



Regresión Logística Fundamentos



Regresión Logística

- Concepto y problema que resuelve
-
- Relación target (y) con features (x)
- Funcionamiento interno
- Obtención de Parámetros
- Métricas
- Aplicación a multiclase
- Regularización

Regresión logística

Se trata de una generalización de la regresión lineal al caso de la clasificación. **OJO! Pese a llamarse "regresión logística" es un método aplicado a la clasificación.**

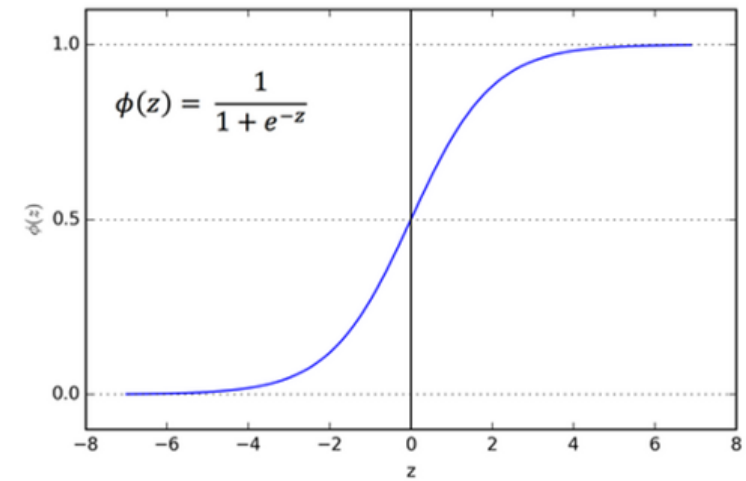


Fig: Sigmoid Function

If 'Z' goes to infinity, Y(predicted) will become 1 and if 'Z' goes to negative infinity, Y(predicted) will become 0.

Función sigmoide

Regresión logística

Se trata de una generalización de la regresión lineal al caso de la clasificación. **OJO! Pese a llamarse "regresión logística" es un método aplicado a la clasificación.**

OBJETIVO: Transformar el resultado de un valor continuo no acotado $(-\infty, +\infty)$ (regresión lineal) en una probabilidad (acotada $(0, 1)$).

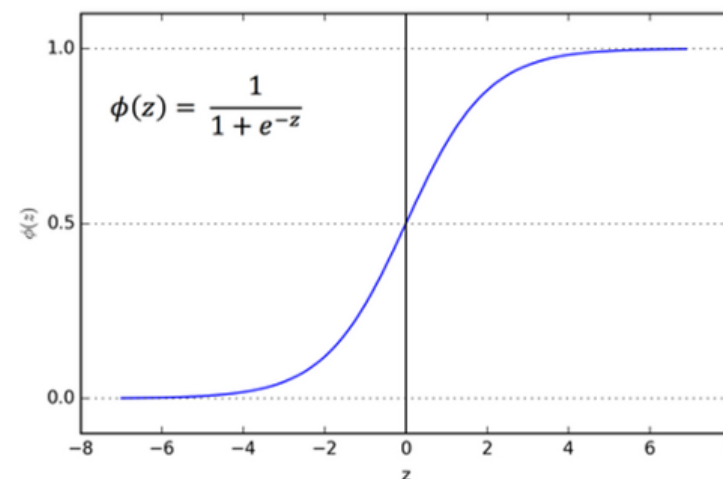


Fig: Sigmoid Function

If 'Z' goes to infinity, Y(predicted) will become 1 and if 'Z' goes to negative infinity, Y(predicted) will become 0.

Función sigmoide

Regresión logística

Se trata de una generalización de la regresión lineal al caso de la clasificación. **OJO! Pese a llamarse "regresión logística" es un método aplicado a la clasificación.**

OBJETIVO: Transformar el resultado de un valor continuo no acotado $(-\infty, +\infty)$ (regresión lineal) en una probabilidad (acotada $(0, 1)$).

La probabilidad obtenida se compara con un umbral y se decide la clase estimada por el modelo.

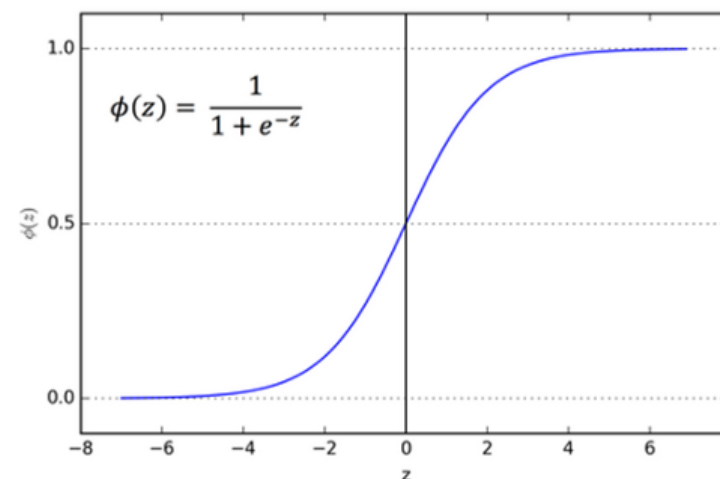
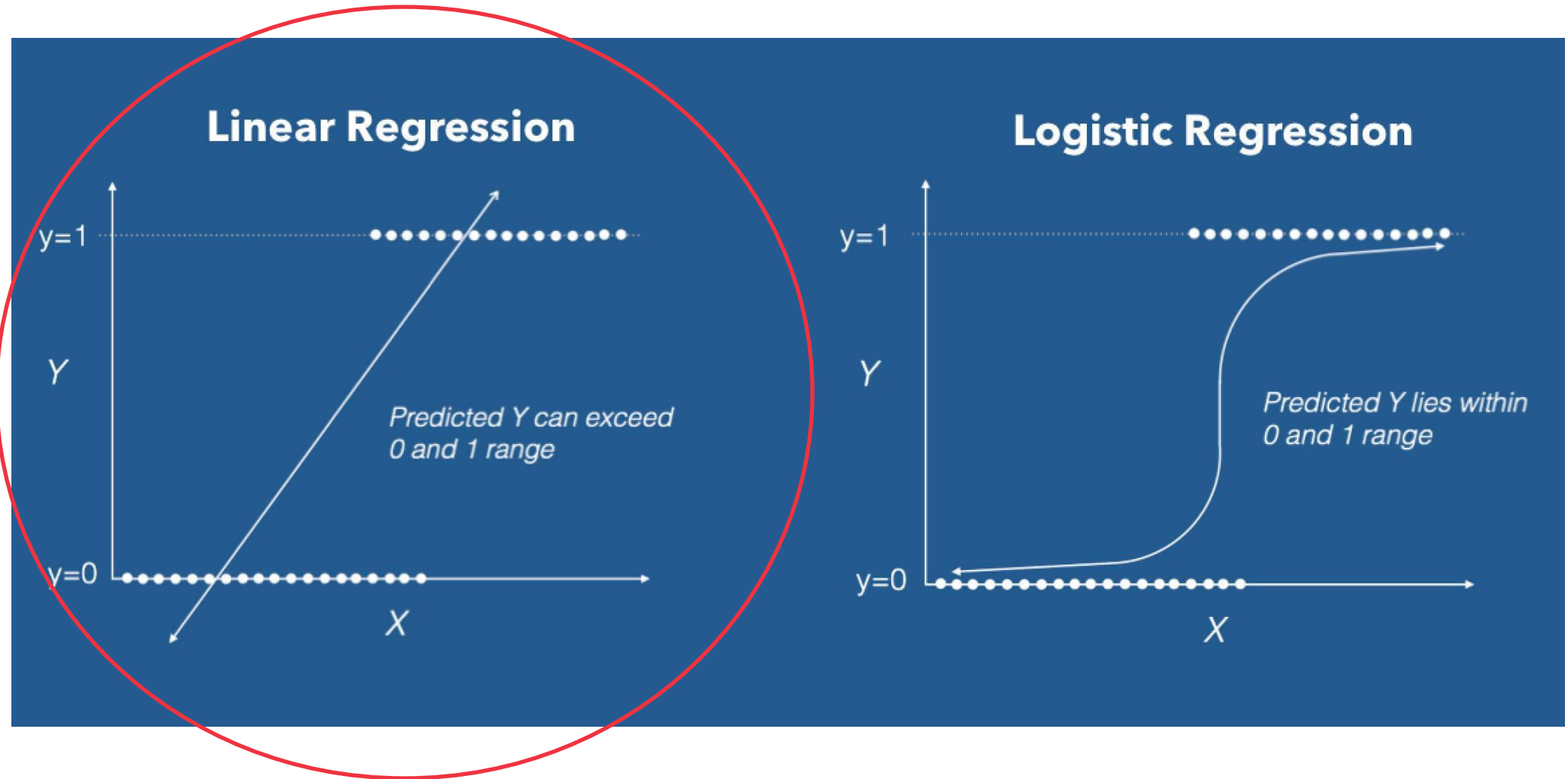


Fig: Sigmoid Function

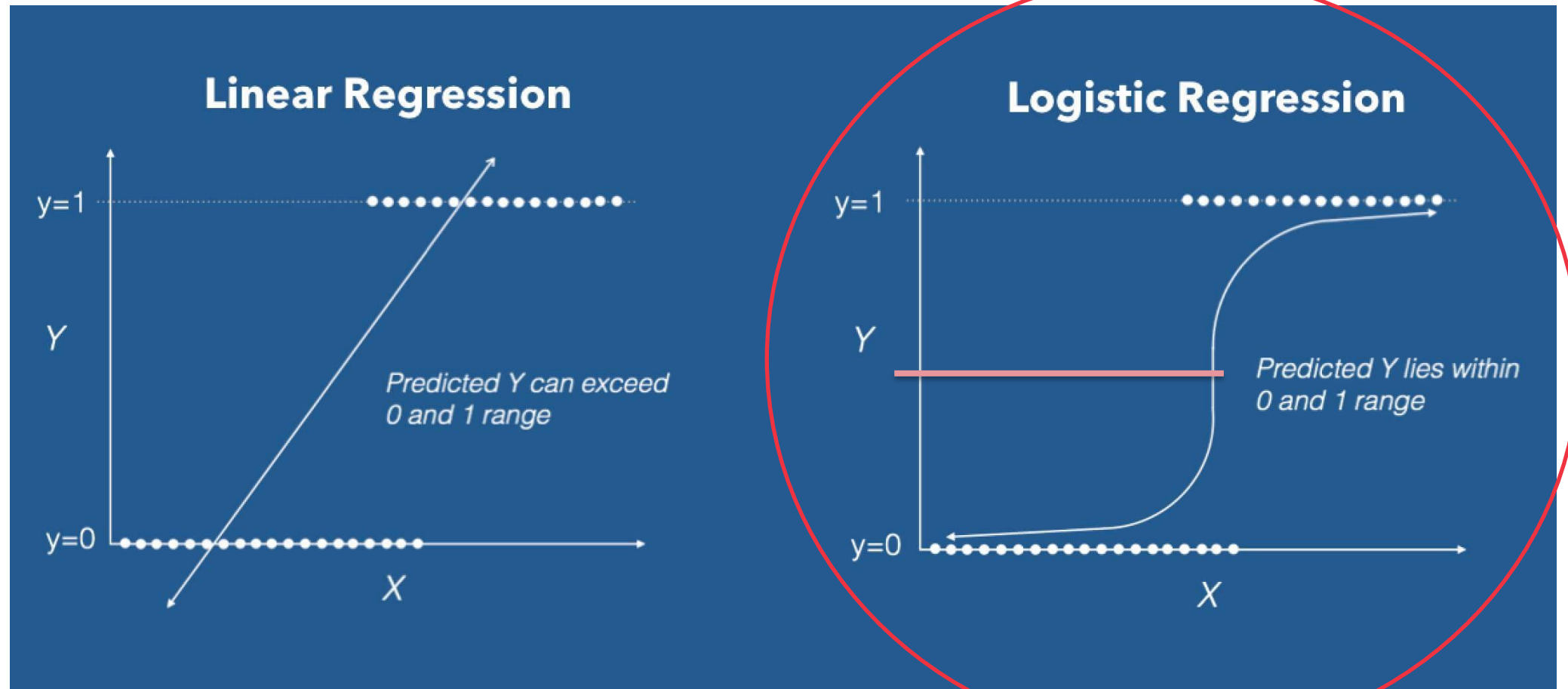
If 'Z' goes to infinity, Y(predicted) will become 1 and if 'Z' goes to negative infinity, Y(predicted) will become 0.

Función sigmoide

Regresión logística vs Regresión lineal



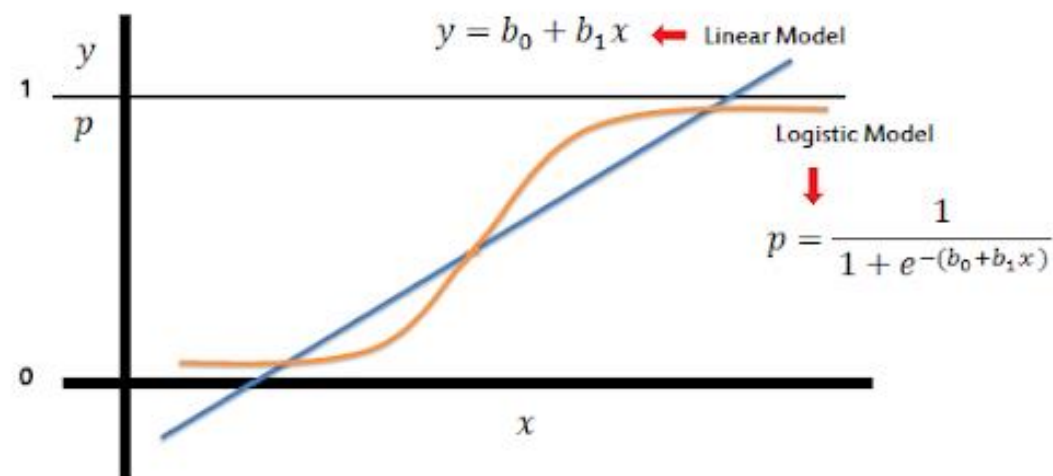
Regresión logística vs Regresión lineal



Pasos de la Regresión Logística

1. Establecimiento de la formula de regresión lineal

$$\hat{y} = a x + b$$



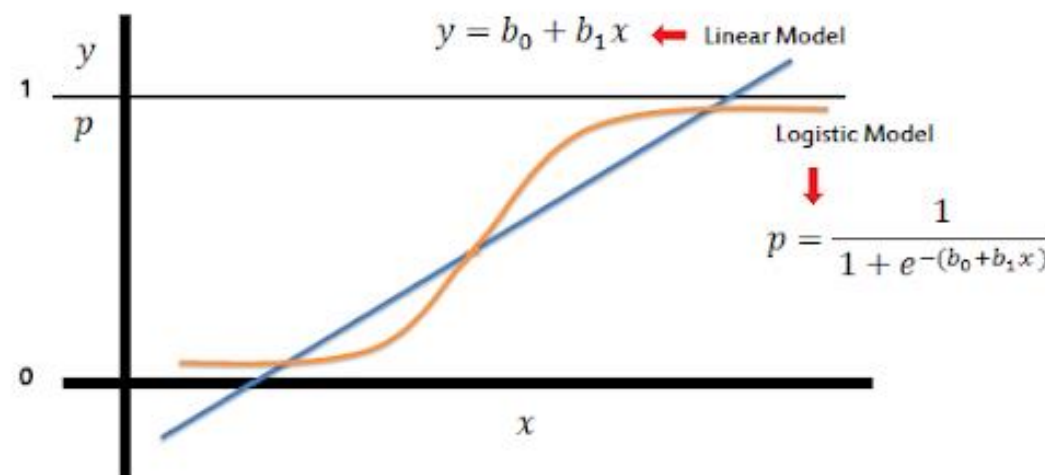
Pasos de la Regresión Logística

1. Establecimiento de la formula de regresión lineal

$$\hat{y} = a x + b$$

2. Conversión a probabilidad con la **función sigmoide** o **función logística**. Esta función tiene un valor de entrada no limitado (todos los números reales) y su output pertenece al interval $(0, 1)$. Los extremos del interval nunca se alcanzan. Compruébalo si quieres ;).

$$\hat{p} = \frac{1}{1 + e^{-(ax+b)}}$$



Pasos de la Regresión Logística

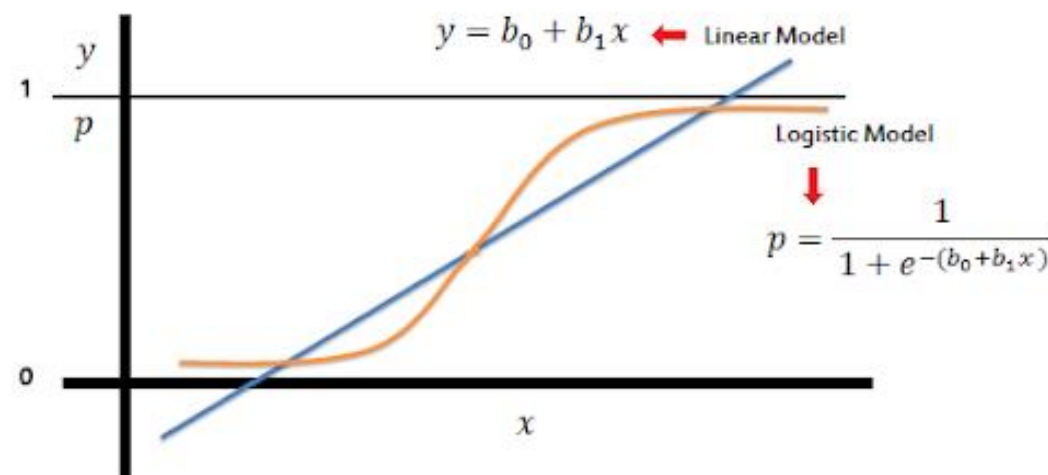
1. Establecimiento de la formula de regresión lineal

$$\hat{y} = a x + b$$

2. Conversión a probabilidad con la **función sigmoide** o **función logística**. Esta función tiene un valor de entrada no limitado (todos los números reales) y su output pertenece al intervalo $(0, 1)$. Los extremos del intervalo nunca se alcanzan. Compruébalo si quieres ;).

$$\hat{p} = \frac{1}{1 + e^{-(ax+b)}}$$

3. Se fija un umbral y si la probabilidad es mayor una clase y si no la otra.



Pasos de la Regresión Logística

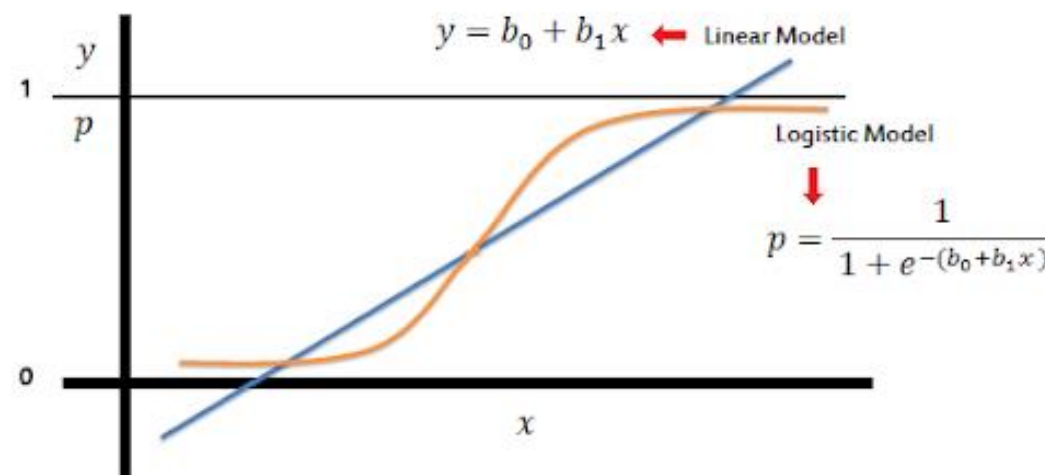
1. Establecimiento de la formula de regresión lineal

$$\hat{y} = a x + b$$

2. Conversión a probabilidad con la **función sigmoide** o **función logística**. Esta función tiene un valor de entrada no limitado (todos los números reales) y su output pertenece al intervalo $(0, 1)$. Los extremos del intervalo nunca se alcanzan. Compruébalo si quieres ;).

$$\hat{p} = \frac{1}{1 + e^{-(ax+b)}}$$

3. Se fija un umbral y si la probabilidad es mayor una clase y si no la otra.



Como hacíamos con la regresión lineal, la regresión logística trata de ajustar los valores para "a" y "b".

Obtención de los parámetros

$$E(a, b) = -\frac{1}{m} \sum_{i=1}^m \left(y_i \log(\hat{P}_i) + (1 - y_i) \log(1 - \hat{P}_i) \right)$$

1. Tiene un único mínimo. Se aplica Gradiente Descendente para obtener los parámetros que lo alcanzan.
2. Los términos con:
 - $y_i = 0$ y $\hat{P}_i \rightarrow 1$
 - $y_i = 1$ y $\hat{P}_i \rightarrow 0$

son muy grandes, mientras que aquellos con:

- $y_i = 1$ y $\hat{P}_i \rightarrow 1$
- $y_i = 0$ y $\hat{P}_i \rightarrow 0$

se aproximan a 0.

Métricas

Accuracy

Precision

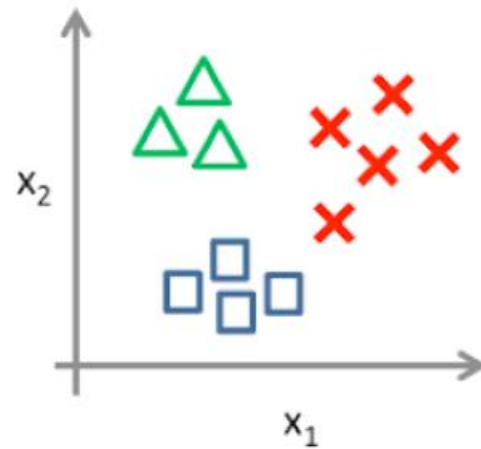
Recall

F1-score

Curva RoC -> AuC/AuRoC

Multiclase/Multicategórico

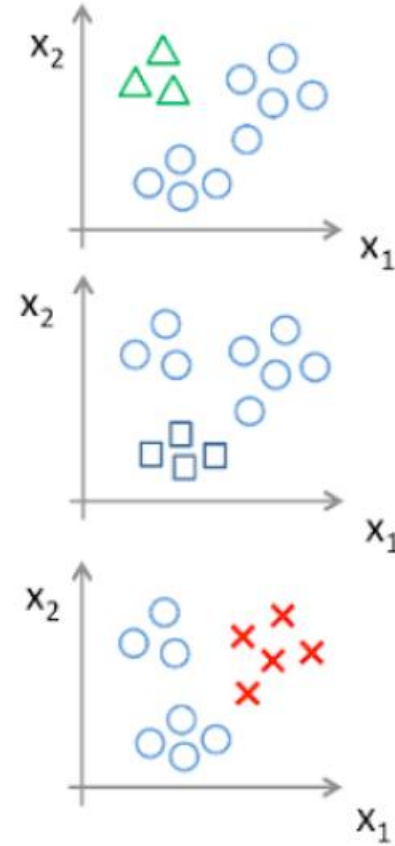
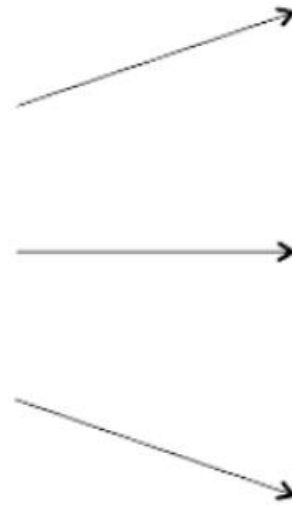
One-vs-all (one-vs-rest):



Class 1: **Green**

Class 2: **Blue**

Class 3: **Red**



Regularización

- *L2 (Ridge)*
- *L1 (Lasso)*
- *Combinación (Elastic Net)*

