

Least Squares Approximation

Aim

To demonstrate how to use the least squares approximation method to fit a curve to a set of data.

Learning Outcomes

At the end of this section you will be able to:

- Understand what least squares approximation is,
- Use least squares approximation to find the best fit linear curve for a given set of data points.

Least squares approximation is a method for calculating the **best fit** linear curve to a given set of data. In the previous section, curve fitting, we showed how it was possible to draw a linear curve that fitted the data provided. Unfortunately in the last section we did not use any particular method to calculate the curve and so there could be slight variations between the curves drawn by two different people. Using the method of *least squares approximation* guarantees that we find the best fit curve for the data, but it also guarantees that everyone has the same linear curve fitted to the data.

A linear curve is represented by the equation

$$y = a + bx.$$

Given a set of n data points

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

we find the values of a and b in the linear equation above by using the following formulas;

$$a = \frac{\sum y - (\sum x)b}{n},$$

$$b = \frac{n \sum xy - (\sum x)(\sum y)}{n \sum x^2 - (\sum x)^2}$$

where \sum represents the sum from 1 to n .

Example

Use the least squares method to fit the best fit curve to the following set of data.

$$(-4.5, 0.7), (-3.2, 2.3), (-1.4, 3.8), (0.8, 5.0), (2.5, 5.5), (4.1, 6.6)$$

We need to come up with the a and b value that best fits the above set of data. To do this we use the formulas on the previous page. To use these formulas we first need to find the following;

n	$\sum x$	$\sum y$	$\sum xy$	$\sum x^2$	$(\sum x)^2$
6	-1.7	23.9	28.98	56.15	2.89

Filling these values into the expressions on the previous page for a and b we get

$$a = 4.16529, \quad \text{and} \quad b = 0.642226.$$

This gives us the equation $y = 4.16529 + 0.642226x$ as the equation of the line that is the best fit to above set of data.

