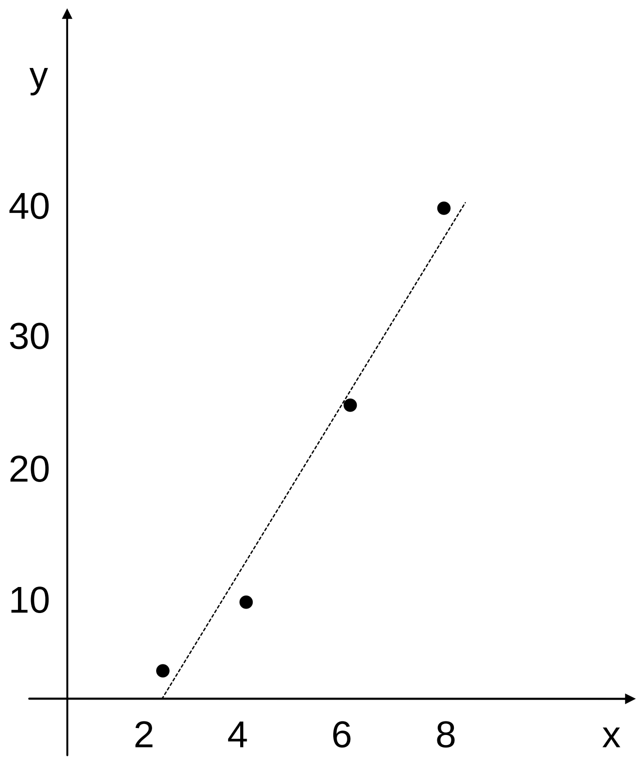
APROXIMAÇÃO POR MÍNIMOS QUADRADOS

Consideremos a seguinte tabela de valores de uma função y = f(x):

i	1	2	3	4
x_i	2	4	6	8
y_i	2	11	28	40

Pretende-se estimar valores da função em pontos não tabelados. Poderíamos utilizar o polinómio interpolador de Lagrange, de grau ≤ 3 (visto termos 4 pontos), ou então um Spline cúbico. Em ambos os casos, e designando por P tal função, ela teria de verificar: $P(x_i) = y_i$, i= 1, 2, 3 e 4.

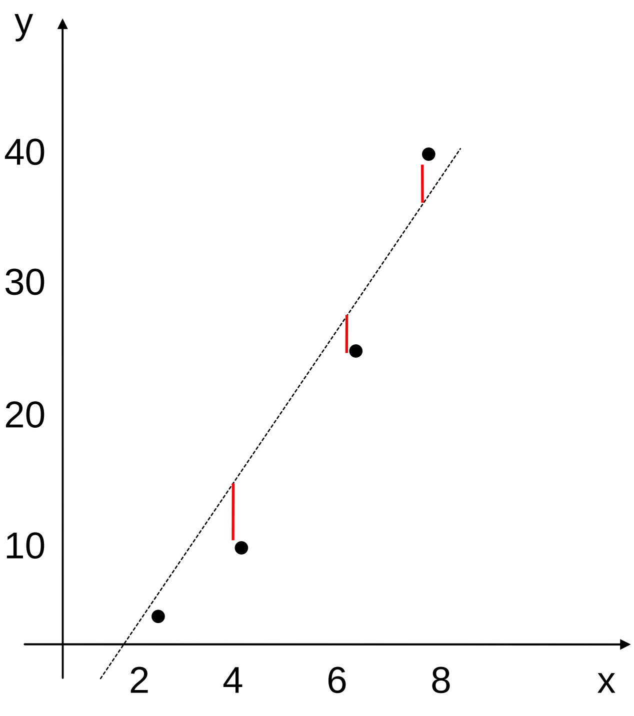
Experimentemos primeiramente, representar em eixos cartesianos o conjunto de pontos da tabela dada



Verifica-se que os pontos se dispõem quase em linha recta (representação gráfica de um polinómio do 1º grau). Se usarmos essa linha recta para aproximar os valores de f(x), essa função "não passará" pelos pontos tabelados, ou melhor dizendo, não será interpoladora de f(x), nesses mesmos pontos.

Em vez de um polinómio interpolador de f(x), pode usar-se a recta que "passe mais próximo" dos pontos tabelados, ou seja que minimize a soma das distâncias dos pontos tabelados à recta. Mas minimizar a soma das distâncias dos pontos tabelados à recta é equivalente a minimizar a soma dos quadrados das distâncias dos pontos tabelados à recta.

A aproximação por **mínimos quadrados** consiste em encontrar a função que "melhor se ajuste", ao conjunto de pontos dado, minimizando o erro resultante do ajustamento, ou seja, pretende-se minimizar a **soma dos quadrados das diferenças entre os valores tabelados e os valores obtidos pela aproximação**.



Designando $ax_i + b$ o i-ésimo valor dado pela aproximação, o problema consiste em encontrar as constantes **a** e **b** que minimizam

$$\sum_{i=1}^{n} [yi - (axi + b)]^2$$

No exemplo apresentado, o problema reduz-se a encontrar as constantes **a** e **b** que minimizam

$$\sum_{i=1}^{4} [y_i - (ax_i + b)]^2 =$$

= [2 - (2a + b)] ^{2} + [11 - (4a + b)] ^{2} + [28 - (6a + b)] ^{2} + [40 - (8a + b)]^2.

Se $\sum_{i=1}^{4} [y_i - (ax_i + b)]^2$ se considerar uma função de duas variáveis, para que um par ordenado (a,b) seja um mínimo local é condição **necessária** (não suficiente) que as derivadas parciais da função em ordem a **a** e **b** se anulem em (a,b), ou seja

$$\frac{\partial}{\partial a} \sum_{i=1}^{4} [y_i - (ax_i + b)]^2 = 0 \qquad \text{e} \qquad \frac{\partial}{\partial b} \sum_{i=1}^{4} [y_i - (ax_i + b)]^2 = 0$$

das quais resultam, para este exemplo, 30a + 5b = 134   e 20a + 4b = 81. A solução deste sistema de equações lineares é a = 6.55 e b = -12.5, logo a função aproximadora pretendida é

$$P(x) = 6.55x - 12.5$$

O problema de encontrar a recta que “melhor se ajuste” a um conjunto de pontos dado, pode então ser formulado como:

$$\text{Min}\sum_{i=1}^{n}\left[y_i - (ax_i + b)\right]^2$$

Para que o mínimo exista é **necessário** que

$$\frac{\partial}{\partial a}\sum_{i=1}^{n}\left[y_i - (ax_i + b)\right]^2 = 2\sum_{i=1}^{n}(y_i - ax_i - b)(-x_i) = 0$$

e

$$\frac{\partial}{\partial b}\sum_{i=1}^{n}\left[y_i - (ax_i + b)\right]^2 = 2\sum_{i=1}^{n}(y_i - ax_i - b)(-1) = 0\,.$$

Estas equações podem ser simplificadas, assumindo a seguinte forma:

$$a\sum_{i=1}^{n}x_i^2 + b\sum_{i=1}^{n}x_i = \sum_{i=1}^{n}x_iy_i$$

e

$$a\sum_{i=1}^{n}x_i + b.n = \sum_{i=1}^{n}y_i$$

a que se costuma chamar **equações normais**, e cuja solução é

$$a = \frac{n(\sum_{i=1}^{n}x_iy_i) - (\sum_{i=1}^{n}x_i).(\sum_{i=1}^{n}y_i)}{n.(\sum_{i=1}^{n}x_i^2) - (\sum_{i=1}^{n}x_i)^2} \qquad \text{e}$$

$$b = \frac{(\sum_{i=1}^{n}x_i^2).(\sum_{i=1}^{n}y_i) - (\sum_{i=1}^{n}x_iy_i).(\sum_{i=1}^{n}x_i)}{n.(\sum_{i=1}^{n}x_i^2) - (\sum_{i=1}^{n}x_i)^2}$$

Consideremos agora o problema mais geral que consiste em aproximar um conjunto de pares ordenados $\{ (x_i, y_i), i=0, 1, 2,..., M \}$, por um polinómio

$$P_n(x) = \sum_{k=0}^{n} a_k x^k$$

de grau $n < M$, usando a técnica dos mínimos quadrados.

O procedimento é análogo ao descrito aquando da utilização de um polinómio de grau 1 (cuja representação gráfica é uma recta) e consiste em encontrar as constantes $a_0, a_1, a_2, ..., a_n$ que minimizam:

$$E = \sum_{i=0}^{M} (y_i - P(x_i))^2$$

$$= \sum_{i=0}^{M} (y_i)^2 - 2\sum_{i=0}^{M} P(x_i).\, y_i + \sum_{i=0}^{M} (P(x_i))^2$$

$$= \sum_{i=0}^{M} y_i^2 - 2\sum_{i=0}^{M} \left(\sum_{j=0}^{n} a_j x_i^j\right).\, y_i + \sum_{i=0}^{M} \left(\sum_{j=0}^{n} a_j x_i^j\right)^2$$

$$= \sum_{i=0}^{M} y_i^2 - 2\sum_{j=0}^{n} a_j.\left(\sum_{i=0}^{M} y_i x_i^j\right) + \sum_{j=0}^{n}\sum_{k=0}^{n} a_j a_k \left(\sum_{i=0}^{M} x_i^{j+k}\right).$$

Tal como no caso linear, para que E tenha um mínimo é **necessário** que

$$\frac{\partial E}{\partial a_j} = 0, j = 0,1,\ldots, n.$$

$$\frac{\partial E}{\partial a_j} = 0 \Leftrightarrow -2\sum_{i=0}^{M} y_i x_i^j + 2\sum_{k=0}^{n} a_k \sum_{i=0}^{M} x_i^{j+k} = 0$$

que constitui um sistema de n+1 incógnitas a_j e n+1 equações:

$$\sum_{k=0}^{n} a_k . \sum_{i=0}^{M} x_i^{j+k} = \sum_{i=0}^{M} y_i x_i^{j}, \qquad j = 0,1,\ldots,n$$

chamadas **equações normais**.

Estas últimas equações podem escrever-se na forma:

$$a_0\sum_{i=0}^{M} x_i^0 + a_1\sum_{i=0}^{M} x_i^1 + a_2\sum_{i=0}^{M} x_i^2 + \cdots + a_n\sum_{i=0}^{M} x_i^n = \sum_{i=0}^{M} y_i x_i^0,$$

$$a_0\sum_{i=0}^{M} x_i^1 + a_1\sum_{i=0}^{M} x_i^2 + a_2\sum_{i=0}^{M} x_i^3 + \cdots + a_n\sum_{i=0}^{M} x_i^{n+1} = \sum_{i=0}^{M} y_i x_i^1,$$

$$\vdots$$

$$a_0\sum_{i=0}^{M} x_i^n + a_1\sum_{i=0}^{M} x_i^{n+1} + a_2\sum_{i=0}^{M} x_i^{n+2} + \cdots + a_n\sum_{i=0}^{M} x_i^{2n} = \sum_{i=0}^{M} y_i x_i^n.$$

Prova-se que as equações normais têm solução única se os valores x_i (i=0, 1, 2,..., M) forem distintos.

Exemplo: Considere a seguinte tabela da função y = f(x):

i	0	1	2	3	4
x_i	0.00	0.25	0.50	0.75	1.00
y_i	1.0000	1.2840	1.6487	2.1170	2.7183

determine a expressão analítica do polinómio de grau dois, que aproxima a função tabelada, utilizando a técnica dos mínimos quadrados.

Resolução: Neste problema n = 2 e M = 4, logo teremos 3 equações normais.

$\sum x_i^0 = 5$ $\sum x_i^1 = 2.5$ $\sum x_i^2 = 1.875$ $\sum x_i^3 = 1.5625$ $\sum x_i^4 = 1.3828$

$\sum y_i^1 = 8.768$ $\sum x_i y_i = 0.4514$ $\sum x_i^2 y_i = 4.4015$

as equações normais para este problema são:

$$a_0 \sum_{i=0}^{4} x_i^0 + a_1 \sum_{i=0}^{4} x_i^1 + a_2 \sum_{i=0}^{4} x_i^2 = \sum_{i=0}^{4} y_i x_i^0$$

$$a_0 \sum_{i=0}^{4} x_i^1 + a_1 \sum_{i=0}^{4} x_i^2 + a_2 \sum_{i=0}^{4} x_i^3 = \sum_{i=0}^{4} y_i x_i$$

$$a_0 \sum_{i=0}^{4} x_i^2 + a_1 \sum_{i=0}^{4} x_i^3 + a_2 \sum_{i=0}^{4} x_i^4 = \sum_{i=0}^{4} y_i x_i^2$$

substituindo valores obtém-se:

$$5.000\, a_0 + 2.500\, a_1 + 1.875\, a_2 = 8.7680$$

$$2.500\, a_0 + 1.875\, a_1 + 1.5625\, a_2 = 5.4514$$

$$1.875\, a_0 + 1.5625\, a_1 + 1.3828\, a_2 = 4.4015$$

cuja solução é $a_0 = 1.0052$, $a_1 = 0.8641$ e $a_2 = 0.8437$.

O polinómio de grau dois, que aproxima a função tabelada, utilizando a técnica dos mínimos quadrados é

$$P_2(x) = 0.8437\, x^2 + 0.8641\, x + 1.0052$$

Embora a técnica dos mínimos quadrados utilize normalmente funções aproximadoras polinomiais, nalguns casos podem utilizar-se outras funções, como por exemplo a função exponencial.

Neste caso a função aproximadora será da forma:

$$y = be^{ax}$$

com **a** e **b** constantes.

Ao aplicar a técnica dos mínimos quadrados, pretende-se minimizar

$$E = \sum_{i=0}^{M} (y_i - b.\,e^{ax_i})^2$$

as equações normais associadas são:

$$\frac{\partial E}{\partial b} = 2\sum_{i=1}^{4} \left[y_i - b.\,e^{ax_i}\right] \left[-e^{ax_i}\right] = 0$$

$$\frac{\partial E}{\partial a} = 2\sum_{i=1}^{4} \left[y_i - b.\,e^{ax_i}\right] \left[-b.\,x_i e^{ax_i}\right] = 0$$

mas não têm solução exacta.

O método usualmente seguido para dados que possuam uma relação exponencial consiste em aplicar a função logaritmo à função aproximadora:

$$y = be^{ax} \Leftrightarrow \ln y = \ln b + ax$$

surgindo assim uma relação linear para ln b e a. O procedimento aplicado aquando da obtenção de uma função aproximadora linear pode agora ser aplicado, com as devidas adaptações.

Exemplo : Considere os dados da seguinte tabela:

i	x_i	y_i
1	1.00	5.10
2	1.25	5.79
3	1.50	6.53
4	1.75	7.45
5	2.00	8.46

Se se representar em eixos cartesianos os valores de x_i e $\ln y_i$, verifica-se a existência de uma relação linear, o que sugere a utilização de uma função aproximadora do tipo:

$$y = be^{ax} \Leftrightarrow \ln y = \ln b + ax$$

i	x_i	$\ln y_i$	x_i^2	$x_i \ln y_i$
1	1.00	1.629	1.0000	1.629
2	1.25	1.756	1.5625	2.195
3	1.50	1.876	2.2500	2.814
4	1.75	2.008	3.0625	3.514
5	2.00	2.135	4.0000	4.270
$\sum$	7.50	9.404	11.875	14.422

$$a = \frac{n(\sum_{i=1}^{n} x_i y_i) - (\sum_{i=1}^{n} x_i).(\sum_{i=1}^{n} y_i)}{n.(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)^2}$$

$$b = \frac{(\sum_{i=1}^{n} x_i^2).(\sum_{i=1}^{n} y_i) - (\sum_{i=1}^{n} x_i y_i).(\sum_{i=1}^{n} x_i)}{n.(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)^2}$$

obtém-se

$$a = \frac{(5)(14.422) - (7.5)(9.404)}{(5)(11.875) - (7.5)^2} = 0.5056$$

$$\ln b = \frac{(11.875)(9.404) - (14.422)(7.5)}{(5)(11.875) - (7.5)^2} = 1.122$$

note-se que ln b = 1.122 $\Leftrightarrow$ b = $e^{1.122}$ = 3.071, pelo que a função aproximadora pretendida é

$$y = 3.071\, e^{1.122\, x}$$