Interpolation

What is curve fitting?

Suppose we have discrete table of data points $(x_i, y_i), i = 0,1,2, \dots, n,$ where x is independent variable and y is dependent variable. Estimating the value of y for any intermediate value of x we need to construct a function or polynomial $y(x)$ that pass through all the given data points and then evaluate $y(x)$ for any given point x. The process of constructing $y(x)$ that fit the discrete table of data points (x_i, y_i), is called curve fitting. A Table of data points may belongs to two categories:

1. Table of values of well-defined functions
2. Data tabulated from the measurements made during an experiments.

There are two types of curve fitting techniques depending on categories of table of data namely—Interpolation and Regression.

Interpolation:

If the table values are from well-defined function such as logarithmic tables, interest tables, steam tables, etc then constructing a function $y(x)$that passes through all the data points and estimating values at non-tabular points is called interpolation. The function is called interpolating function.

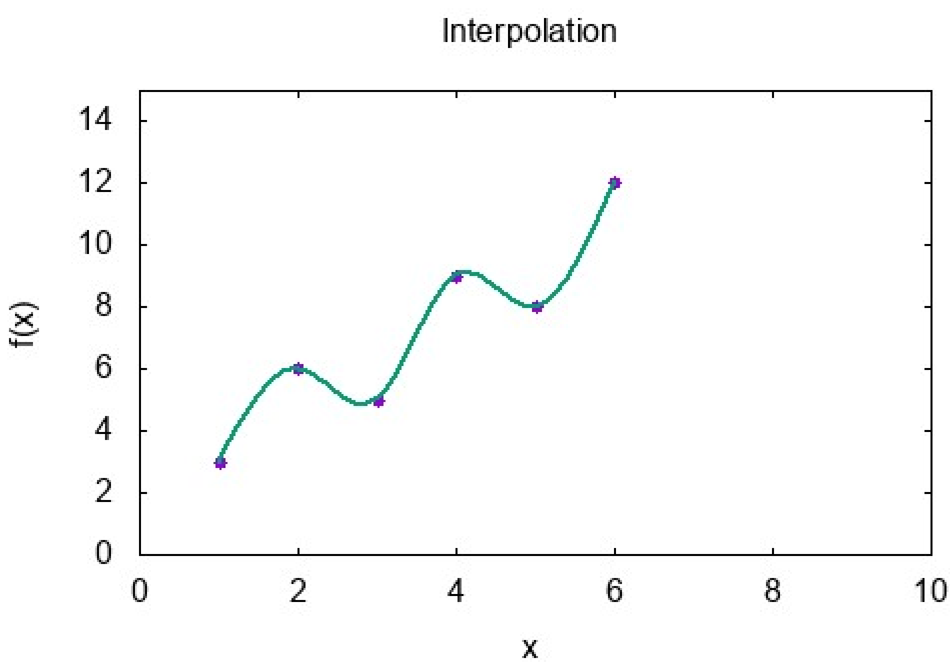

Figure 1: Explanation of Interpolation

So Interpolation is:

- Estimating intermediate values between precise data points.
- We first fit a function that exactly passes through the given data points and then evaluate intermediate values using this function.

If a function, say f(x) is constructed, such that it passes through all the set of data points and then evaluating f(x) for the specified value of x is known as interpolation.

This method of constructing a function gives the estimation of values at non tabular print.

Regression:
If the table values are from the measurements made during an experiment, then constructing a function that not exactly fit all the data points but it represent the general trend of the data is called regression. The function is called approximate function and popular approach for finding the function to best fit the data is known as least-square regression.

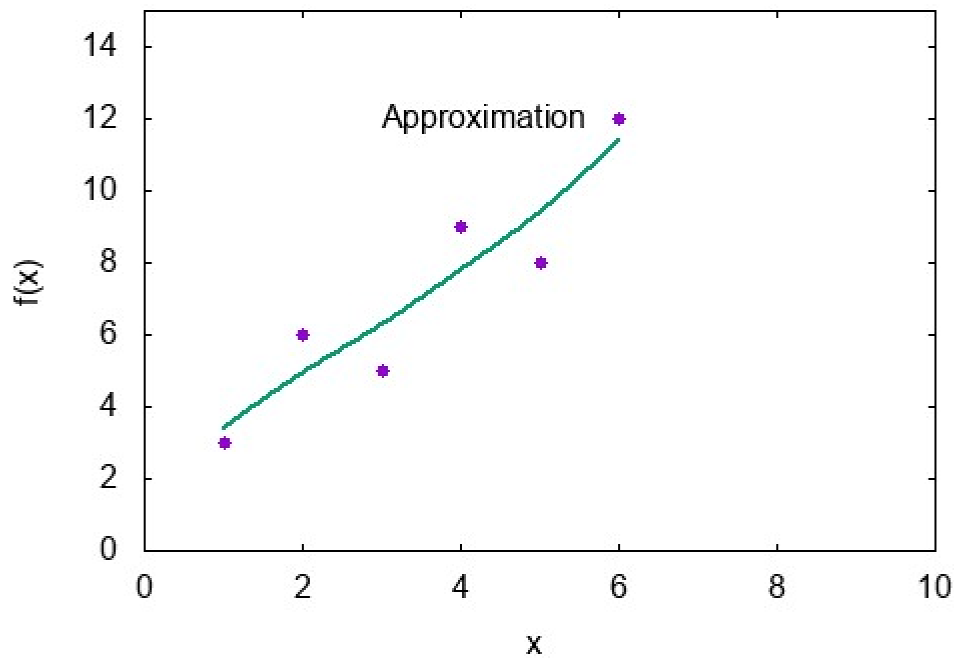

Figure 2: Explanation of Regression

What is the difference between interpolation and regression?

--In function regression (or approximation), we want to find function or polynomial y_i such that for all i, $f(x_i) \approx y_i$ i.e. the curve should not be over fitted and error should be minimum so that we can accurately predict the future values.

--In function interpolation, we want to fit function y_i such that $f(x_i) = yi$ i.e. the curve fits the x/f(x) relationship perfectly and do not need to worry about the error.

Polynomial Interpolation: A unique n**th** order polynomial passes through n points.
Most common form of n_{th} order polynomials are:

1. $P(x) = a_0 + a_1x + a_2x^2 + \ldots\ldots.. + a_{n-1}x^{n-1} + a_nx^n$. … (i), this is known as n_{th} order polynomial in **Power Form**.
2. $P(x) = a_0 + a_1(x-c) + a_2(x-c)^2 + \ldots\ldots.. + a_{n-1}(x-c)^{n-1} + a_n(x-c)^n$ …(ii), this is known as n_{th} order polynomial in **Shifted Power Form**, and

3. $P(x) = a_0 + a_1(x-c1) + a_2(x-c1)(x-c2) + \ldots\ldots.. + a_{n-1}(x-c1)(x-c2)\ldots(x-cn)$ ...(iii), this is known as n^{th} order polynomial in **Newton Form**, which reduces to shifted power form, when $c_1 = c_2 = c_3 = \ldots = c_n$ and reduces to simple power form, when $c_i = 0$ for all i.

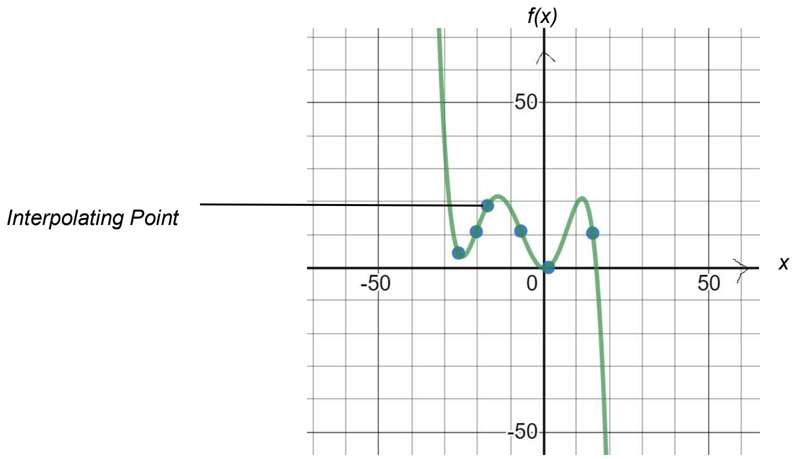

Figure 3: Polynomial Interpolation

- **Spline Interpolation:** Pass different curves (mostly 3**rd** order) through different subsets of the data points.

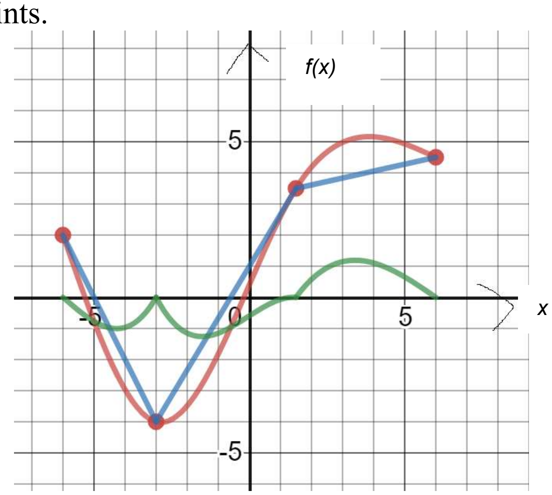

Figure 4: Four Points Cubic Spline Interpolation

- **Application of interpolation:**
  1. Analyzing experimental data
  2. Plotting smooth curves to represent experimental data
  3. Derive many different formulas

- The justification for replacing a given function by a polynomial or by a trigonometric series relies on two theorems proved by Weierstrass (1885).
  1. If $f(x)$ is continuous in an interval (a, b) (i.e. $a \leq x \leq b$), then given any $\varepsilon > 0$, there exists a polynomial $p(x)$ such that

$| f(x) - p(x) | < \varepsilon$ for all x in (a, b)

2. Every continuous function of period 2π can be represented by a finite trigonometric series of the form $g(x) = a_0 + a_1 \sin x + a_2 \sin 2x + \dots\dots\dots\dots + a_n \sin nx + b_1 \cos x + b_2 \cos 2x + \dots\dots\dots b_n \cos nx$, where, $| f(x) - g(x) | < \varepsilon$ for all values of x in the interval considered and ε represents any pre assigned positive quantity.

This means that it is possible to find a polynomial $p(x)$ whose graph remain within the regain bounded by $y = f(x) + \varepsilon$ and $y = f(x) - \varepsilon$ for all x between a and b, however small ε may be.

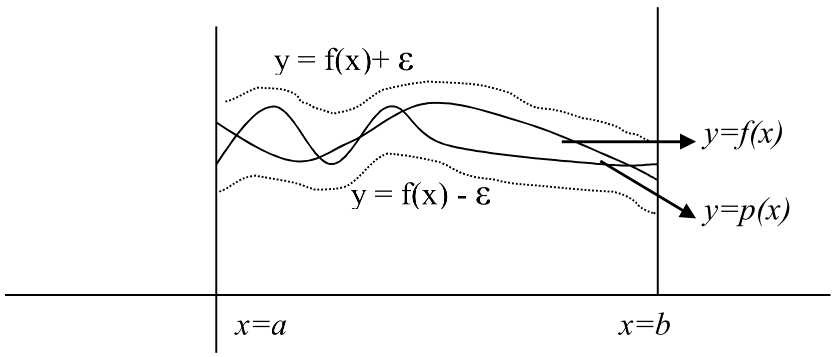

Figure 5: Explanation of Weierstrass Theorem

Various methods of interpolation:

1. Newton Gregory forward and backward interpolation formulas

2. Lagrange Interpolation

3. Newton divide difference formula

The Newton Gregory forward and backward interpolation formula and central difference formula are applicable only when values of x are given at equal intervals.

When the values of independent variable x are given at unequal intervals then we can use Lagrange and New Newton divide difference formulas.

References:

1. BalaGurushamy, E. Numerical Methods. New Delhi : Tata McGraw-Hill, 2000.
2. Steven C.Chapra, Raymon P. Cannale. Numerical Methods for Engineers. New Delhi : Tata McGRAW-HILL, 2003. ISMN 0-07-047437-0.
3. Rao, G. Shankar. Numerical Analysis. New Age International Publisher, 2002. 3rd edition