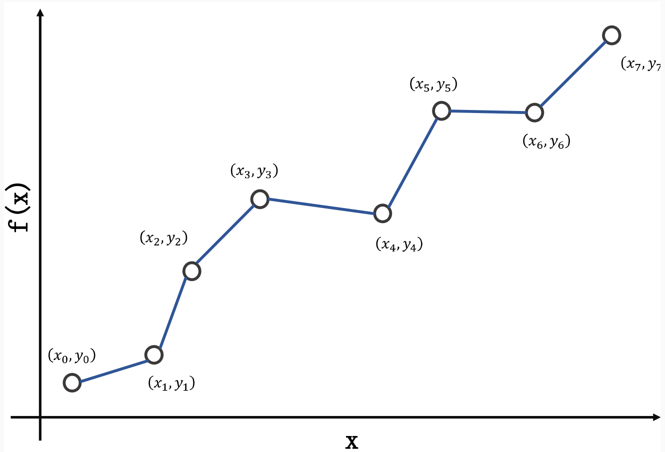
Interpolation

- The process of finding a simple function from a given set of discrete data points in such a way that it passes through all these data points.
- The function thus found can be used to estimate the intermediate response of the variables in the defined parameter range.
- There are several methods for interpolation over a given data points.
- Polynomial functions are very popular interpolation because they are easy to evaluate, and have nice mathematical properties such as continuity, differentiability, and integrability.

Linear Interpolation

In linear interpolation, a straight line is used to connect two adjacent data points.



Linear Interpolation

- Formula for finding a line between two adjacent data points (x_0, y_0) and (x_1, y_1) is

$$\frac{y - y_0}{y_1 - y_0} = \frac{x - x_0}{x_1 - x_0}$$

i.e.,

$$y = y_0 + \frac{x - x_0}{x_1 - x_0}(y_1 - y_0)$$

- Or in general form

$$f_k(x) = y_k + \frac{x - x_k}{x_{k+1} - x_k}(y_{k+1} - y_k) \qquad 1 \leq k \leq n$$

Linear interpolation is the quickest and easiest interpolation method. But it suffers several drawbacks such as lack of precision, non-differentiability of the fit at the point $x = x_k$ etc.

Lagrange Interpolating Polynomial

- Consider $(k+1)$−data points $(x_0, y_0), (x_1, y_1), \ldots, (x_k, y_k)$ such that $x_i \neq x_j$ for any $i, j = 0, 1, \ldots, k$.
- The interpolation polynomial in the Lagrange form is a linear combination

$$f(x) = \sum_{i=0}^{k} y_i P_i(x)$$

Where $P_i(x) = \prod_{j=0,\ j\neq i}^{k} \frac{x - x_j}{x_i - x_j} = \frac{x - x_0}{x_i - x_0}\frac{x - x_1}{x_i - x_1} \cdots \frac{x - x_k}{x_i - x_k}$

Each $P_i(x)$ is a polynomial of degree k.

and $P_i(x = x_m) = \begin{cases} 1 & \text{if } m = i \\ 0 & \text{if } m \neq i \end{cases}$

Lagrange interpolations Pseudo Code

Algorithm 1 Lagrange Interpolating Polynomial

Input: $y = 0, (x_i, y_i), i = 1, 2, 3, \ldots, k+1$

Output: y

for $i = 1$ to $k+1$ **do**
  $P = y_i$
  for $j = 1$ to $k+1$ **do**
    if $(i \neq j)$ **then**
      $P = P * (x - x_j)/(x_i - x_j)$
    end if
  end for
  $y = y + P$
end for
return y

Curve Fitting

For a given data set, one may be interested to get the relationship between various problem variables.

- For this we try to find a suitable function using these data points. The process of finding a function that fits best to a series of data points is called curve fitting.
- One of the biggest advantage of finding the best fit to a given data set is that, just by substituting the value in the equation of the fitting function, we can estimate the values at the points that are not observed.
- Curve fitting is different from interpolation. In interpolation, the unobserved values are approximated using the interpolating function between two observed points. However, in curve fitting one can approximate the values outside the observation range.

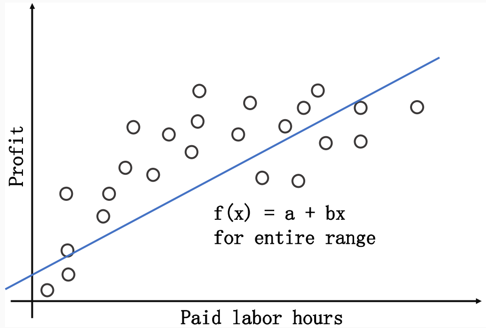
Linear curve Fitting (Linear Regression)

Fitting a straight line to a given data set using the method of least squares is one of the most popular curve fitting method.

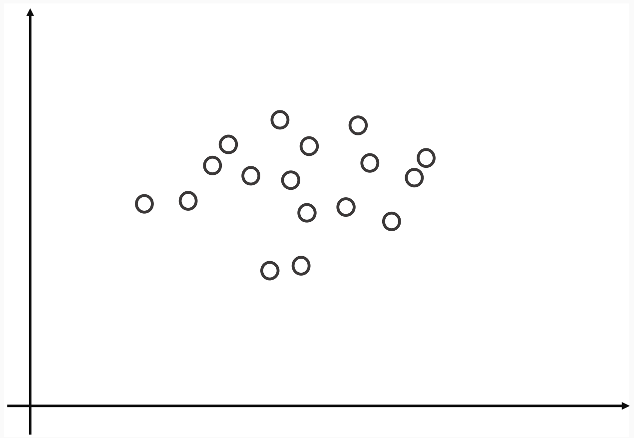
The equation of this least square line can be written as

$$f(x) = a + bx$$

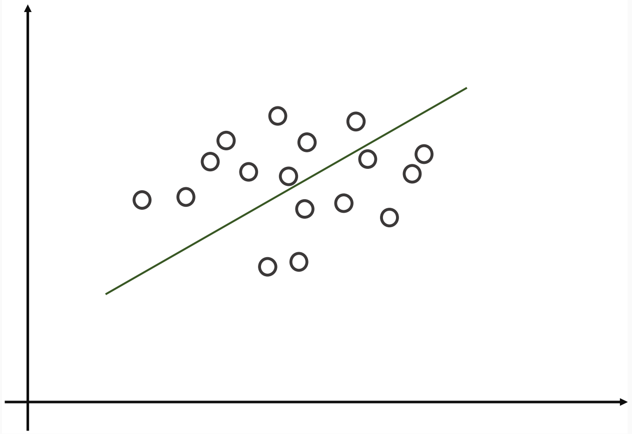
where a and b are unknowns coefficients and needs to be found such that $f(x)$ fits data "well" (least square error in this case).



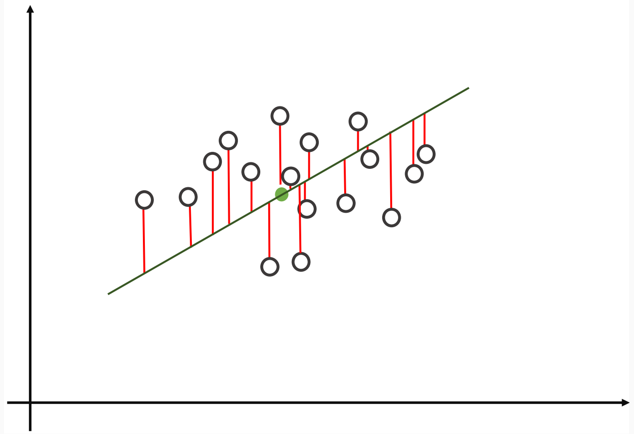
Linear Curve Fitting



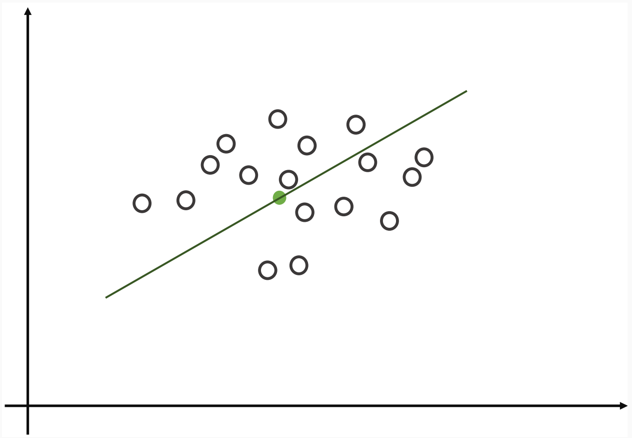
Linear Curve Fitting



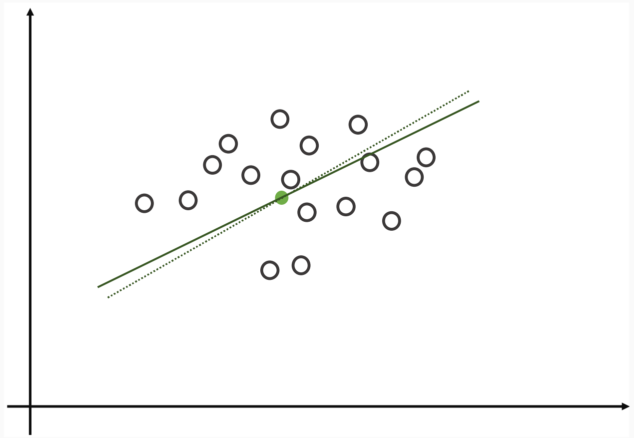
Linear Curve Fitting



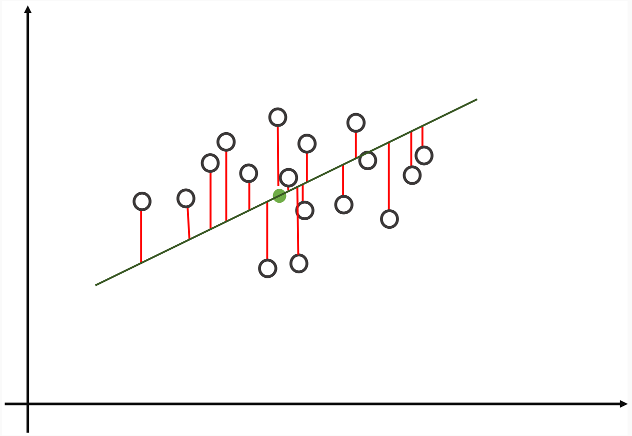
Linear Curve Fitting



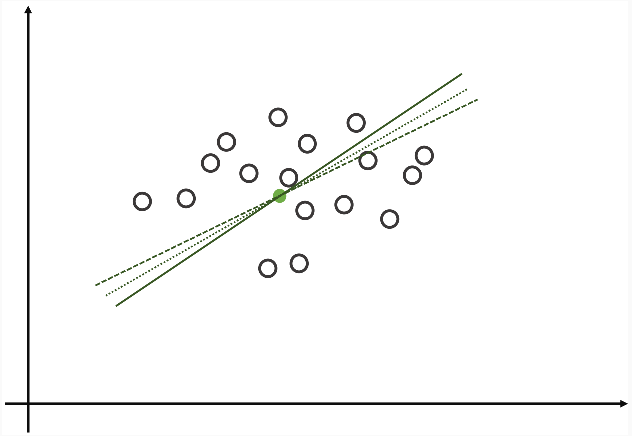
Linear Curve Fitting



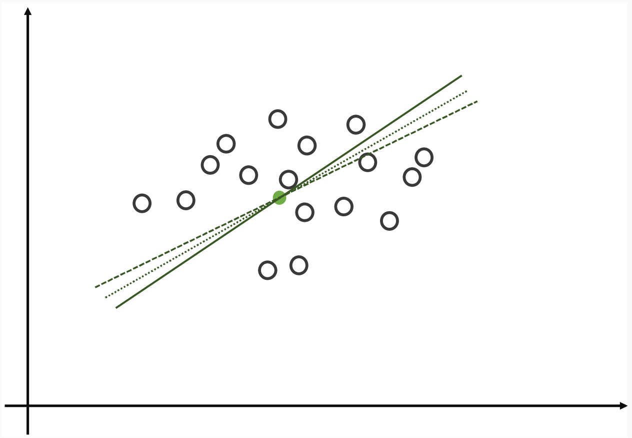
Linear Curve Fitting



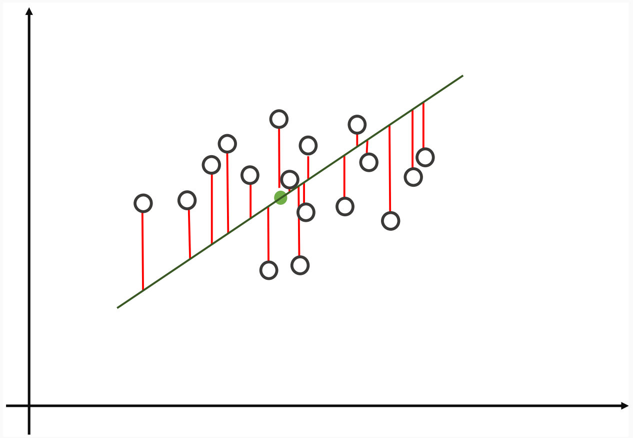
Linear Curve Fitting



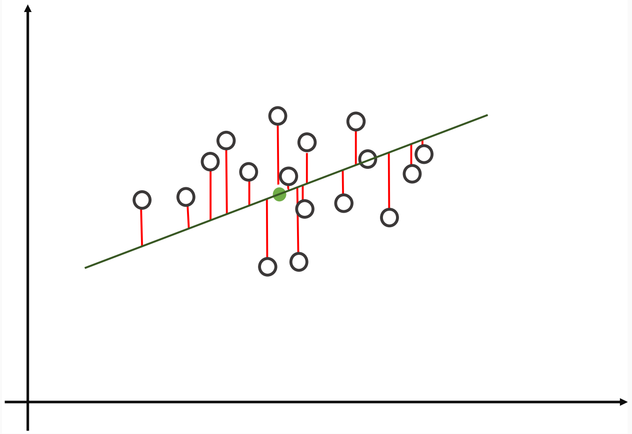
Linear Curve Fitting



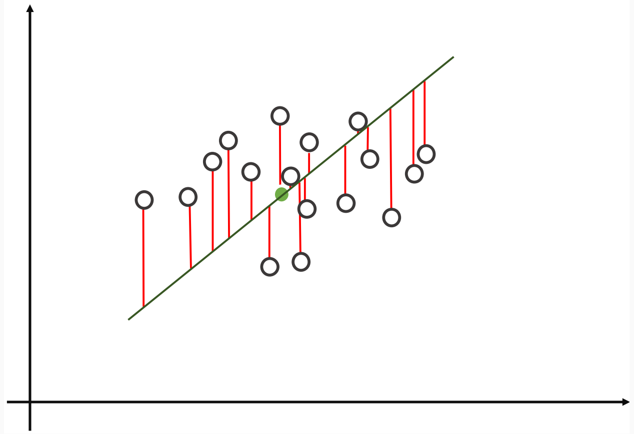
Linear Curve Fitting



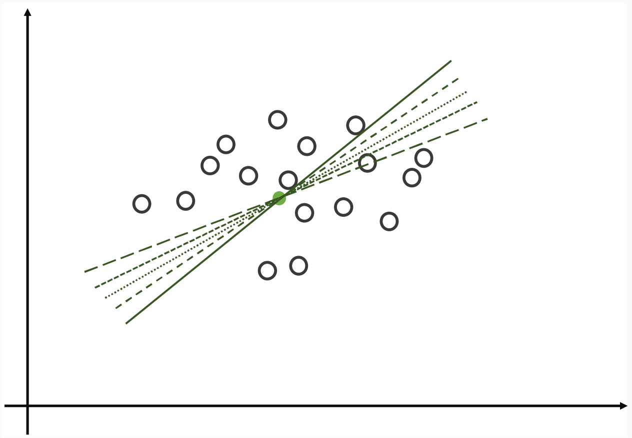
Linear Curve Fitting



Linear Curve Fitting



Linear Curve Fitting



Quantifying error in a curve fit

In previous plots, the line which provides the minimum error is the 'best fit' linear function.

How to quantify the error?

Quantifying the error in the least square sense is one of the most convenient and popular way. Steps to least square error are given below

- Let $f(x) = a + bx$ be the linear fit to the given data set.
- Take a point (x_i, y_i) from the given data set.
- Find the value $f(x_i)$ for the chosen point using the linear fit $f(x) = a + bx$.
- For each point $(x_i, y_i), i = 1, 2, \ldots, k$, find the deviation $d_i (= y_i - f(x_i))$ of $f(x_i)$ from y_i.

Quantifying error in a curve fit

Find the sum of squared deviations E for all data points

$$E = \sum_{i=1}^{n} (d_i)^2 = \sum_{i=1}^{n} (y_i - f(x_i))^2 = \sum_{i=1}^{n} (y_i - (a + bx_i))^2$$

Least squares approach tries to find the unknowns a and b such that sum of the square of the deviation is minimum. This leads to the following optimization problem.

$$\text{minimize } E = \sum_{i=1}^{n} (y_i - (a + bx_i))^2$$

Finding the least square linear fit

To find the best least square fit, one need to solve the following minimization problem.

$$\text{minimize } E = \sum_{i=1}^{n} (y_i - (a + bx_i))^2$$

Taking the partial derivative of the E with respect to a and b respectively and equating them to zero, we get

$$\frac{\partial E}{\partial a} = -2 \sum_{i=1}^{n} \left(x_i \left(y_i - a - bx_i \right) \right) = 0$$

$$\frac{\partial E}{\partial b} = -2 \sum_{i=1}^{n} (y_i - a - bx_i) = 0$$

Finding the least square linear fit

This leads to following system of linear equations

$$b \sum x_i^2 + a \sum x_i = \sum (x_i y_i)$$

$$b \sum x_i + a \cdot n = \sum y_i$$

The above system can be written in the following matrix form

$$\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum (x_i y_i) \end{bmatrix}$$

Which is of the standard form

$$AX = B$$

and can be solved easily

$$X = A^{-1} B \qquad \text{(how?)}$$

Example: Linear Curve Fitting

Example 1: Fit a linear curve to the following data

i	1	2	3	4	5	6
x	0	0.5	1	1.5	2	2.5
y	-2.03	-0.15	2.14	5.56	6.21	8.88

Solution of the following system using the above data would give the least square linear fit

$$\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum (x_i y_i) \end{bmatrix}$$

It can be easily verified that

$$\sum x_i = 7.5 \qquad \sum x_i^2 = 13.75$$

$$\sum y_i = 20.61 \qquad \sum x_i y_i = 45.06$$

Example: Linear Curve Fitting

Substituting the above values, we get

$$\begin{bmatrix} 6 & 7.5 \\ 7.5 & 13.75 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 20.61 \\ 45.06 \end{bmatrix}$$

Now $X = A^{-1}B$ implies

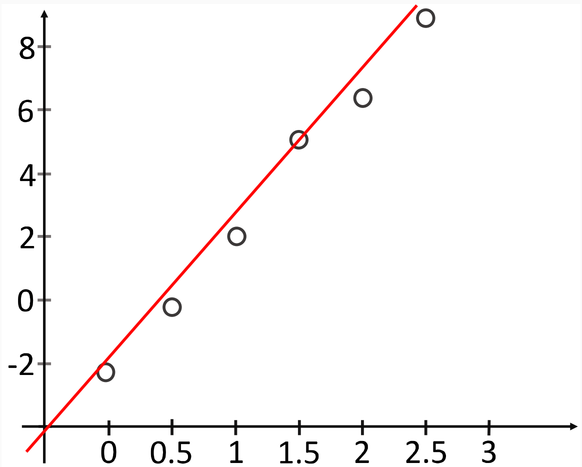
$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 6 & 7.5 \\ 7.5 & 13.75 \end{bmatrix}^{-1} \begin{bmatrix} 20.61 \\ 45.06 \end{bmatrix}$$

will give

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} -2.07 \\ 4.41 \end{bmatrix}$$

Therefore our fit curve is $f(x) = -2.07 + 4.41x$

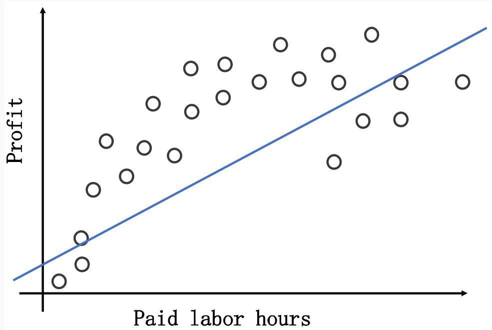
Example: Linear Curve Fitting



This fits the line for the given set of data points.

Polynomial Curve Fitting

What will we do when a straight line is not suitable for the data set? See example below



Straight line fit will not be suitable in such cases and one needs to try some other function for the least square fit.

Polynomial Curve Fitting

Second order polynomial could be next choice if the linear function does not fit data well.

- In linear regression, we were finding coefficients a and b so that error can be minimized.
- The same idea can be extended to fit a second order polynomial to find the 'best' fit.
- Let $f(x) = a_0 + a_1x + a_2x^2$ be the second order polynomial fit to the given data set.
- Take a point (x_i, y_i) from the given data set.
- Find the value $f(x_i)$ for the chosen point using the second order polynomial fit $f(x) = a_0 + a_1x + a_2x^2$.
- For each point (x_i, y_i), $i = 1, 2, \ldots, k$, find the deviation $d_i(= y_i - f(x_i))$ of $f(x_i)$ from y_i.

Finding the least square quadratic fit

To find the best least square fit, one need to solve the following minimization problem.

$$\text{minimize } E = \sum_{i=1}^{n}(y_i - (a_0 + a_1 x_i + a_2 x_i^2))^2$$

Taking the partial derivative of the E with respect to a_0, a_1 and a_2 respectively and equating them to zero, we get

$$\frac{\partial E}{\partial a_0} = -2\sum_{i=1}^{n}\left(y_i - (a_0 + a_1 x_i + a_2 x_i^2)\right) = 0$$

$$\frac{\partial E}{\partial a_1} = -2\sum_{i=1}^{n}\left(x_i\left(y_i - (a_0 + a_1 x_i + a_2 x_i^2)\right)\right) = 0$$

$$\frac{\partial E}{\partial a_2} = -2\sum_{i=1}^{n}\left(x_i^2\left(y_i - (a_0 + a_1 x_i + a_2 x_i^2)\right)\right) = 0$$

Finding the least square quadratic fit

This leads to following system of linear equations

$$a_0 \cdot n + a_1 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 = \sum y_i$$

$$a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 + a_2 \sum_{i=1}^{n} x_i^3 = \sum_{i=1}^{n}(x_i y_i)$$

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

$$\Rightarrow \begin{bmatrix} n & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum(x_i y_i) \\ \sum(x_i^2 y_i) \end{bmatrix}$$

Fitting a general polynomial using Least squares approach

In general, for a polynomial of degree m

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \ldots + a_m x^m = \sum_{k=1}^{m} a_k x^k$$

The least square polynomial fit in this case for a polynomial of degree m can found by solving following linear system of equations.

$$\begin{bmatrix} n & \sum x_i & \sum x_i^2 & \cdots & \sum x_i^m \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \cdots & \sum x_i^{m+1} \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 & \cdots & \sum x_i^{m+2} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum x_i^m & \sum x_i^{m+1} & \sum x_i^{m+2} & \cdots & \sum x_i^{m+m} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum(x_i y_i) \\ \sum(x_i^2 y_i) \\ \vdots \\ \sum(x_i^m y_i) \end{bmatrix}$$

Example Polynomial Curve Fitting

Example 2: Fit a second order polynomial curve for the following data

i	1	2	3	4	5	6
x	0	0.5	1	1.5	2	2.5
y	0.5	1.1	2.2	3.1	4.5	8.4

Solution of the following system using the above data would give the following least square quadratic fit

$$\begin{bmatrix} n & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum(x_i y_i) \\ \sum(x_i^2 y_i) \end{bmatrix}$$

Example Polynomial Curve Fitting

$$\begin{array}{lll} \sum x_i = 7.50 & \sum x_i^2 = 13.75 & \sum x_i^3 = 28.13 \\ \sum x_i^4 = 61.19 & \sum y_i = 19.80 & \sum y_i = 19.80 \\ \sum x_i^2\ y_i = 79.95 & & \end{array}$$

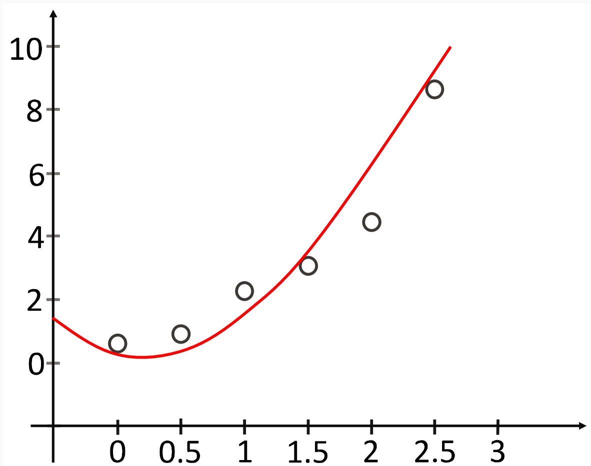
$$\Rightarrow \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 6 & 7.5 & 13.75 \\ 7.5 & 13.75 & 28.13 \\ 13.75 & 28.16 & 61.19 \end{bmatrix}^{-1} \begin{bmatrix} 19.8 \\ 37.4 \\ 79.95 \end{bmatrix}$$

will give us

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 0.75 \\ -0.30 \\ 1.27 \end{bmatrix}$$

Therefore fitted function is $f(x) = 0.75 - 0.3 \cdot x + 1.27 \cdot x^2$

Example Polynomial Curve Fitting



This fits the curve for the given set of data.