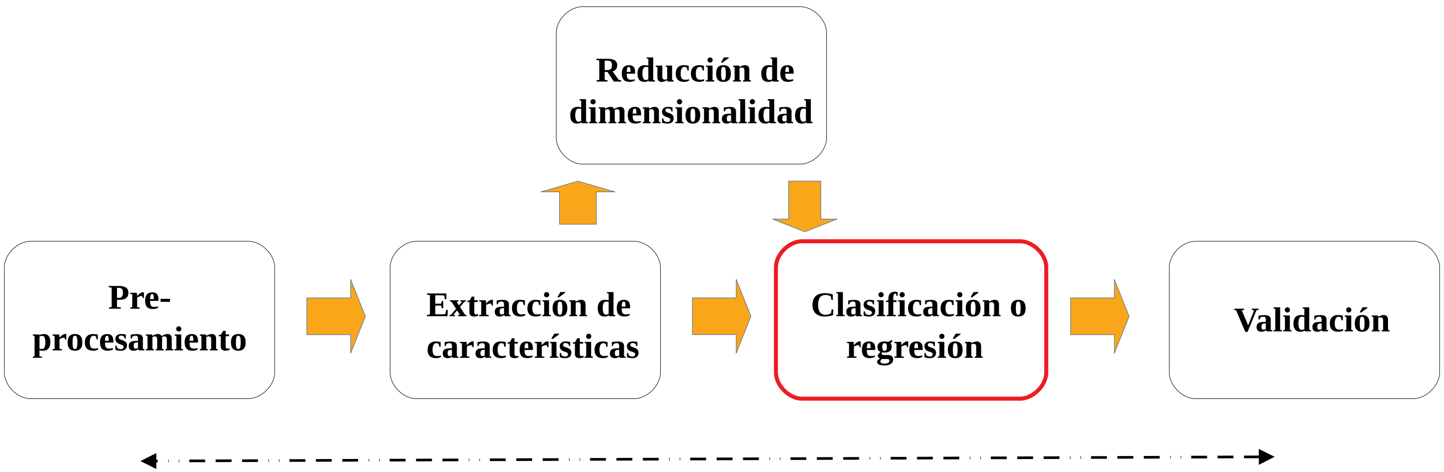
Aprendizaje automático

El problema de ajuste de curvas

Etapas básicas de un sistema de aprendizaje automático supervisado



- El objetivo principal del aprendizaje automático es **crear un modelo que proporcione predicciones precisas** para un conjunto particular de casos.

Dependencia entre variables

La **dependencia entre variables** se puede estudiar como:

- Relación **funcional matemática**, ej. velocidad y distancia recorrida por objeto.
- **Relación estadística**: dados los valores de variables independientes no se puede determinar con exactitud el valor de la variable dependiente, aunque si se puede llegar a **determinar un cierto comportamiento que permite construir un modelo** (ej. la temperatura y la humedad).

Dependencia entre variables

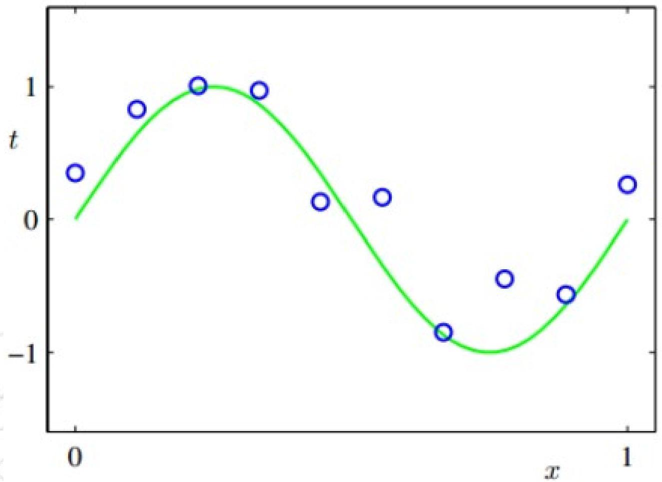
La **relación estadística** se puede plantear como:

- El grado de dependencia existente entre las variables: teoría de la correlación.
- Estructura de la dependencia que mejor exprese la relación: teoría de regresión.

El problema de ajuste de curvas

- El objetivo del ajuste de curvas es encontrar un modelo que se ajuste al conjunto de observaciones (x, y) el cual posibilite **la predicción de la salida** para una nueva muestra x.
- La curva se define de forma **paramétrica**, es decir se deben encontrar los valores de los parámetros que hacen que el **error o pérdida se minimice**.

Ejemplo



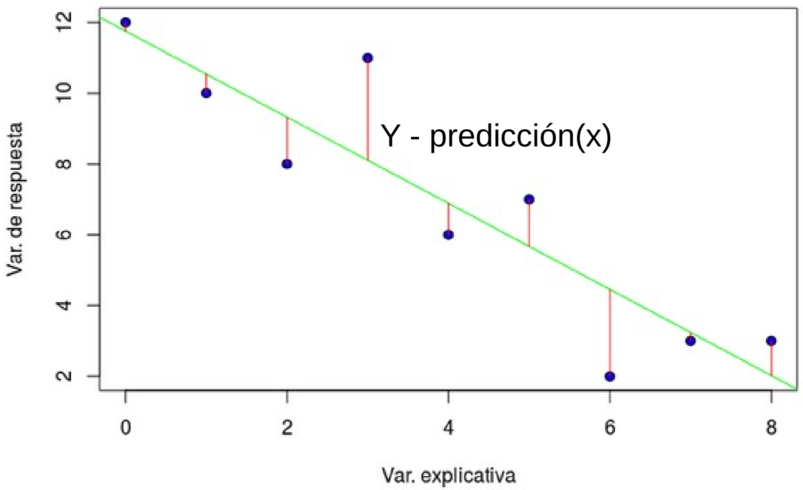
Mínimos cuadrados

- El método de mínimos cuadrados es una técnica comúnmente utilizada en el análisis de regresión.
- Es un método matemático utilizado para encontrar **la línea de mejor ajuste** que representa la relación entre variables independientes y una dependiente.
- El criterio para lograr el mejor ajuste es **el mínimo error cuadrático.**

Mínimos cuadrados

- Busca **minimizar la suma de los cuadrados de las diferencias** entre los puntos generados por la función elegida y los correspondientes valores en los datos.
- Si se tiene conocimiento sobre X ese conocimiento puede ayudar a determinar Y.
- Ejemplo en R^2

$$E(y_i / x_i) = \alpha + \beta x_i$$



Buscamos un alfa y beta que minimicen la pérdida.

$$L_2 Loss = \sum_{(x,y)\in D} (y - prediction(x))^2$$

Con D = conjunto de muestras.

Ajuste de curvas

- Ejemplo de ajuste a una recta, parábola y cúbica

x_i	f_i
0.4	1.00
2.5	0.50
4.3	2.00
5.0	2.55
6.0	4.00

$N = 5$

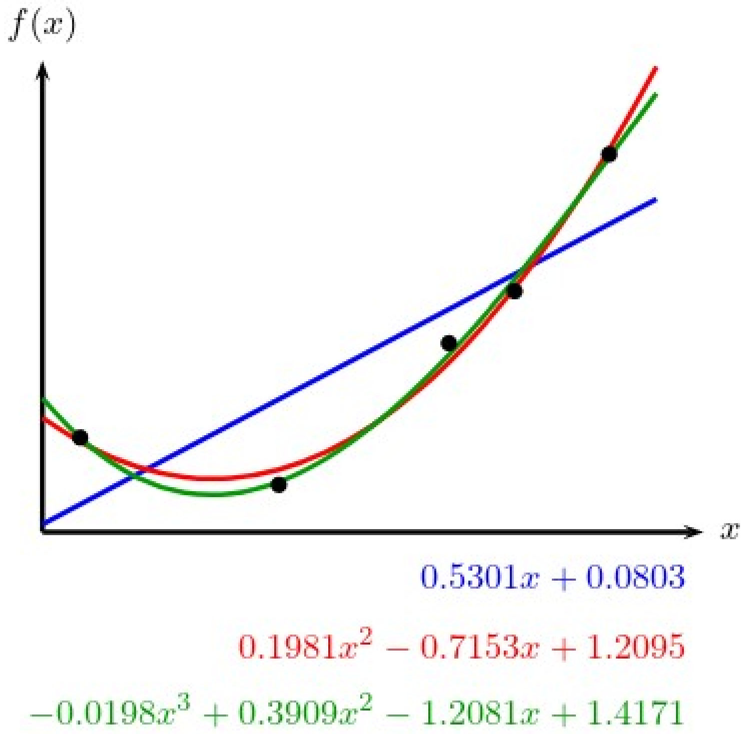


Imagen: http://depa.fquim.unam.mx/amyd/archivero/1_14800.PDF

Mínimos cuadrados

El problema de los mínimos cuadrados en este caso se definirá para encontrar, dada la matriz de rango completo $A \in \mathbb{R}^{m \times n}$, y por ende, invertible y un vector $\vec{b} \in \mathbb{R}^{m \times 1}$, el vector $\vec{x} \in \mathbb{R}^{n \times 1}$ más cercano al **espacio de columnas** de la matriz A, el cual recordamos es denotado como $C(A)$ y corresponde al espacio generado por las columnas de la matriz A, combinadas linealmente por los componentes x_i del vector $\vec{x}$:

$$C(A) = \left\{ \vec{v} \in \mathbb{R}^m : \vec{v} = A\,\vec{x},\ \vec{x} \in \mathbb{R}^n,\ A \in \mathbb{R}^{m \times n} \right\}$$

Se debe encontrar el vector x más cercano al espacio de columnas de A.

Mínimos cuadrados

Asumiendo que A es de rango completo y que $n < m$ se tiene que la proyección del vector $\vec{b} \in \mathbb{R}^{m\times 1}$ **al cuadrado para simplificar su minimización** en el espacio de columnas de la matriz A está dado por:

$$\operatorname{proy}\left(\vec{b}; A\right) = \operatorname{argmin}_{\vec{v}\in C(A)} \left\|\vec{v} - \vec{b}\right\|_2^2 = \operatorname{argmin}_{\vec{x}} \left(A\,\vec{x} - \vec{b}\right) \cdot \left(A\,\vec{x} - \vec{b}\right)$$

Con v = Ax.

Después de calcular el gradiente de la ecuación e igualarlo a cero despejamos el vector x:

$$\Rightarrow \vec{x} = A^{+}\,\vec{b}$$

Referencias

- Bishop (2006). Pattern Recognition and Machine Learning. Recuperado de http://users.isr.ist.utl.pt/~wurmd/Livros/school/Bishop%20-%20Pattern%20Recognition%20And%20Machine%20Learning%20-%20Springer%20%202006.pdf
- Arias Puon, M., Cameras Hau, G. y Guerrero Sanchez, G. (Nd.), Numerictron (Metodos Numericos). Instituto Tecnologico de Tuxtla Gutierrez. Recupreado de: https://sites.google.com/site/numerictron/unidad-4/4-3-regresion-por-minimos-cuadrados-lineal-y-cuadratica