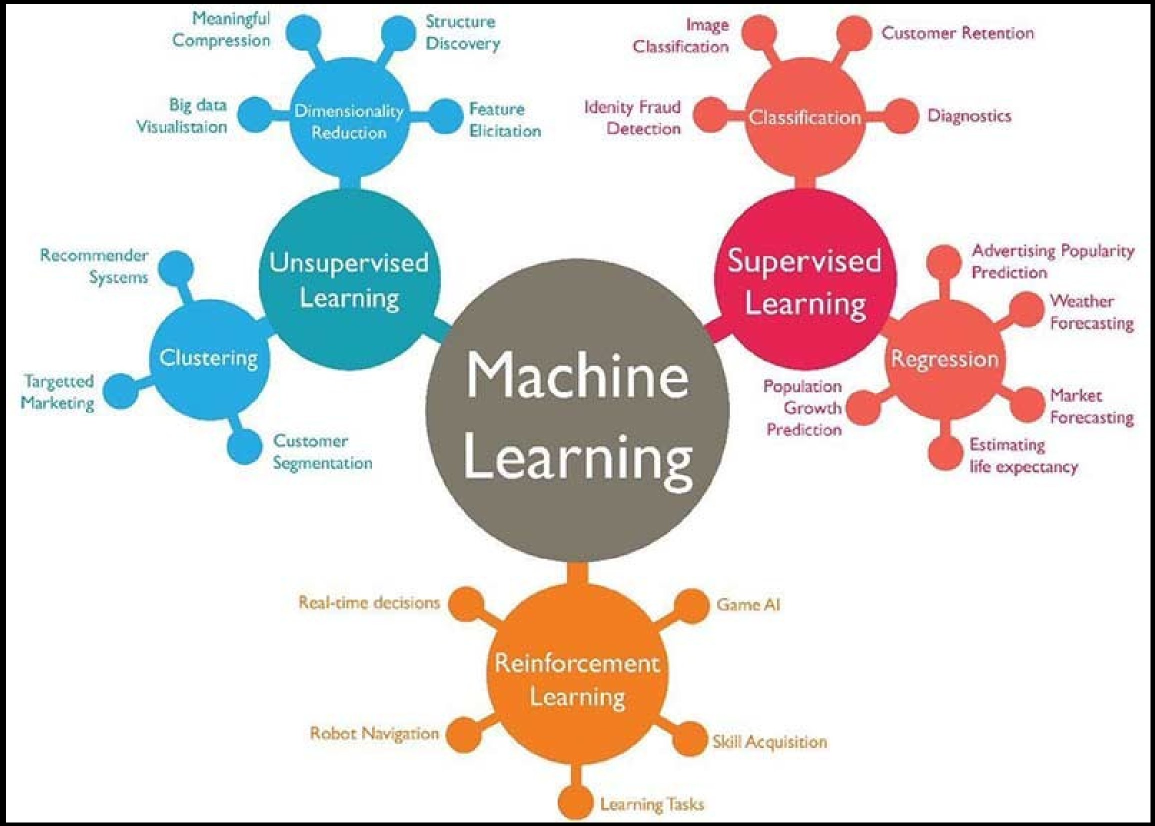
Machine Learning – Regresión logística

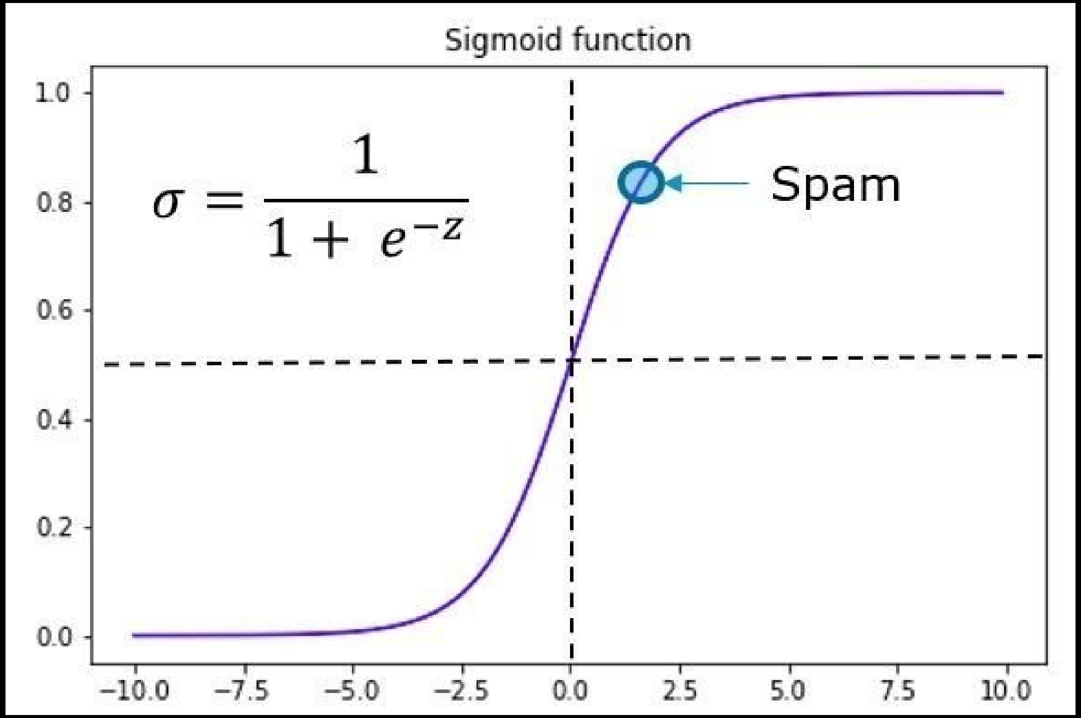
Algoritmo de clasificación

- Aprendizaje supervisado:
    - Regresión
    - Clasificación
- Aprendizaje no supervisado:
    - Clusterización
    - Reducción de dimensionalidad
- Aprendizaje por refuerzo

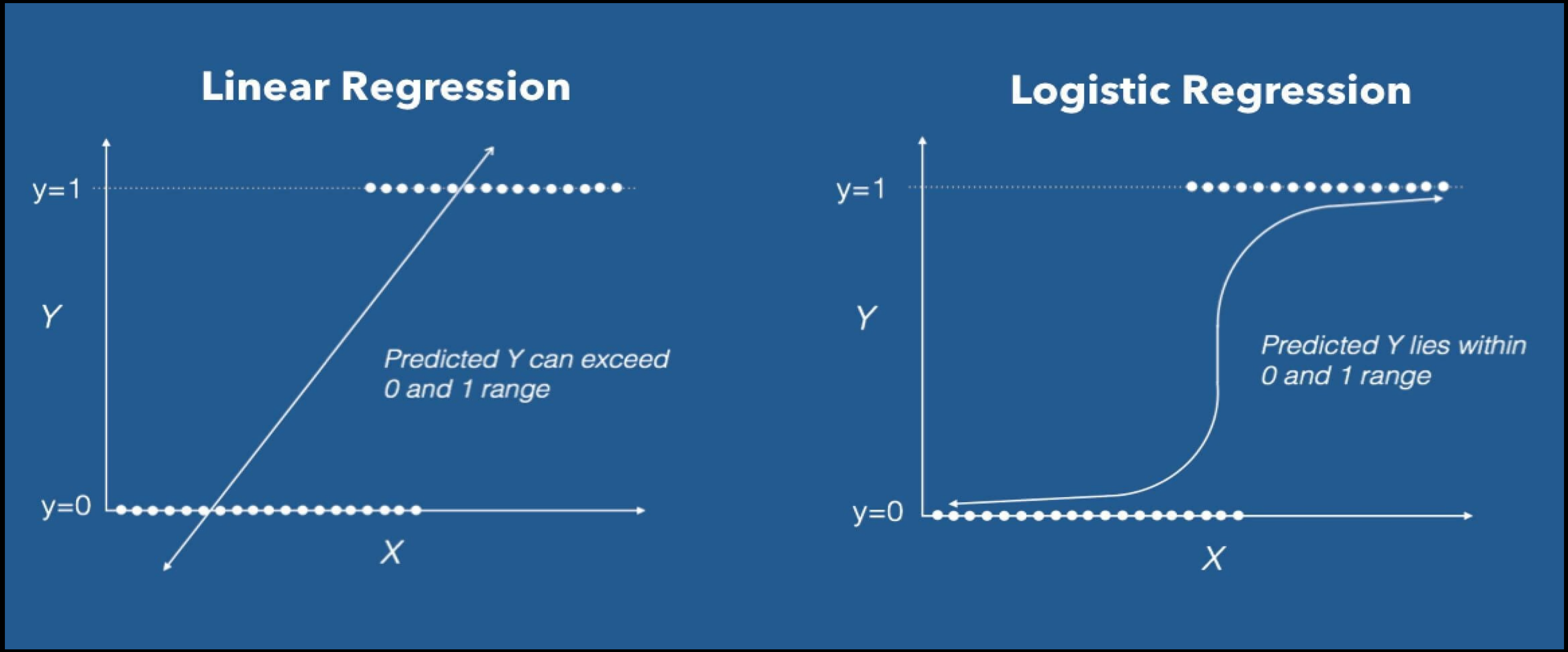


Regresión logística

- Que no te confunda el nombre! Es clasificación.
- Es un tipo de análisis de regresión utilizado para predecir el resultado de una variable categórica (y) en función de las variables independientes (x).
- Util para modelar la probabilidad de un evento ocurriendo como función de otros factores.
- Se usa principalmente en clasificación binaria
- Devuelve las probabilidades de pertenencia a la clase 1



Regresión logística vs Regresión lineal



Las predicciones no estarían en el rango [0, 1]

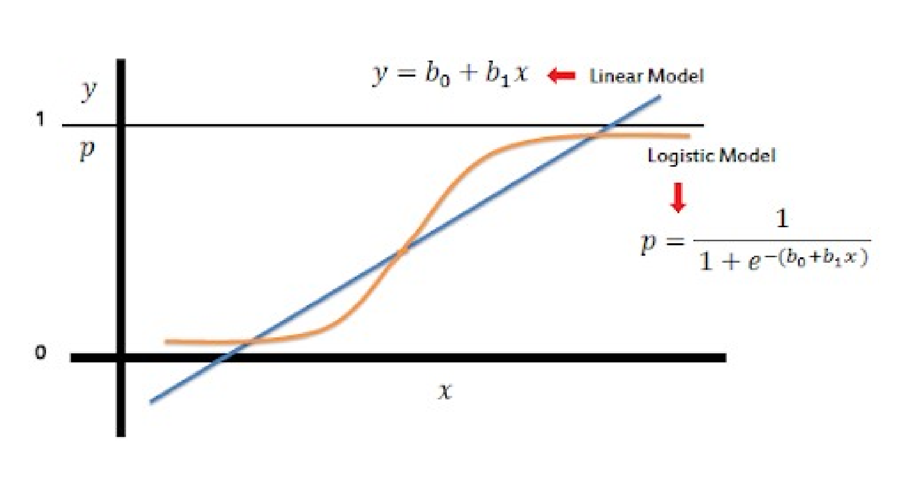
Función sigmoide para acotar el resultado en [0, 1]

Regresión logística

- Su fórmula se obtiene a partir de la función sigmoide. Esta función tiene un valor de entrada y la operación que realiza ofrece un valor comprendido entre [0,1].

- En la resolución de problemas de clasificación, este valor de entrada es toda la ecuación de la regresión lineal.

- Como hacíamos con la regresión lineal, la regresión logística trata de ajustar los valores para “a” y “b”



Formulación matemática

1. Partimos de la función sigmoide

$$\text{sigmoide} = \sigma(y) = \frac{1}{1+e^{-y}}$$

2. Calculamos la probabilidad de que y = 1

$$P(y=1|X=x) = \\ = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\ldots+\beta_p x_p)}} = \\ = \frac{1}{\frac{e^{\beta_0+\beta_1 x_1+\ldots+\beta_p x_p}}{e^{\beta_0+\beta_1 x_1+\ldots+\beta_p x_p}} + \frac{1}{e^{\beta_0+\beta_1 x_1+\ldots+\beta_p x_p}}} = \\ = \frac{1}{\frac{1+e^{\beta_0+\beta_1 x_1+\ldots+\beta_p x_p}}{e^{\beta_0+\beta_1 x_1+\ldots+\beta_p x_p}}} = \\ = \frac{e^{\beta_0+\beta_1 x_1+\ldots+\beta_p x_p}}{1+e^{\beta_0+\beta_1 x_1+\ldots+\beta_p x_p}}$$

$$P(y=1|X=x) = \frac{e^{\beta_0+\beta_1 x_1+\ldots+\beta_p x_p}}{1+e^{\beta_0+\beta_1 x_1+\ldots+\beta_p x_p}}$$

3. Tenemos los coeficientes en los exponentes, por lo que no resulta un modelo lineal. Calculamos la probabilidad de que y = 0

$$P(y=0|X=x) = 1 - P(y=0|X=x) = 1 - \frac{e^{\sum \beta_i x_i}}{1+e^{\sum \beta_i x_i}} = \frac{1}{1+e^{\sum \beta_i x_i}}$$

Dividimos ambas probabilidades y aplicamos ln

$$ln\left(\frac{p(y=1|X=x)}{p(y=0|X=x)}\right) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$$

4. Ya tenemos una ecuación lineal.
El término de la izquierda se conoce como log of odds

De tal manera que cuando tengamos una respuesta, obtendremos la probabilidad de pertenencia a la clase 1:

p(y = 0) = 1 - p(y = 1)

Gradient Descent

Al igual que en Regresión Lineal, la Regresión Logística tiene su función de coste, que hay que minimizar para obtener los pesos (w) de la regresión que minimizan los errores.

¿Cómo solventamos esto? → De nuevo, **Gradient Descent**.

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

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

$$m = number\ of\ samples$$

[Gradient Descent para Logistic Regression]()

Condiciones

- No colinealidad
- Independencia de las observaciones (aplica principalmente en Time Series)
- Grandes muestras funcionan siempre mejor
- Ausencia de outliers
- La relación entre variable respuesta y variables independientes no tiene por qué ser lineal
- Los errores no tienen por qué distribuirse normalmente

Coeficientes

- Ahora las betas las interpreto como cuánto aumenta el *log odds* cuanto aumenta en una unidad el regresor. No muy intuitivo…

$$ln(\frac{p(y=1|X=x)}{p(y=0|X=x)}) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$$

- Si lo que queremos es obtener el feature importance, nos sirve la técnica empleada en regresión lineal: estandarizamos las features y después realizmaos un ranking de las betas.

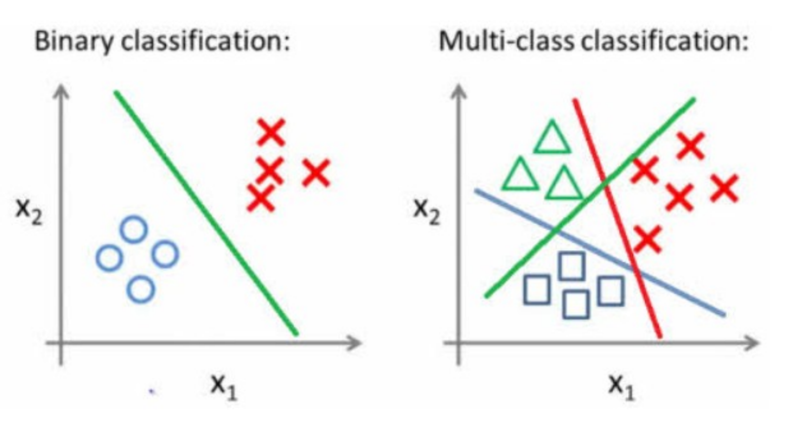
¿Qué pasa si tenemos multiclase?

Se utiliza la estrategia OvA → One vs Rest

1. Si hay N clases, montaremos N modelos
2. Cada uno de estos modelos me dará la probabilidad de pertenencia a cada clase
3. La clase ganadora de la predicción será la que tenga una mayor probabilidad

Por ejemplo, si tenemos un clasificador de tipos de vino (blanco, tinto, rosado), tendremos tres modelos:

4. Tinto vs [blanco, rosado]
5. Blanco vs [tinto, rosado]
6. Rosado vs [tinto, blanco]



Bibliografía

https://learning.oreilly.com/library/view/hands-on-machine-learning/9781492032632/ch04.html#idm45022189758376

https://www.cienciadedatos.net/documentos/py17-regresion-logistica-python.html

https://medium.com/analytics-vidhya/understanding-logistic-regression-b3c672deac04