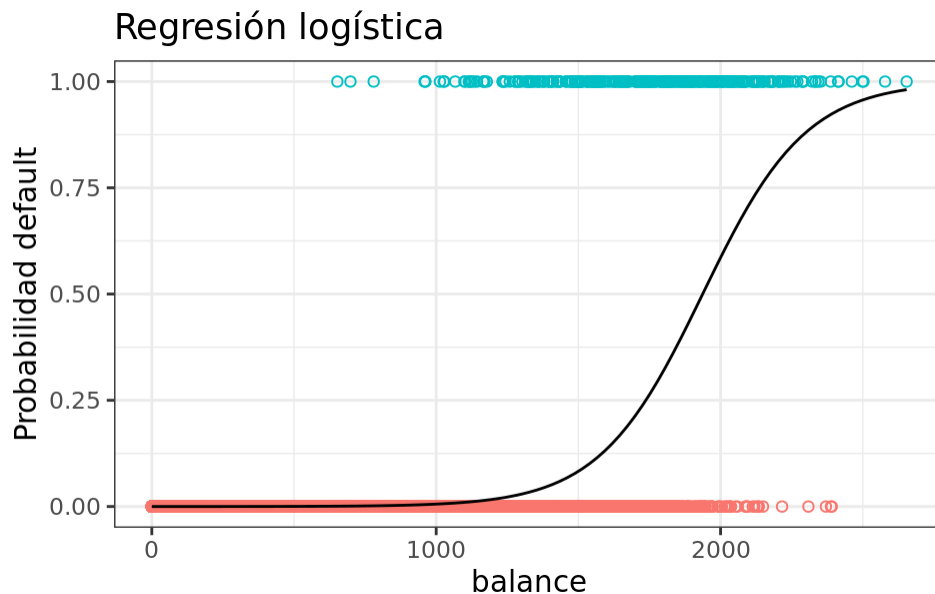
Inteligencia Artificial

Ing. Luis Espino

## Regresión Logística Binaria Simple y MultiClase

### Regresión Logística Binaria Simple

La Regresión Logística Simple, desarrollada por David Cox en 1958, es un método de regresión que permite estimar la probabilidad de una variable cualitativa binaria en función de una variable cuantitativa. Una de las principales aplicaciones de la regresión logística es la de clasificación binaria, en el que las observaciones se clasifican en un grupo u otro dependiendo del valor que tome la variable empleada como predictor. Por ejemplo, clasificar a un individuo desconocido como hombre o mujer en función del tamaño de la mandíbula.



### ¿Por qué regresión logística y no lineal?

Si una variable cualitativa con dos niveles se codifica como 1 y 0, matemáticamente es posible ajustar un modelo de regresión lineal por mínimos cuadrados  $\beta_0 + \beta_1 x$ . El problema de esta aproximación es que, al tratarse de una recta, para valores extremos del predictor, se obtienen valores de Y menores que 0 o mayores que 1, lo que entra en contradicción con el hecho de que las probabilidades siempre están dentro del rango [0,1].

Para evitar estos problemas, la regresión logística transforma el valor devuelto por la regresión lineal ( $\beta_0 + \beta_1 X$ ) empleando una función cuyo resultado está siempre comprendido entre 0 y 1. Existen varias funciones que cumplen esta descripción, una de las más utilizadas es la función logística (también conocida como función sigmoide):

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

Donde  $z$  es la función de la recta de regresión Lineal  $\beta_0 + \beta_1 X$

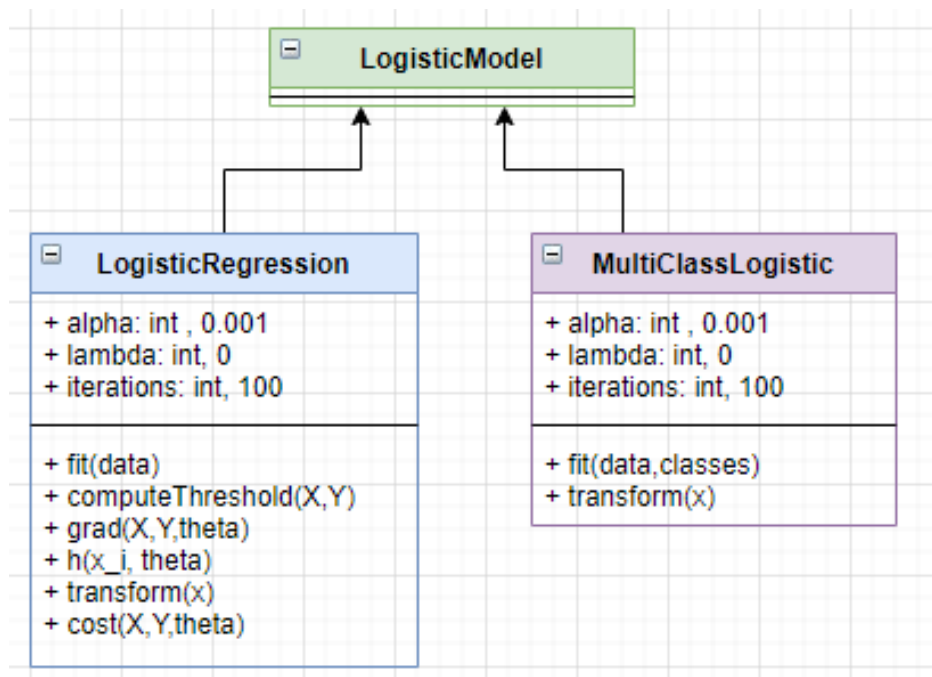
### Regresión Logística Multiclase

La regresión logística múltiple es una extensión de la regresión logística simple. Se basa en los mismos principios que la regresión logística simple (explicados anteriormente) pero ampliando el número de predictores. Los predictores pueden ser tanto continuos como categóricos. El valor de la probabilidad de  $Y$  se puede obtener con la inversa del logaritmo natural:

$$p(Y) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_i}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_i}}$$

### IMPLEMENTACION

#### Class Diagram



## Logistic Regression Binary

El algoritmo se implemento utilizando diferentes métodos los cuales los mas importantes son:  $\text{fit}(\text{data})$ ,  $\text{h}(\text{x}_i, \text{theta})$  y  $\text{transform}(\text{x})$ .

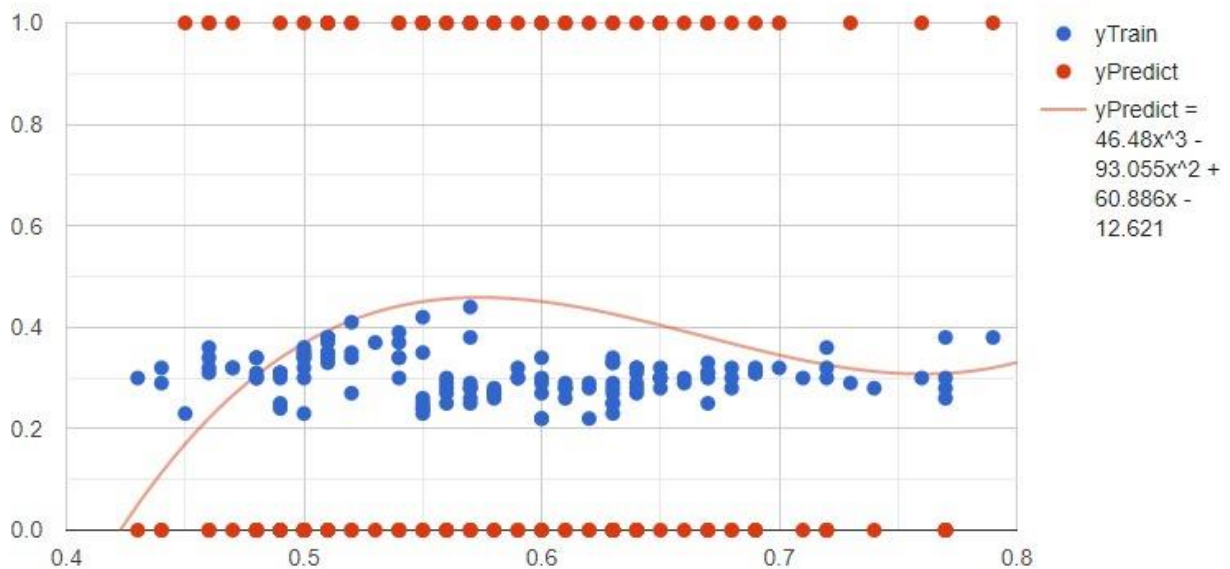
```
var logistic = new LogisticRegression()
var fit = logistic.fit(matriz)
var yPredict = []
for(var i=0; i < matriz.length; ++i){
    var p = logistic.transform(matriz[i]);
    var predicted = p >= logistic.threshold ? 1 : 0;
    console.log("logistic binary logistic testing: actual: " + matriz[i][2] + " predicted: " + predicted);
    yPredict.push(predicted)
}
```

Se utiliza una muestra de entrada para la prueba de la regresión logística simple:

```
var xTrain = [6.8,4.3,5.4,5.8,6.2,5.6,5.2,6.3,5.6,5,7.2,6.4,5.5,6.7,6.7,4.9,7.7,6.7,6,6.3,4.8,6,7.7,4.8,4.8  
var yTrain = [2.8,3,3.4,2.7,2.2,2.5,2.7,2.8,2.7,3.2,3.6,3.2,3.5,3.1,3.3,2.5,3.8,3.1,3.4,2.9,3,2.7,2.6,3.4,3  
var zActualValue = [0,0,0,0,0,0,0,1,0,0,1,0,0,1,1,1,0,0,1,0,0,1,0,0,0,0,0,0,0,0,0,0,1,0,0,0,1,0,1,0,1,0,  
var matrix= []
```

Como podemos observar en la imagen anterior, se envia un parámetro anterior en binario donde se puede representar de diferentes formas como Gano o perdió, Si o No , etc.

**Y la grafica que nos devuelve dicha probabilidad es la siguiente**



## Logistic Regression Multiclass

El algoritmo se implemento utilizando diferentes métodos los cuales los mas importantes son: `fit(data,classes)` y `transform(X)`.

```
var fit = logistic.fit(matriz)
var yPredict = []

for(var i=0; i < matriz.length; ++i){
  var predicted = logistic.transform(matriz[i]);
  console.log("linear svm binary classifier testing: actual: " + matriz[i][2] + " predicted: " + predicted);
  yPredict.push(predicted)
}
```

Se utiliza una muestra de entrada para la prueba de la regresión logística multiclase:

```
<script type="text/javascript">
var xTrain = [6.7,5,5,5.1,6.7,6.1,6.8,4.4,5,5.8,4.8,6.5,5.7,7.2,4.9,5.6,6.9,5.4,7.2,5.7,6.2,6.1,5.9,4.8,5.8,5.7,
var yTrain = [3,2.3,3.5,2.5,3.3,2.8,2.8,3.2,3.5,2.7,3.4,3,3.8,3,3.1,2.5,3.1,3.4,3.6,2.6,2.9,3,3.2,3,2.7,2.8,2.4,
var zActualValue = ["Lluvia","Nublado","Sol","Nublado","Lluvia","Nublado","Nublado","Sol","Sol","Lluvia","Sol","
var matriz = []
```

A diferencia de la funcion de regression logística simple, esta se inserta un valor no binario si no que múltiples acciones o clases, en este caso estamos determinando si es probable que Lluvia, que haya Sol o que este Nublado.

Y la grafica que nos devuelve dicha probabilidad es la siguiente

