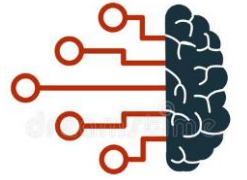


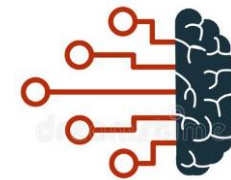
Aprendizaje Supervisado

Regresión

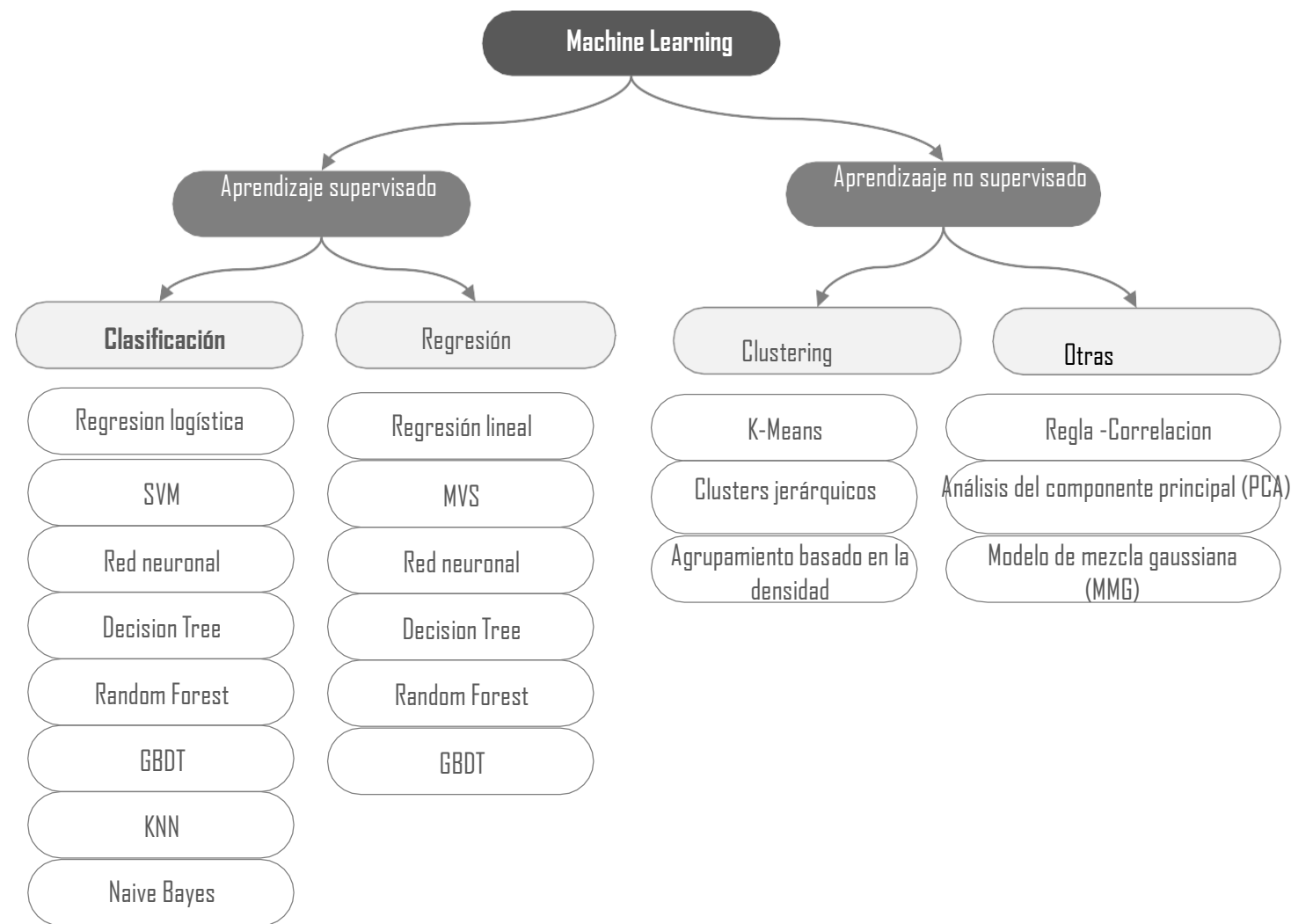
Contenido

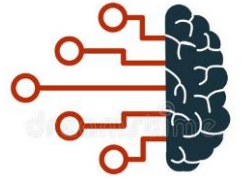


1. Introducción
2. Regresión
3. Árboles de Decisión
4. SVM
5. KNN



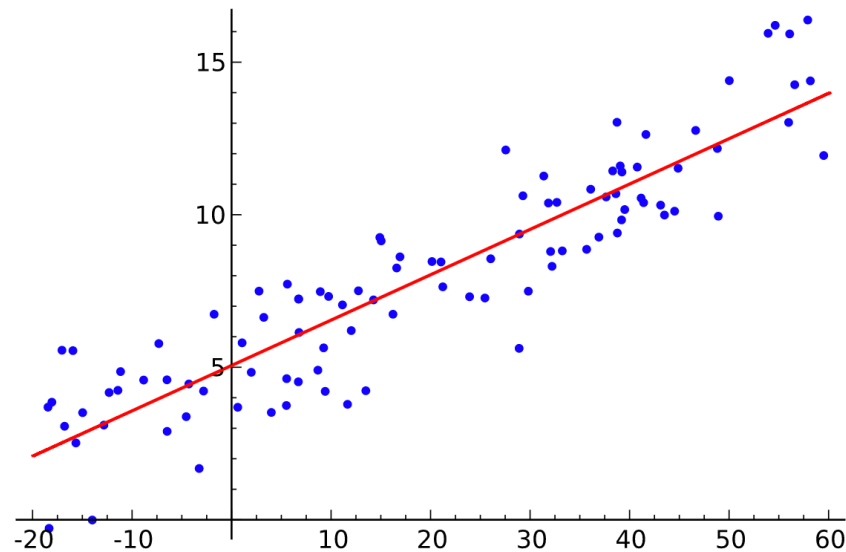
Algoritmos de Aprendizaje automático



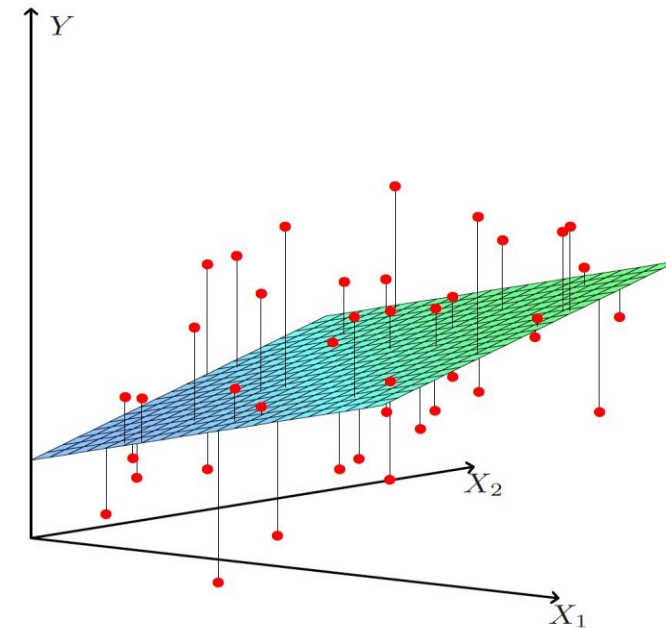


Regresión lineal

- Regresión lineal: un método de análisis estadístico para determinar las relaciones cuantitativas entre dos o más variables a través del análisis de regresión en estadísticas matemáticas.
- La regresión lineal es un tipo de aprendizaje supervisado.

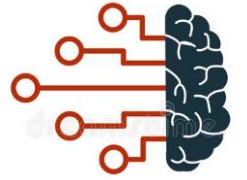


Regresión lineal unaria



Regresión lineal multidimensional

Regresión lineal



La función del modelo de regresión lineal es la siguiente, donde w indica el parámetro de ponderación, b indica el sesgo y x indica el atributo de la muestra.

$$h_w(x) = w^T x + b$$

La relación entre el valor predicho por el modelo y el valor real es la siguiente, donde y indica el valor real y ε indica el error.

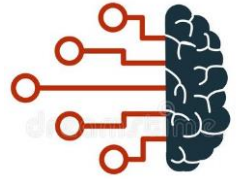
$$y = w^T x + b + \varepsilon$$

El error ε está influenciado por muchos factores de forma independiente. Según el teorema del límite central, el error ε sigue una distribución normal. Según la función de distribución normal y la estimación de máxima verosimilitud, la función de pérdida de la regresión lineal es la siguiente:

$$J(w) = \frac{1}{2m} \sum (h_w(x) - y)^2$$

Para que el valor previsto se acerque al valor real, debemos minimizar el valor de pérdida. Podemos usar el método de descenso de gradiente para calcular el parámetro de peso w cuando la función de pérdida alcanza el mínimo y luego completar la construcción del modelo.

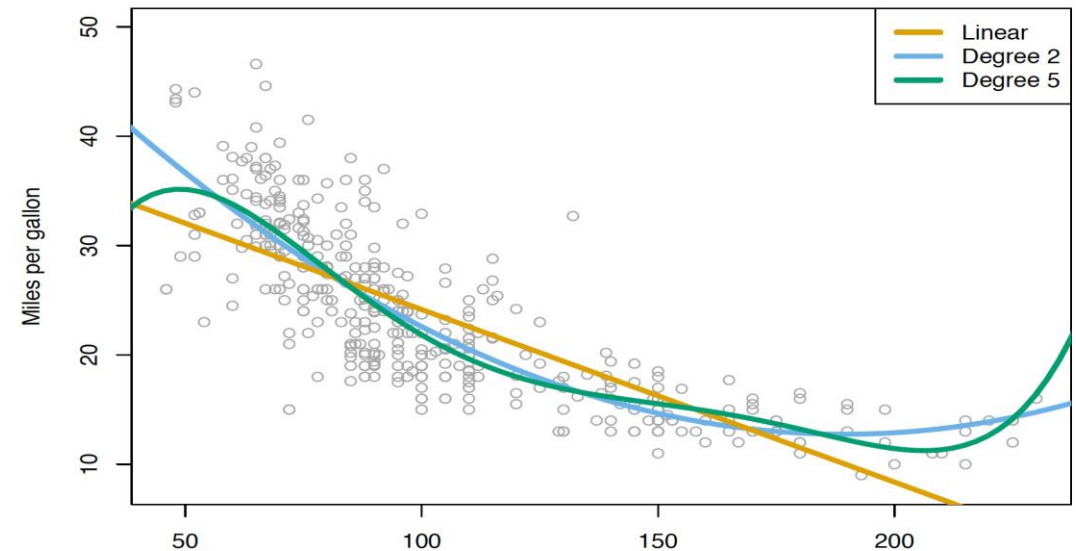
Regresión polinómica



La regresión polinómica es una extensión de la regresión lineal. En general, la complejidad de un conjunto de datos excede la posibilidad de que se pueda ajustar por una línea recta. Es decir, la falta de adecuación obvia ocurre si se utiliza el modelo de regresión lineal original. La solución es utilizar la regresión polinómica.

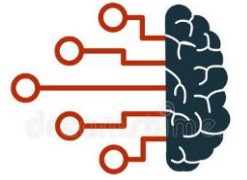
$$h_w(x) = w_1x + w_2x^2 + \dots + w_nx^n + b$$

- donde, la n -ésima potencia es una dimensión de regresión polinomial (grado).
- La regresión polinomial pertenece a la regresión lineal ya que la relación entre sus parámetros de peso w sigue siendo lineal, mientras que su no linealidad se refleja en la dimensión de la característica.



Comparación entre regresión lineal y regresión polinómica

Regresión - prevención de sobreajuste



Los términos de regularización se pueden utilizar para reducir el sobreajuste. El valor de w no puede ser demasiado grande o demasiado pequeño en el espacio de muestra. Puede agregar una pérdida de suma cuadrada en la función de destino.

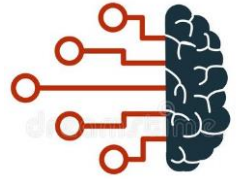
$$J(w) = \frac{1}{2m} \sum (h_w(x) - y)^2 + \lambda \sum \|w\|_2^2$$

Términos de regularización (norma): El término de regularización aquí se denomina L2-norma. La regresión lineal que utiliza esta función de pérdida también se denomina regresión Ridge.

$$J(w) = \frac{1}{2m} \sum (h_w(x) - y)^2 + \lambda \sum \|w\|_1$$

La regresión lineal con pérdida absoluta se denomina regresión de Lasso.

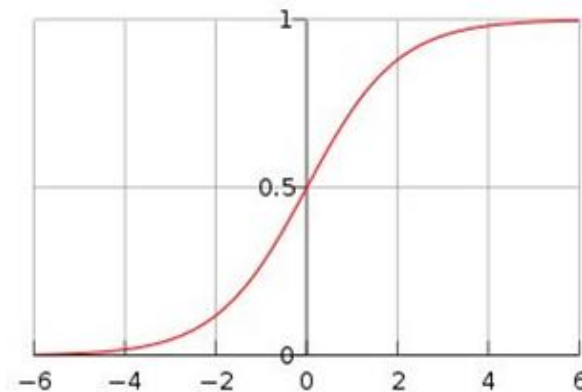
Regresión logística



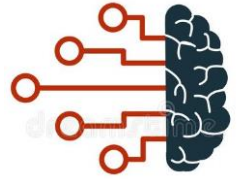
Regresión logística: el modelo de regresión logística se utiliza para resolver problemas de clasificación. El modelo se define de la siguiente manera:

$$P(Y = 1|x) = \frac{e^{wx+b}}{1 + e^{wx+b}}$$
$$P(Y = 0|x) = \frac{1}{1 + e^{wx+b}}$$

donde w indica el peso, b indica el sesgo, y $wx+b$ se considera como la función lineal de x . Compare los dos valores de probabilidad anteriores. La clase con un valor de probabilidad más alto es la clase de x .



Regresión logística

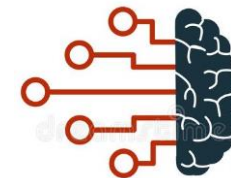


Tanto el modelo de regresión logística como el modelo de regresión lineal son modelos lineales generalizados. La regresión logística introduce factores no lineales (la función sigmoide) basados en la regresión lineal y establece umbrales, por lo que puede tratar los problemas de clasificación binaria.

Según la función modelo de regresión logística, la función de pérdida de la regresión logística se puede estimar de la siguiente manera utilizando la estimación de máxima verosimilitud:

$$J(w) = -\frac{1}{m} \sum (y \ln h_w(x) + (1 - y) \ln(1 - h_w(x)))$$

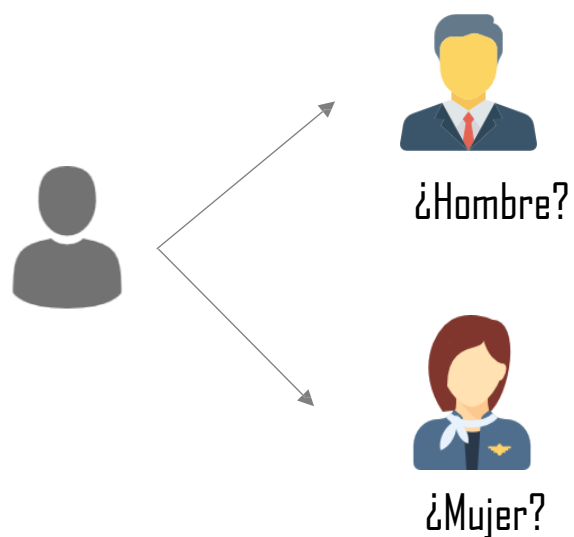
donde w indica el parámetro de peso, m indica el número de muestras, x indica la muestra e y indica el valor real. Los valores de todos los parámetros de peso w también se pueden obtener a través del algoritmo de descenso de gradiente.



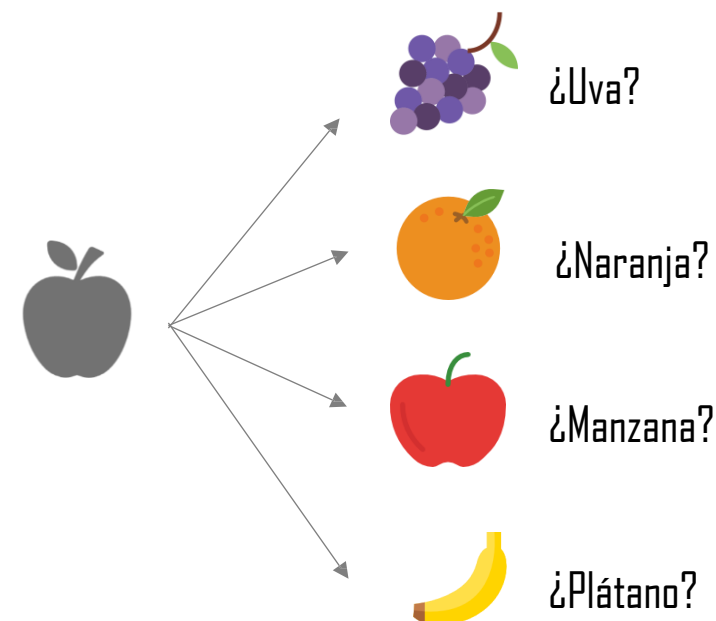
Extensión Regresión logística- Función Softmax

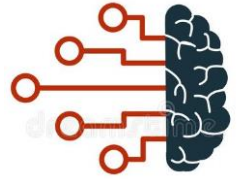
La regresión logística sólo se aplica a problemas de clasificación binaria. Para problemas de clasificación multi-clase, utilice la función Softmax.

Problema de clasificación binaria



Problema de clasificación de varias clases

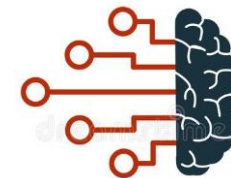




Extensión Regresión logística- Función Softmax

- La regresión Softmax es una generalización de la regresión logística que podemos usar para la clasificación de clase K.
- La función Softmax se utiliza para asignar un vector K-dimensional de valores reales arbitrarios a otro vector K-dimensional de valores reales, donde cada elemento vectorial está en el intervalo (0, 1).
- La función de probabilidad de regresión de Softmax es la siguiente:

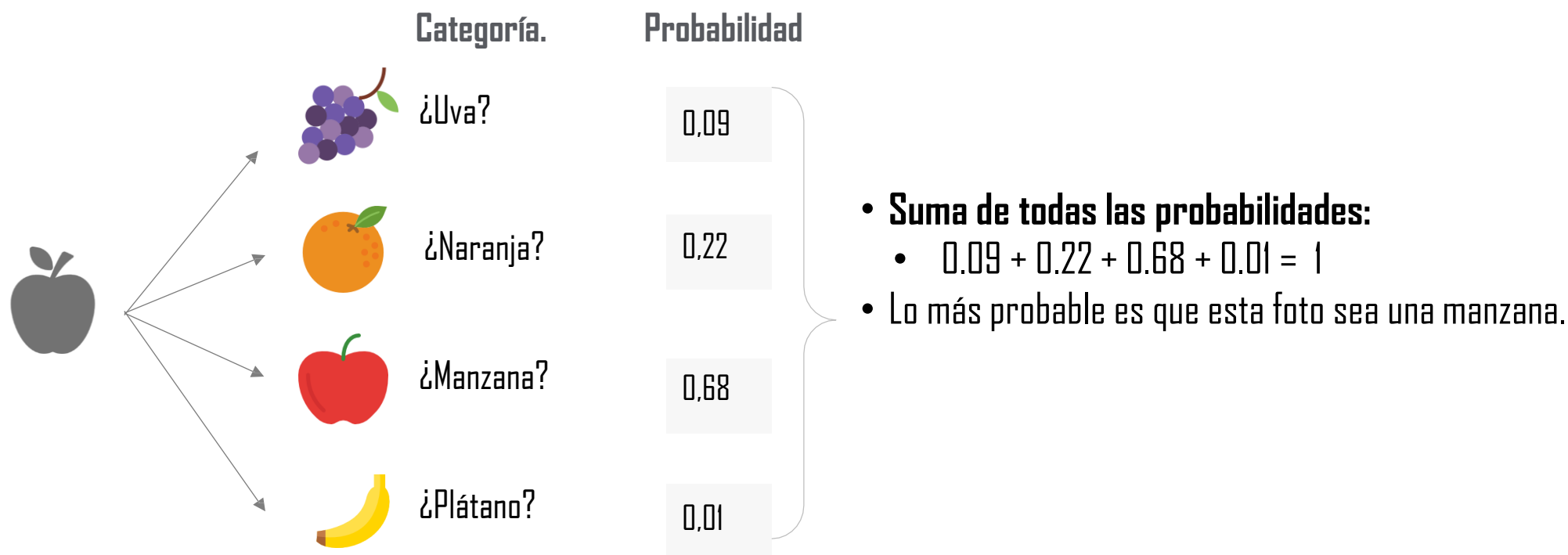
$$p(y = k \mid x; w) = \frac{e^{w_k^T x}}{\sum_{l=1}^K e^{w_l^T x}}, k = 1, 2, \dots, K$$

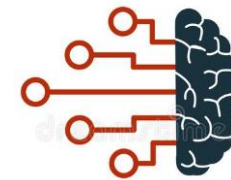


Extensión Regresión logística- Función Softmax

Softmax asigna una probabilidad a cada clase en un problema de varias clases. Estas probabilidades deben sumar 1.

Softmax puede producir un formulario perteneciente a una clase particular. Ejemplo:





Gracias.....