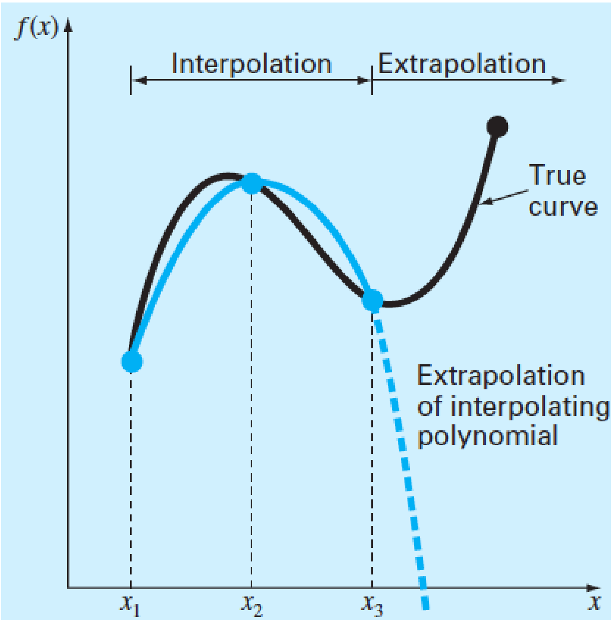
Interpolation and Extrapolation

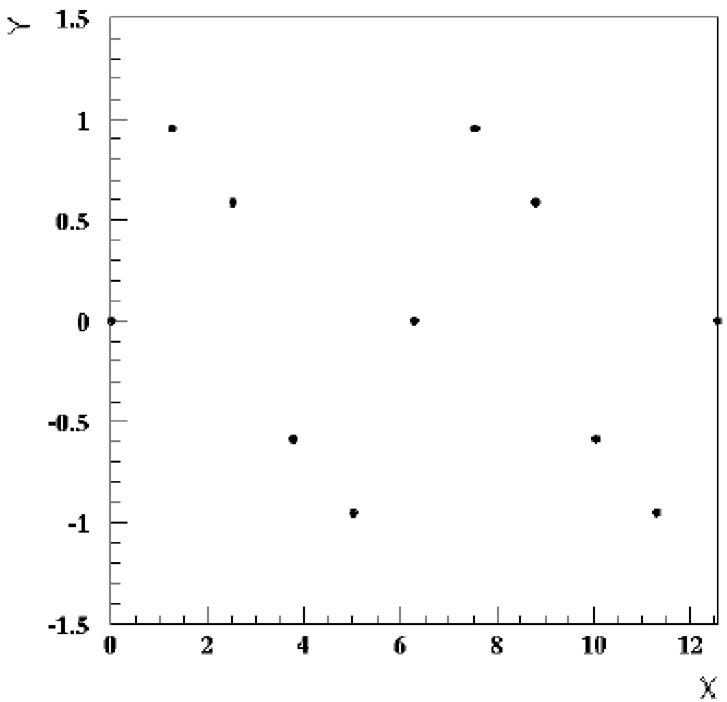
In the framework of the undergraduate course:
"Applied Mathematics"



Prof. Dr. Mohamed Samer
Engineering in Biosystems, Energy and Environment
Department of Agricultural Engineering
Faculty of Agriculture, Cairo University
E-Mails: msamer@agr.cu.edu.eg; samer@cu.edu.eg
Website: http://scholar.cu.edu.eg/samer/biocv

Interpolation, Smoothing, Extrapolation

A typical numerical application is to find a smooth parametrization of available data so that results at intermediate (or extended) positions can be evaluated.



What is a good estimate for y for x=4.5, or x=15 ?

Options: if have a model, y=f(x), then fit the data and extract model parameters. Model then used to give values at other points.
If no model available, then use a smooth function to interpolate

Interpolation, Extrapolation, and Inverse Interpolation

Interpolation is a method of constructing new data points within the range of discrete set of known data points.

In engineering and science one often has a number of data points, are obtained by sampling or experimentation, and try to construct a function which closely fits those data points. This is called “Curve Fitting” or “Regression” analysis. Interpolation is a specific case of curve fitting in which the function must go exactly through the data points.

Interpolation

Interpolation is a useful mathematical and statistical tool used to estimate values between two points. In this lesson, you will learn about this tool, its formula and how to use it.

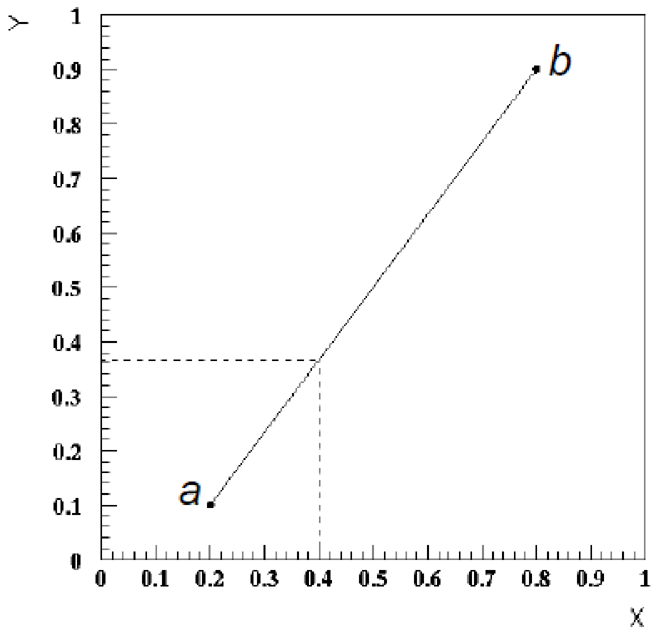
What Is Interpolation?

Interpolation is the process of finding a value between two points on a line or curve. To help us remember what it means, we should think of the first part of the word, 'inter,' as meaning 'enter,' which reminds us to look 'inside' the data we originally had. This tool, interpolation, is not only useful in statistics, but is also useful in science, business or any time there is a need to predict values that fall within two existing data points.

Interpolation

Start with interpolation. Simplest - linear interpolation. Imagine we have two values of *x*, x_a and x_b, and values of *y* at these points, y_a, y_b. Then we interpolate (estimate the value of *y* at an intermediate point) as follows:

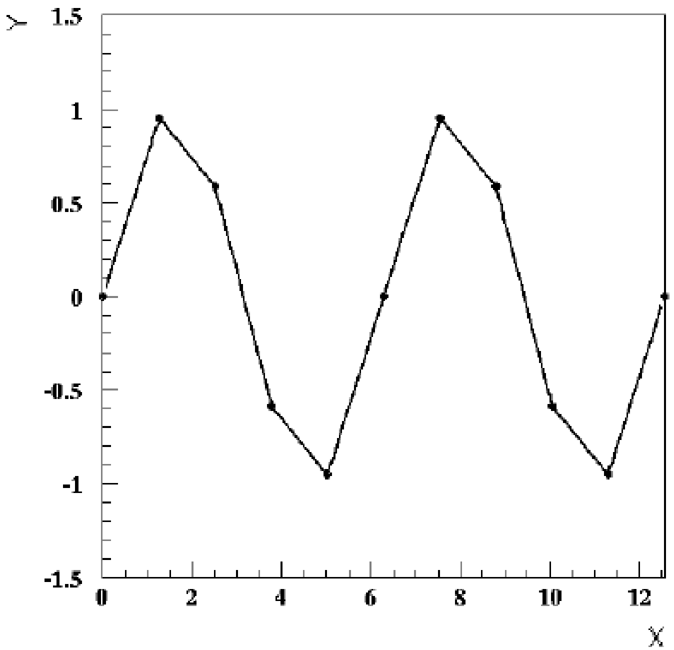
$$y = y_a + \frac{(y_b - y_a)}{(x_b - x_a)}(x - x_a)$$



Interpolation

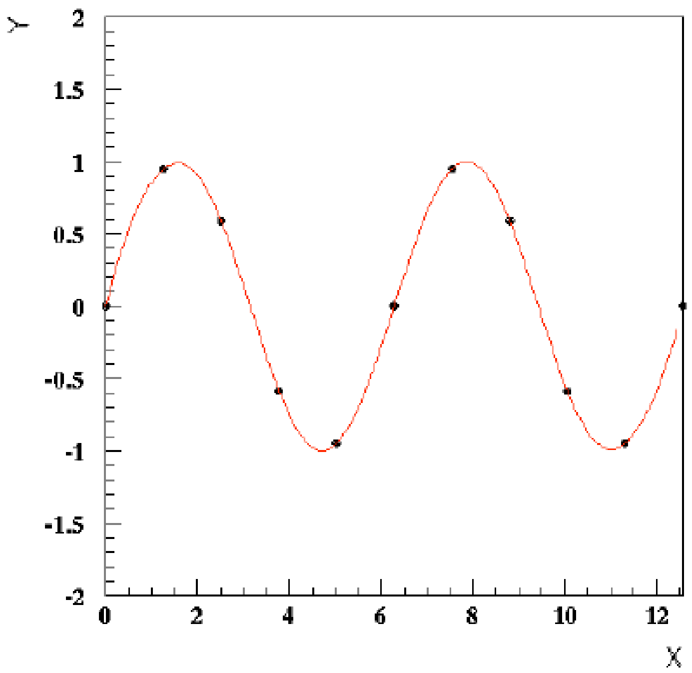
Back to the initial plot:

$$y = y_a + \frac{(y_b - y_a)}{(x_b - x_a)}(x - x_a)$$

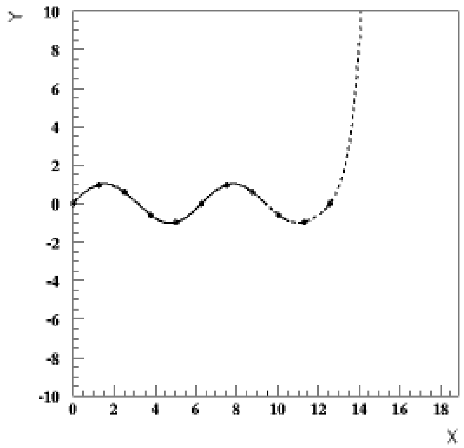


Not very satisfying. Our intuition is that functions should be smooth. Try reproducing with a higher order polynomial. If we have *n+1* points, then we can represent the data with a polynomial of order *n*.

Interpolation



Fit with a 10th order polynomial. We go through every data point (11 free parameters, 11 data points). This gives a smooth representation of the data and indicates that we are dealing with an oscillating function. However, extrapolation is dangerous !



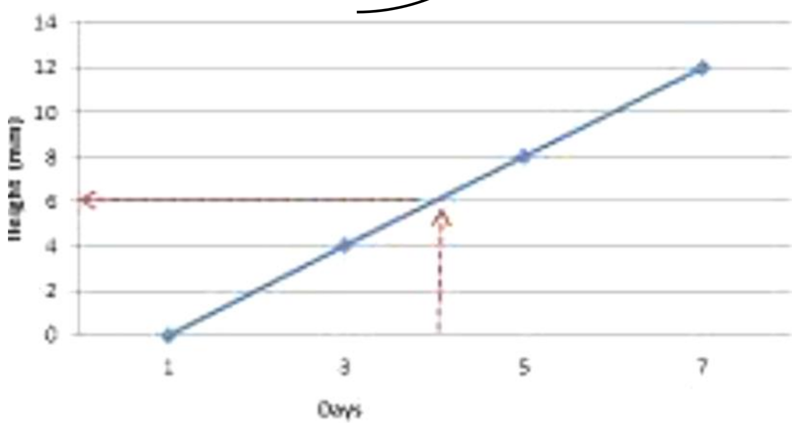
Example

Interpolation Example

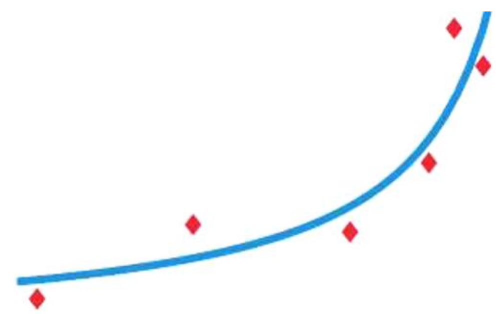
Here's an example that will illustrate the concept of interpolation. A gardener planted a tomato plant and she measured and kept track of its growth every other day. This gardener is a curious person, and she would like to estimate how tall her plant was on the fourth day.

Her table of observations looked like this:

Day	Height (mm)
1	0
3	4
5	8
7	12
9	16



But what if the plant was not growing with a convenient linear pattern? What if its growth looked more like this?



Based on the chart, it's not too difficult to figure out that the plant was probably 6 mm tall on the fourth day. This is because this disciplined tomato plant grew in a linear pattern; there was a linear relationship between the number of days measured and the plant's height growth. **Linear pattern** means the points created a straight line. We could even estimate by plotting the data on a graph.

What would the gardener do in order to make an estimation based on the above curve? Well, that is where the interpolation formula would come in handy.

Interpolation Formula

The interpolation formula looks like this:

$$y - y_1 = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1)$$

Now, let's explore how to use this formula. The two sets of points between which the estimate can be found are:

$$(x_1, y_1) \text{ and } (x_2, y_2)$$

Going back to the tomato plant example, the first set of values for day three are (3,4), the second set of values for day five are (5,8), and the value for x is 4 since we want to find the height, y, on the fourth day. After substituting these values into the formula, calculate the estimated height of the plant on the fourth day.

$$y - y_1 = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1)$$

$$y - 4 = \frac{8 - 4}{5 - 3}(x - 3)$$

$$y - 4 = \frac{4}{2}(x - 3)$$

$$y - 4 = 2(x - 3)$$

$$y = 2(x - 3) + 4$$

$$y = 2(4 - 3) + 4$$

$$y = 2(1) + 4$$

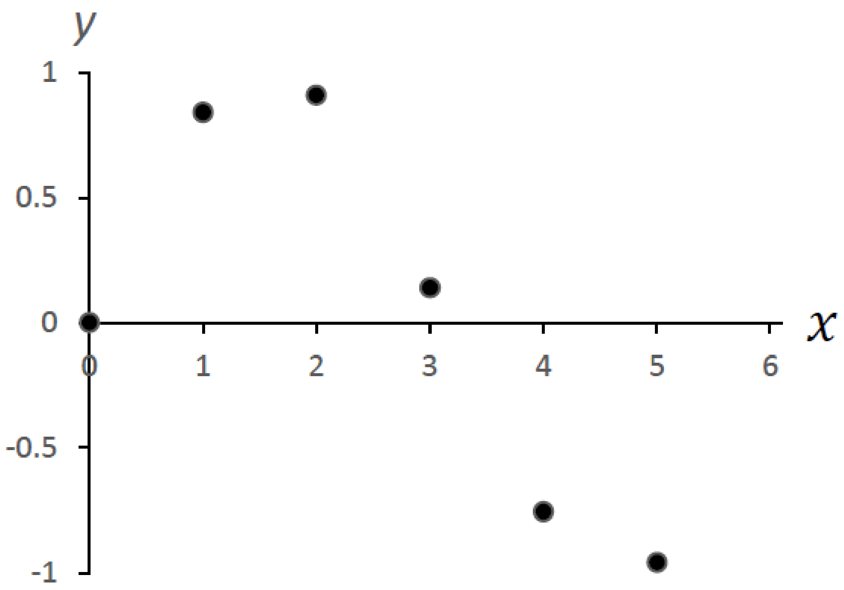
$$y = 6$$

Hence, the estimated height of the plant on the fourth day is 6 mm.

Linear and Polynomial Interpolation

Example: Suppose we have a table which gives some values of an unknown function f.

x	$f(x)$
0	0
1	0.8415
2	0.9093
3	0.1411
4	-0.7568
5	-0.9589
6	-0.2794



Note: There are many different interpolation methods.

1. Linear interpolation

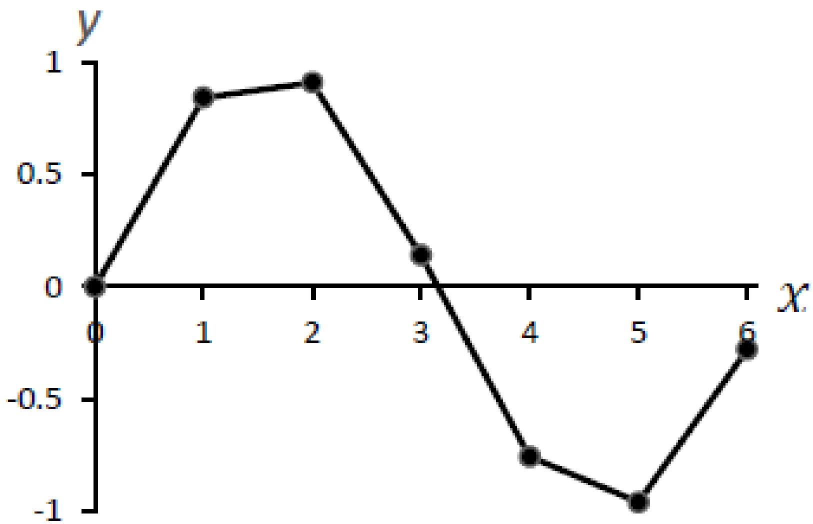
One of the simplest methods is linear interpolation. Consider the above example of determining $f(2.5)$. since 2.5 is midway between 2 and 3. It is reasonable to take $f(2.5)$ midway between $f(2) = 0.9093$ and $f(3) = 0.1411$, which yields 0.5252.

Generally, linear interpolation takes two data points, say (x_a, y_a) and (x_b, y_b), and the interpolante is given by:

$$y = y_0 + (x - x_0)\frac{(y_b - y_a)}{(x_b - x_a)} \quad \text{at the point } (x, y)$$

Linear interpolation is:

- quick, and
- Easy, but
- It is not very precise.
- Another disadvanage is that the interpolation is not differenable at the point x_k.



In words, the error is proportional to the square of the distance between the data points.

2. Polynomial interpolation

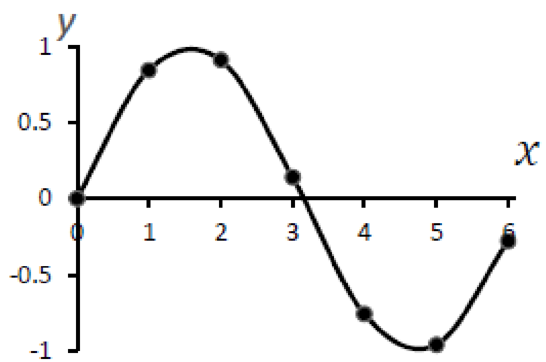
Polynomial interpolation is a generalization of linear interpolation. Note that the linear interpolant is a linear function, we now replace this interpolants by a polynomial of higher degree.

The following six degree polynomial goas through all the seven points:

$$f(x) = -0.00012521x^6 - 0.003130x^5 + 0.07321x^4 - 0.3577x^3 + 0.2255x^2 + 0.9038x$$

Substituting $x = 2.5$, we find that

$$f(2.5) = 0.5965$$



As before, we take the tabulated points to be $y_i \equiv y(x_i)$. If the interpolating polynomial is written as

$$y = c_0 + c_1 x + c_2 x^2 + \cdots + c_{N-1} x^{N-1}$$

then the c_i's are required to satisfy the linear equation

$$\begin{bmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^{N-1} \\ 1 & x_1 & x_1^2 & \cdots & x_1^{N-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{N-1} & x_{N-1}^2 & \cdots & x_{N-1}^{N-1} \end{bmatrix} \cdot \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{N-1} \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{N-1} \end{bmatrix}$$

Generally, if we have n data points, there is exactly one polynomial of degree at most $n-1$ going through all the data points.

The interpolant Error is proportional to the distance between the data points to the power n.

Furthermore, the interpolant is a polynomial and thus infinitely differentiable. We see that polynomial interpolation solves all the problems of linear interpolation.

However, polynomial interpolation also has some disadvantages:

$\Rightarrow$ May not be exact after all, especially at the end points.

$\Rightarrow$ These disadvantages can be avoided by using spline interpolation.

Inverse Interpolation

- So far, when n points were given as $(f(x_i), x_i)$ we tried to determine the value of the function $f(x_k)$ for an intermediate value of the independent variable, x_k
- The inverse problem of this is to determine the value of the independent variable x for given value of f(x).
- Such a process is known as inverse interpolation
- Consider a function f(x)=1/x
- The following table values tabulate some points

x	1	2	3	4	5	6	7
$f(x)$	1	0.5	0.3333	0.25	0.2	0.1667	0.1429

- Now the inverse interpolation can be though of as

$f(x)$	0.1429	0.1667	0.2	0.25	0.3333	0.5	1
x	7	6	5	4	3	2	1

- By rendering the table as an inverse interpolation, i.e changing the axes of plotting
- However, we notice that the spacing on abscissa is not uniform

$f(x)$	0.1429	0.1667	0.2	0.25	0.3333	0.5	1
x	7	6	5	4	3	2	1

- The non-uniform values as independent variables leads to oscillations in interpolating polynomial
- The approach to this is to fit nth order polynomial and then find the solution of this polynomial for given value of f(x)

- For solving the following inverse interpolation problem let us consider 3 points with quadratic polynomial :

$f(x)$	0.1429	0.1667	0.2	0.25	0.3333	0.5	1
x	7	6	5	4	3	2	1

$$f_2(x) = 0.041667x^2 - 0.375x + 1.08333$$

- Now to find x for f(x)=0.3 we need to solve the above quadratic equation

Extrapolation

In Statistics, **Extrapolation** is a process of estimating the value beyond the distinct range of the given variable. The meaning of this term is the process of estimating something if the present situation continues. It is an important concept not only in Mathematics but also in other disciplines like Psychology, Sociology, Statistics etc with some categorical data. Here, we will discuss in detail about the definition, formula and examples of extrapolation.

Extrapolation Definition

Extrapolation is defined as an estimation of a value based on extending the known series or factors beyond the area that is certainly known. In other words, extrapolation is a method in which the data values are considered as points such as $x_1, x_2, \ldots, x_n$. It commonly exists in statistical data very often, if that data is sampled periodically and it approximates the next data point. One such example is when you are driving, you usually **extrapolate** about the road conditions beyond your sight.

What is Extrapolation Statistics?

Extrapolation is a statistical method beamed at understanding the unknown data from the known data. It tries to predict future data based on historical data. For example, estimating the size of a population a few years in the future based on the current population size and its rate of growth.

Extrapolation Methods

Linear Extrapolation

For any linear function, linear extrapolation provides a good result when the point to be predicted is not too far from the given data. It is usually done by drawing the tangent line at the endpoint of the given graph and that will be extended beyond the limit.

Polynomial Extrapolation

A polynomial curve can be created with the help of entire known data or near the endpoints. This method is typically done by means of Lagrange interpolation or Newton's system of finite series that provides the data. The final polynomial is used to extrapolate the data using the associated endpoints.

Conic Extrapolation

A conic section can be created with the help of five points nearer to the end of the given data. The conic section will curve back on itself if it is a circle or ellipse. But for parabola or hyperbola, the curve will not back on itself because it is relative to X-axis.

Extrapolation Formula

Let us consider the two endpoints in a linear graph (x_1, y_1) and (x_2, y_2) where the value of the point “x” is to be extrapolated, then the extrapolation formula is given as

$$y(x)=y_1+\{[(x-x_1)/(x_2-x_1)]\times(y_2-y_1)\}$$

Extrapolation Example

Question:

The two given points that lie on the straight line is (1, 5) and (4, 10). Determine the value of y at x = 5 on the straight line using a linear extrapolation method.

Solution:

Given: $x_1 = 1$,$y_1 = 5$

and $x_2 = 4$, $y_2 = 10$

The linear extrapolation formula is given as;

$$y(x)=y_1+\{[(x-x_1)/(x_2-x_1)]\times(y_2-y_1)\}$$

Substitute the known values in the given formula,

y(5) = 5 +((5-1)/(4-1))(10-5)

y(5) = 5 + (4/3) (5)

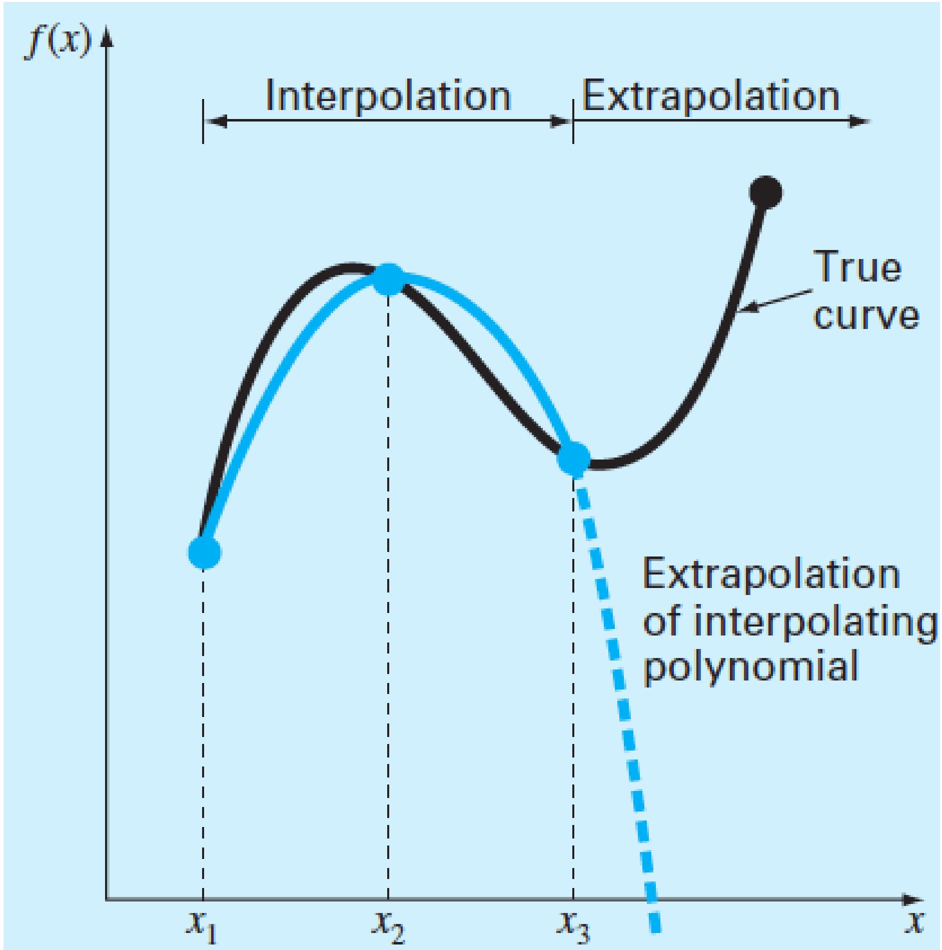
y(5) = 5 + 6.65

y(5)= 11.65

Therefore y(5) = 11.65

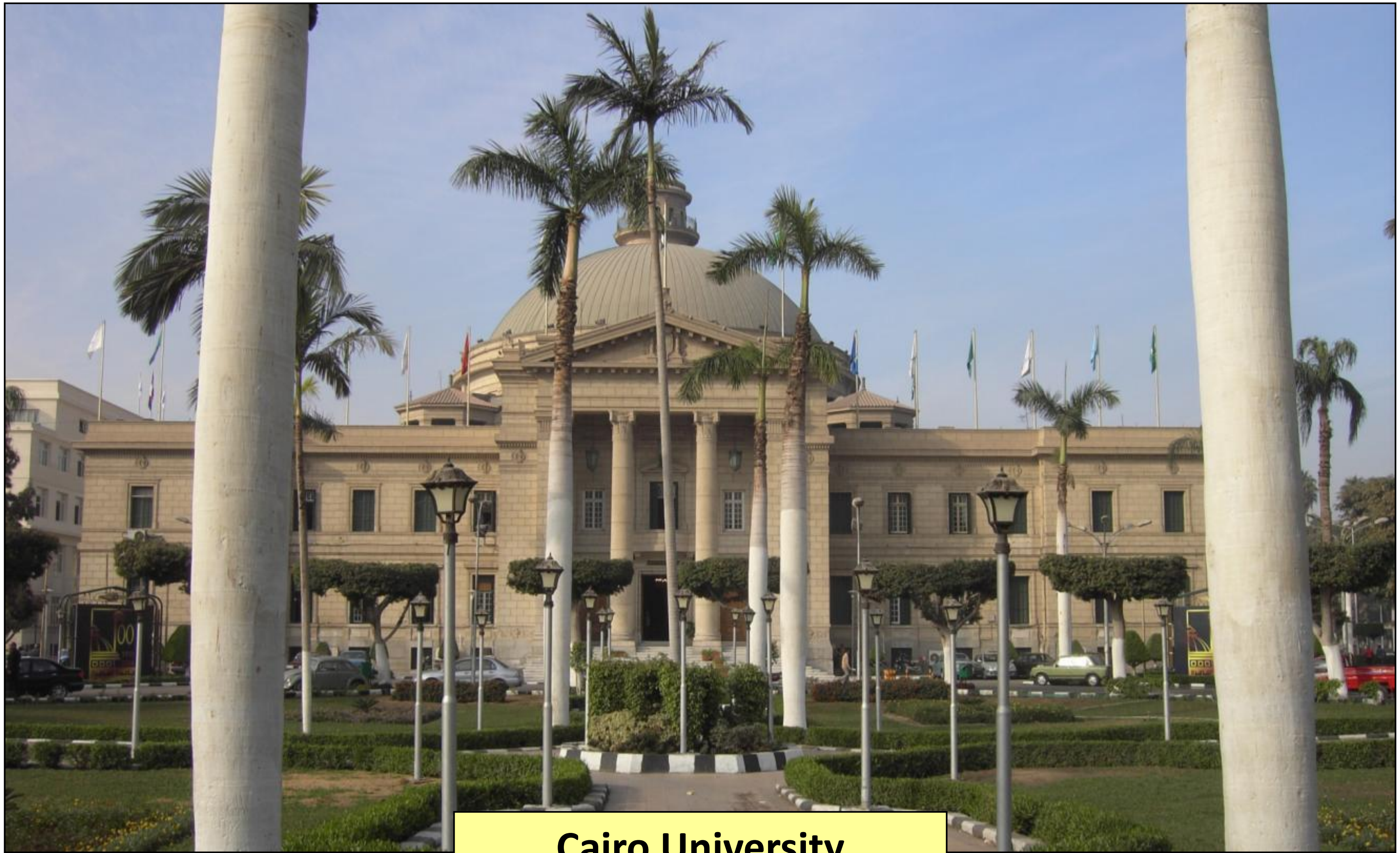
Extrapolation is the process of estimating a value of f(x) that lies outside the range of the known base points, $x_1, x_2, \ldots, x_n$.

The process of extrapolation represents a step into the unknown because the process extends the curve beyond the known region.
As such, the true curve could easily diverge from the prediction.
Therefore we need to be careful whenever a case arises where one must extrapolate.



Exercises and Solutions

Thank You!



Cairo University