



Mínimos cuadrados es una técnica de optimización matemática que, dados un conjunto de pares de valores numéricos (x, y) , intenta encontrar la función que mejor se ajuste a los datos de entrada usando como criterio la minimización del error cuadrático medio.

$$e_k = y_k - f(x_k)$$

Error en un punto

$$\sqrt{\frac{\sum_{k=1}^n (e_k)^2}{n}}$$

Error cuadrático medio

En este ejercicio usaremos la técnica de mínimos para calcular los coeficientes a y b de una recta de regresión $y = ax + b$ desde una serie de puntos. Los datos de entrada estarán en un fichero separado por comas (CSV) en el que cada línea y contiene cuatro informaciones (género, edad, altura, peso).

Para medir la utilidad de la recta de regresión obtenida usaremos la métrica MAE (*Mean Absolute Error*) que consiste en la media de los valores absolutos de los errores cometidos en cada punto. Dada una serie de puntos $[(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)]$ los coeficientes a y b se calculan con las siguientes expresiones:

```
sum_x = x1 + x2 + ... + xn
sum_y = y1 + y2 + ... + yn
sum_xy = x1*y1 + x2*y2 + ... + xn*yn
sum_xx = x1*x1 + x2*x2 + ... + xn*xn

a = (sum_x*sum_y - n*sum_xy) /
    (sum_x*sum_x - n*sum_xx)
b = (sum y - a*sum x) / n
```

Cálculo de coeficientes

El ejercicio consiste en la implementación de un módulo Python con una serie de funciones que permitirán leer los datos, filtrar las colecciones de registros según distintos criterios, calcular la recta de regresión, calcular la métrica MAE y visualizar el resultado.

