Part 5. Curve Fitting
Chapter 17. Least Square Regression

Lecture 17

Linearization of Nonlinear Relationships & Nonlinear Regression Models

17.1.5,17.1.6. 17.2, 17.5

Homeyra Pourmohammadali

Nonlinear Regression

Some popular nonlinear regression models:

Exponential model: $$y = ae^{bx}$$

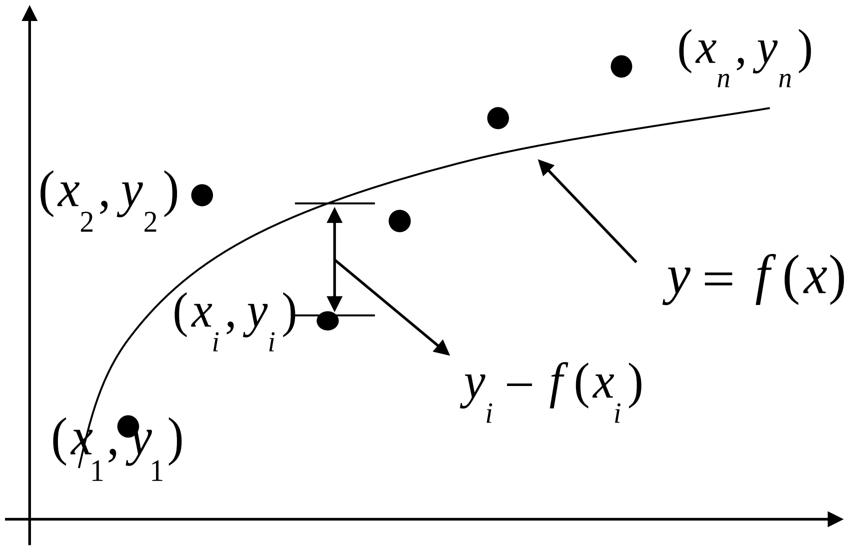
Power model: $$y = ax^{b}$$

Saturation growth model: $$y = \frac{ax}{b + x}$$

Polynomial model: $$y = a_0 + a_1 x + \ldots + a_m x^{m}$$

Nonlinear Regression

Given n data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ best fit $y = f(x)$ to the data, where $f(x)$ is a nonlinear function of x.



Nonlinear regression model for discrete y vs. x data

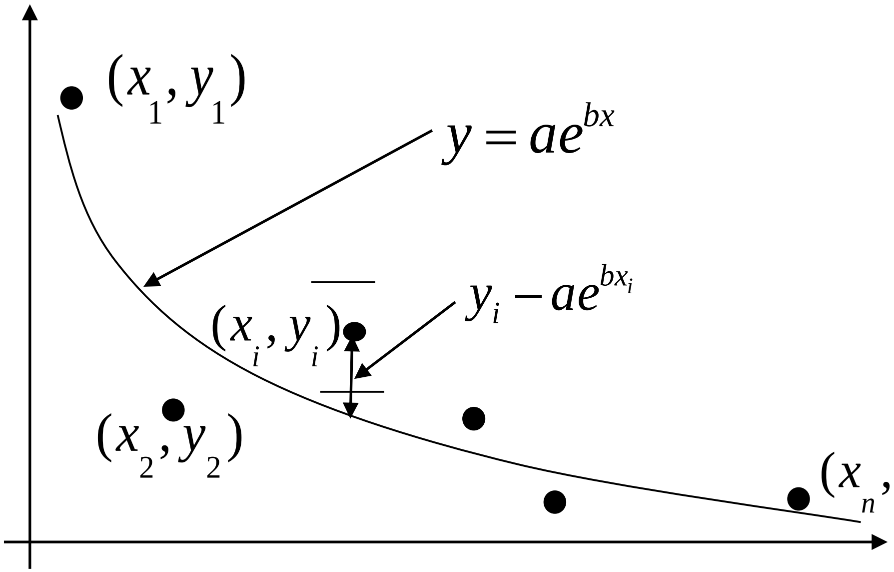
Regression

Exponential Model

$$y = ae^{bx}$$

Exponential Model

To best fit $y = ae^{bx}$ to the data.



Exponential model of nonlinear regression for y vs. x data

Exponential Model

Two methods to solve for constants of $y = ae^{bx}$

1. Linearize the nonlinear equation

   - take *ln* from both sides of equation

   $ln\, y = ln\, a + bx$

   $Y = A + b\,X \quad , \quad A = ln\, a, \quad Y = ln\, y$

2. Directly find constants by minimizing S_r

Constants of Exponential Model

The sum of the square of the residuals is defined as

$$S_r = \sum_{i=1}^{n} \left(y_i - ae^{bx_i} \right)^2$$

Differentiate with respect to *a* and *b*

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

$$\frac{\partial S_r}{\partial b} = \sum_{i=1}^{n} 2 \left(y_i - ae^{bx_i} \right) \left(-ax_i e^{bx_i} \right) = 0$$

Constants of Exponential Model

Rewriting the equations, we obtain

$$-\sum_{i=1}^{n} y_i e^{bx_i} + a\sum_{i=1}^{n} e^{2bx_i} = 0$$

$$\sum_{i=1}^{n} y_i x_i e^{bx_i} - a\sum_{i=1}^{n} x_i e^{2bx_i} = 0$$

Constants of Exponential Model

Solving the first equation for ***a*** yields

$$a = \frac{\sum_{i=1}^{n} y_i e^{bx_i}}{\sum_{i=1}^{n} e^{2bx_i}}$$

Substituting a back into the previous equation

$$\sum_{i=1}^{n} y_i x_i e^{bx_i} - \frac{\sum_{i=1}^{n} y_i e^{bx_i}}{\sum_{i=1}^{n} e^{2bx_i}} \sum_{i=1}^{n} x_i e^{2bx_i} = 0$$

The constant ***b*** can be found through numerical methods such as bisection method.

Example 1. Nonlinear Regression of Exponential Equation.

Find constants of $y = ae^{bx}$ regression model using below data set and estimate y at $x = 24$.

x	0	1	3	5	7	9
y	1.000	0.891	0.708	0.562	0.447	0.355

$$\sum_{i=1}^{n} y_i x_i e^{bx_i} - \frac{\sum_{i=1}^{n} y_i e^{bx_i}}{\sum_{i=1}^{n} e^{2bx_i}} \sum_{i=1}^{n} x_i e^{2bx_i} = 0$$

Can use MATLAB to solve for b

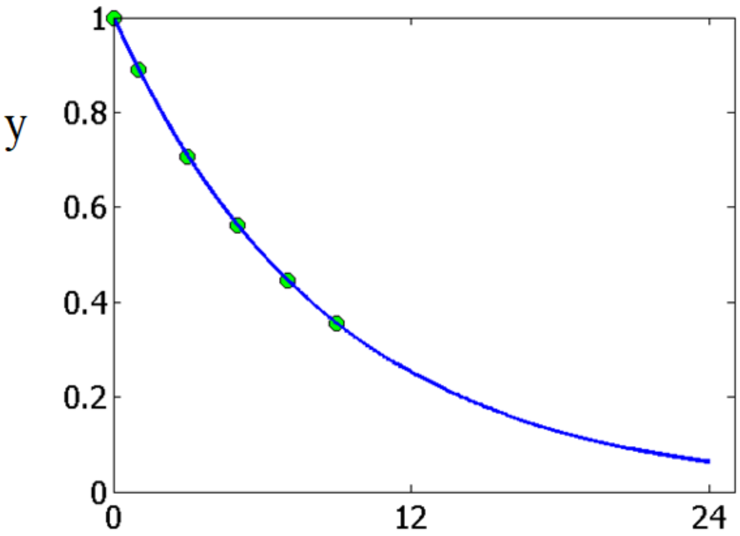
$$b = -0.1151$$

Example 1. Nonlinear Regression of Exponential Equation.

$$a = \frac{\sum_{i=1}^{n} y_i e^{bx_i}}{\sum_{i=1}^{n} e^{2bx_i}} = 0.9998$$

Exponential regression model:

$$y = 0.9998\, e^{-0.1151x}$$



Example 1. Nonlinear Regression of Exponential Equation.

Estimate of *y* at *x=24*:

$$y = 0.9998 \times e^{-0.1151(24)}$$

$$= 6.3160 \times 10^{-2}$$

Regression

Power Model

$$y = ax^b$$

Power Model

1. Linearize the nonlinear equation

   - take *log* from both sides of equation

   $\log y = \log a + b \log x$

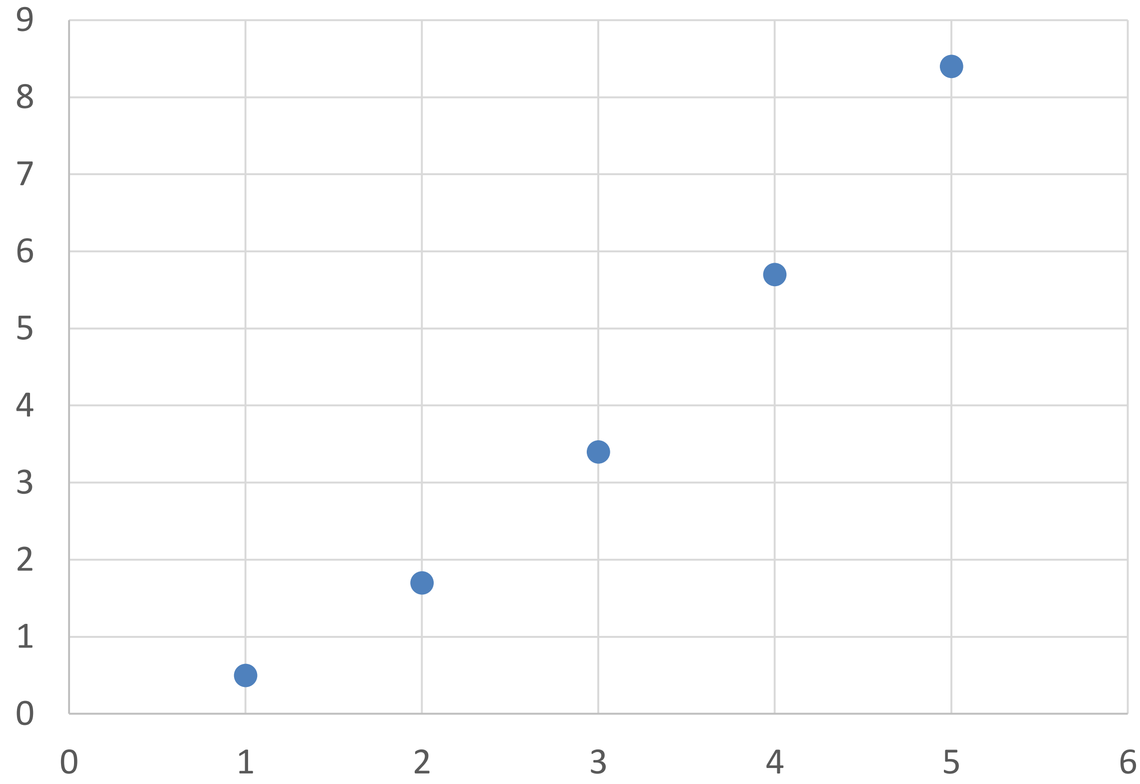
   $Y = A + b X \quad , \quad A = \log a, \quad Y = \log y$

As a homework, try finding a and b by minimizing S_r

Example 2. Nonlinear Regression of Power Equation.

Find constants of $y = ax^b$ regression model using below data set.

x	y
1	0.5
2	1.7
3	3.4
4	5.7
5	8.4



Example 2. Nonlinear Regression of Power Equation.

$$\log y = \log a + b \log x$$

$$Y = A + b\,X$$

$$b = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2}$$

$$A = \bar{y} - a_1 \bar{x}$$

x	y	log(x)	log(y)
		X	Y
1	0.5	0	-0.301
2	1.7	0.301	0.230
3	3.4	0.477	0.531
4	5.7	0.602	0.756
5	8.4	0.699	0.924

A = - 0.3 , b = 1.75

$a = 10^{A} = 10^{-0.3} = 0.5$ → $y = ax^{b}$ → $y = 0.5\,x^{1.75}$

Regression

Saturation Growth Model

$$y = \frac{ax}{b + x}$$

Saturation Growth Model

1. Linearize the nonlinear equation $$y = \frac{ax}{b+x}$$

   - Inverse both sides of equation

   $1/y = (1/a)\,[\,(b/x) + 1\,]$

   $1/y = (1/a) + (b/a)(1/x)$

   $Y = A + B\,X$

As a homework, try finding a and b by minimizing S_r

Regression

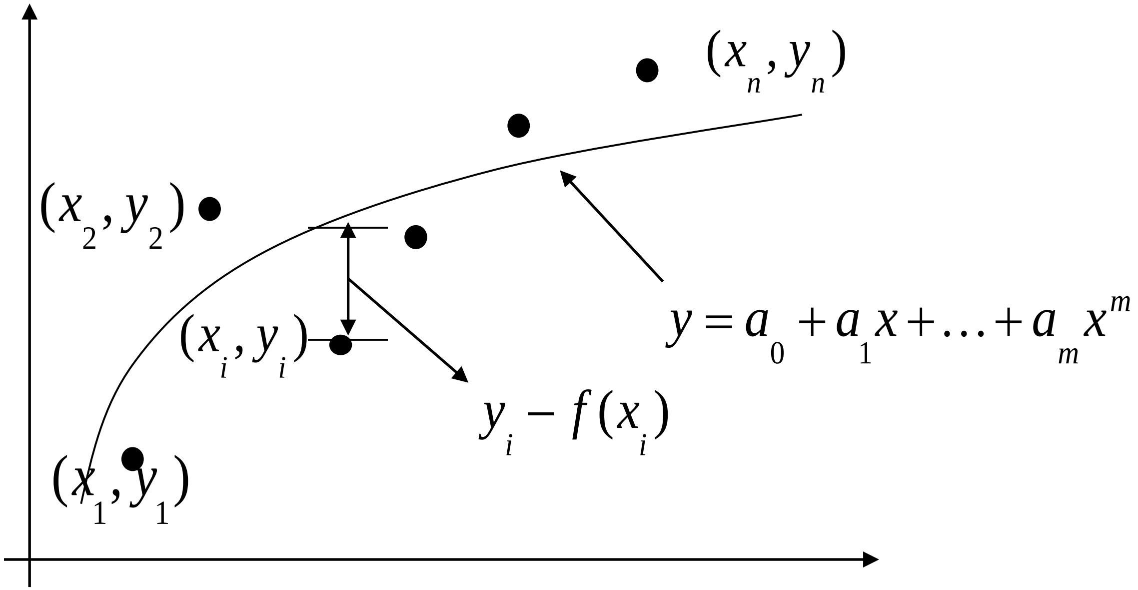
Polynomial Model

$$y = a_0 + a_1 x + \ldots + a_m x^m$$

Polynomial Model

To best fit $y = a_0 + a_1 x + \ldots + a_m x^m$ $\quad (m \le n-2)$

to a given data set.



Polynomial model for nonlinear regression of y vs. x data

Polynomial Model

The residual at each data point is given by

$$e_i = y_i - a_0 - a_1 x_i - \ldots - a_m x_i^m$$

The sum of the square of the residuals then is

$$S_r = \sum_{i=1}^{n} e_i^2$$

$$= \sum_{i=1}^{n} \left(y_i - a_0 - a_1 x_i - \ldots - a_m x_i^m \right)^2$$

Polynomial Model

Finding constants by minimizing S_r

$$\frac{\partial S_r}{\partial a_0} = \sum_{i=1}^{n} 2.\left(y_i - a_0 - a_1 x_i - \ldots - a_m x_i^m\right)(-1) = 0$$

$$\frac{\partial S_r}{\partial a_1} = \sum_{i=1}^{n} 2.\left(y_i - a_0 - a_1 x_i - \ldots - a_m x_i^m\right)(-x_i) = 0$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

$$\frac{\partial S_r}{\partial a_m} = \sum_{i=1}^{n} 2.\left(y_i - a_0 - a_1 x_i - \ldots - a_m x_i^m\right)(-x_i^m) = 0$$

Polynomial Model

These equations in matrix form are given by

$$\begin{bmatrix} n & \left(\sum_{i=1}^{n} x_i\right) & . & . & . & \left(\sum_{i=1}^{n} x_i^m\right) \\ \left(\sum_{i=1}^{n} x_i\right) & \left(\sum_{i=1}^{n} x_i^2\right) & . & . & . & \left(\sum_{i=1}^{n} x_i^{m+1}\right) \\ . & . & . & . & . & . \\ \left(\sum_{i=1}^{n} x_i^m\right) & \left(\sum_{i=1}^{n} x_i^{m+1}\right) & . & . & . & \left(\sum_{i=1}^{n} x_i^{2m}\right) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ . \; . \; . \\ a_m \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \\ . \; . \; . \\ \sum_{i=1}^{n} x_i^m y_i \end{bmatrix}$$

The above equations are then solved for $a_0, a_1, \ldots, a_m$

2^{nd}–Order Polynomial Model

$$y = a_0 + a_1 x + a_2 x^2$$

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - a_0 - a_1 x_i - a_2 x_2^2 \right)^2$$

As a homework, try finding a and b by minimizing S_r

Solve 3 equations for 3 unknown constants

3^{rd}–Order Polynomial Model

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3$$

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - a_0 - a_1 x_i - a_2 x_2^2 - a_2 x_2^2 - a_3 x_3^3 \right)^2$$

As a homework, try finding a and b by minimizing S_r

Solve 4 equations for 4 unknown constants

$$r^2 = \frac{S_t - S_r}{S_t}$$

Error for Polynomial Regression

Use general form of correlation coefficient

$$r^2 = \frac{S_t - S_r}{S_t}$$

S_t: total sum of the squares around the mean for the dependent variable

S_r: sum of the squares of residuals around the regression line is S_r

Part 5. Curve Fitting
Chapter 18. Interpolation

Lecture 18

Linear, Quadratic, Cubic (Spline) Interpolation
18.1, 18.3,18.6

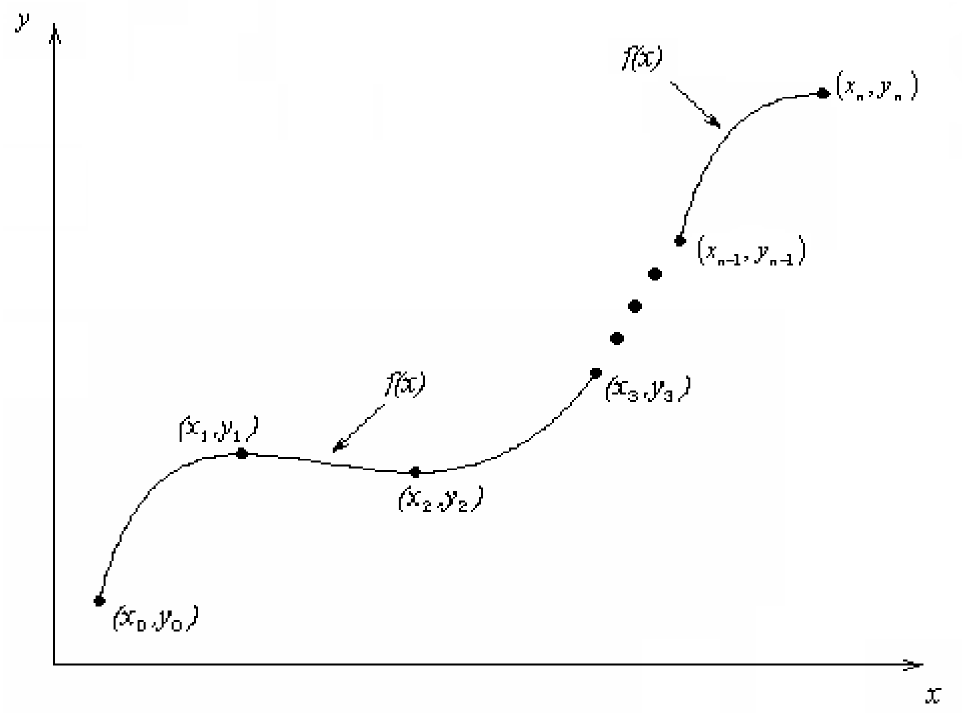
Homeyra Pourmohammadali

Interpolation

Given (x_0,y_0), (x_1,y_1), …… (x_n,y_n), find the value of 'y' at a value of 'x' that is not given.

Interpolants

Polynomials are the most common choice of interpolants because they are easy to evaluate, differentiate and integrate.



Interpolation Using Polynomials

- The most common method is using n^{th}-order polynomial:

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

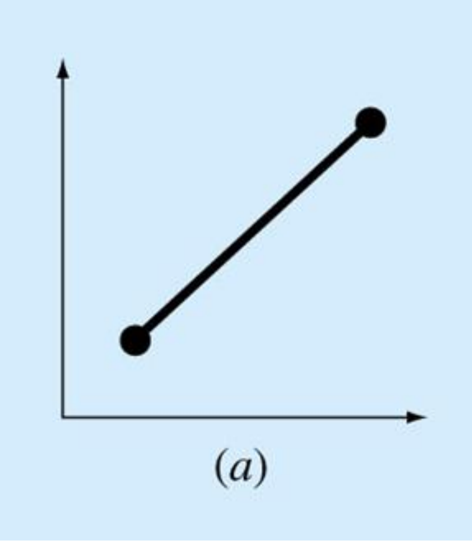
where a_0, a_1,…. a_n are real constants.

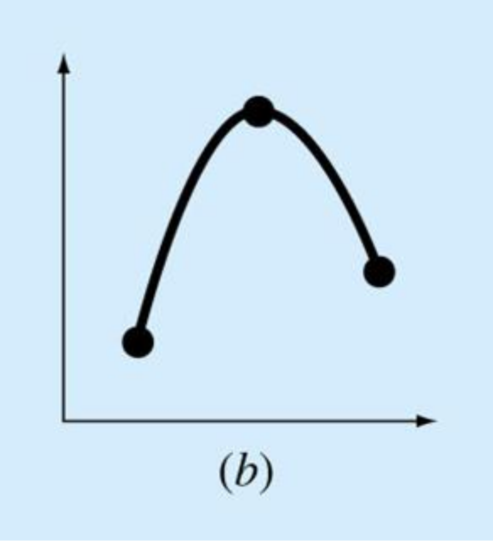
- There is one and only one n^{th}-order polynomial that fits ***n+1*** points

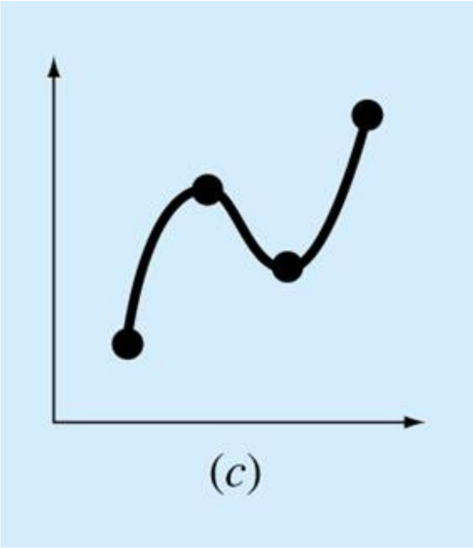
- Set up ‘n+1’ equations to find ‘n+1’ constants.

- To find the value ‘y’ at a given value of ‘x’, simply substitute the value of ‘x’ in the above polynomial.

Interpolating Polynomials







1st-order Polynomial

(linear interpolation)

Connecting 2 points

2nd -order Polynomial

(quadratic or parabolic interpolation)

Connecting 3 points

3rd- Order Polynomial

(cubic or spline interpolation)

Connecting 4 points

Linear Interpolation (1st-Order Polynomial)

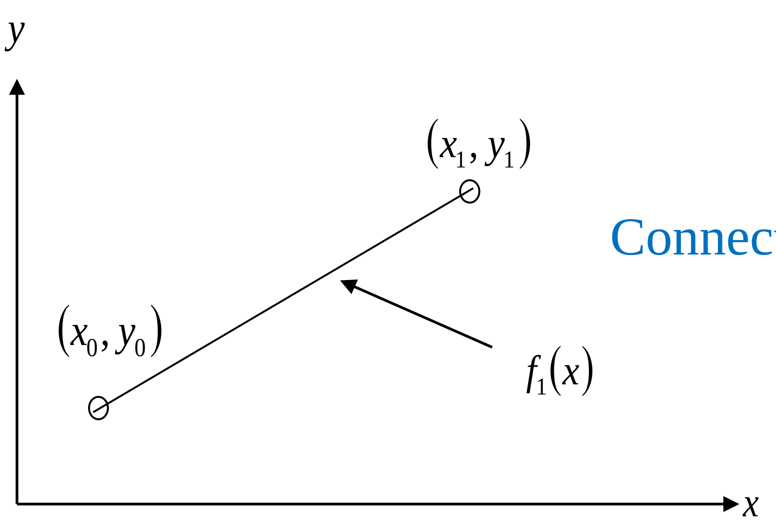
$$f(x) = a_0 + a_1 x$$

Example 1. Direct Method for Linear Interpolation (1st Order Polynomial)

Velocity as a function of time.

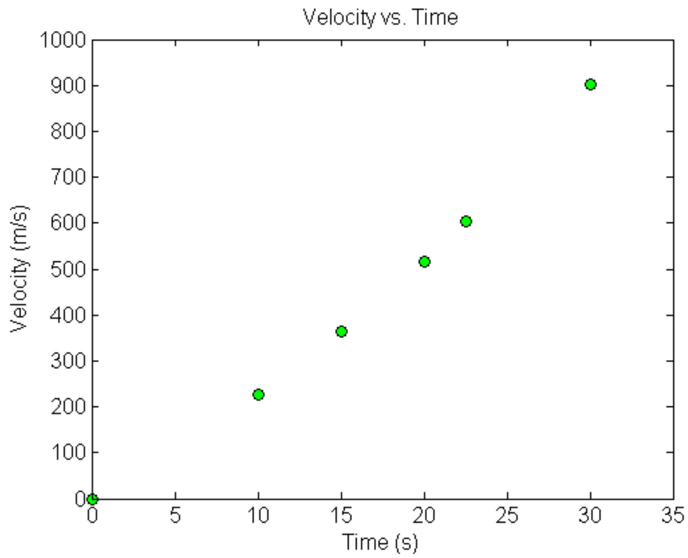
$t,(\text{s})$	$v(t),(\text{m/s})$
0	0
10	227.04
15	362.78
20	517.35
22.5	602.97
30	901.67

Find the velocity at *t =16* seconds, using the given data points for upward velocity of a rocket is given as a function of time.



Linear interpolation.

Connecting 2 points



Estimate using this line $$v(t) = a_0 + a_1 t$$

Example 1. Direct Method for Linear Interpolation (1st Order Polynomial)

$$v(t) = a_0 + a_1 t$$

Step 1. Choose 2 data points that are closest to *t = 16 s* that also bracket *t =16 s*.

Step 2. Evaluate function (velocity) at these 2 points (times) to find a_0 and a_1

t, (s)	$v(t)$, (m/s)
0	0
10	227.04
15	362.78
20	517.35
22.5	602.97
30	901.67

$$v(15) = a_0 + a_1(15) = 362.78$$

$$v(20) = a_0 + a_1(20) = 517.35$$

$$a_0 = -100.93 \qquad a_1 = 30.914$$

Step 3: Set equation of interpolant

$$v(t) = -100.93 + 30.914t,\ \ 15 \le t \le 20.$$

Step 4. Evaluate velocity at t =16 s $\quad v(16) = -100.93 + 30.914(16) = 393.7 \text{ m/s}$

Quadratic (Parabolic) Interpolation (2nd-Order Polynomial)

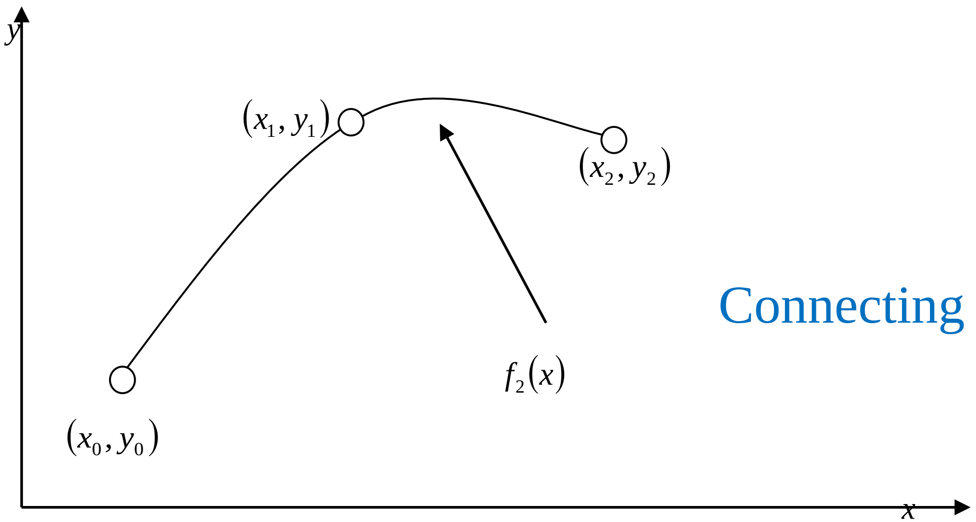
$$f(x) = a_0 + a_1 x + a_2 x^2$$

Example 2. Direct Method for Quadratic Interpolation (2^{nd}-Order Polynomial)

Velocity as a function of time.

$t, (\mathrm{s})$	$v(t), (\mathrm{m/s})$
0	0
10	227.04
15	362.78
20	517.35
22.5	602.97
30	901.67

Find the velocity at $t = 16$ seconds, using the given data points for upward velocity of a rocket is given as a function of time.



Quadratic interpolation.

Connecting 3 points

Estimate using 2^{nd} degree polynomial

$$v(t) = a_0 + a_1 t + a_2 t^2$$

Example 2. Direct Method for Quadratic Interpolation (2nd-Order Polynomial)

$$v(t) = a_0 + a_1 t$$

Step 1. Choose 3 data points that are closest to $t = 16$ s that also bracket $t = 16$ s.

t, (s)	$v(t)$, (m/s)
0	0
10	227.04
15	362.78
20	517.35
22.5	602.97
30	901.67

Step 2. Evaluate function (velocity) at these 3 points (times) to find a_0, a_1 and a_2

$$v(10) = a_0 + a_1(10) + a_2(10)^2 = 227.04$$

$$v(15) = a_0 + a_1(15) + a_2(15)^2 = 362.78$$

$$v(20) = a_0 + a_1(20) + a_2(20)^2 = 517.35$$

$$a_0 = 12.05$$

$$a_1 = 17.733$$

$$a_2 = 0.3766$$

Step 3: Set equation of interpolant

$$v(t) = 12.05 + 17.733t + 0.3766t^2, \quad 10 \le t \le 20$$

Step 4. Evaluate velocity at t =16 s

$$v(16) = 12.05 + 17.733(16) + 0.3766(16)^2$$

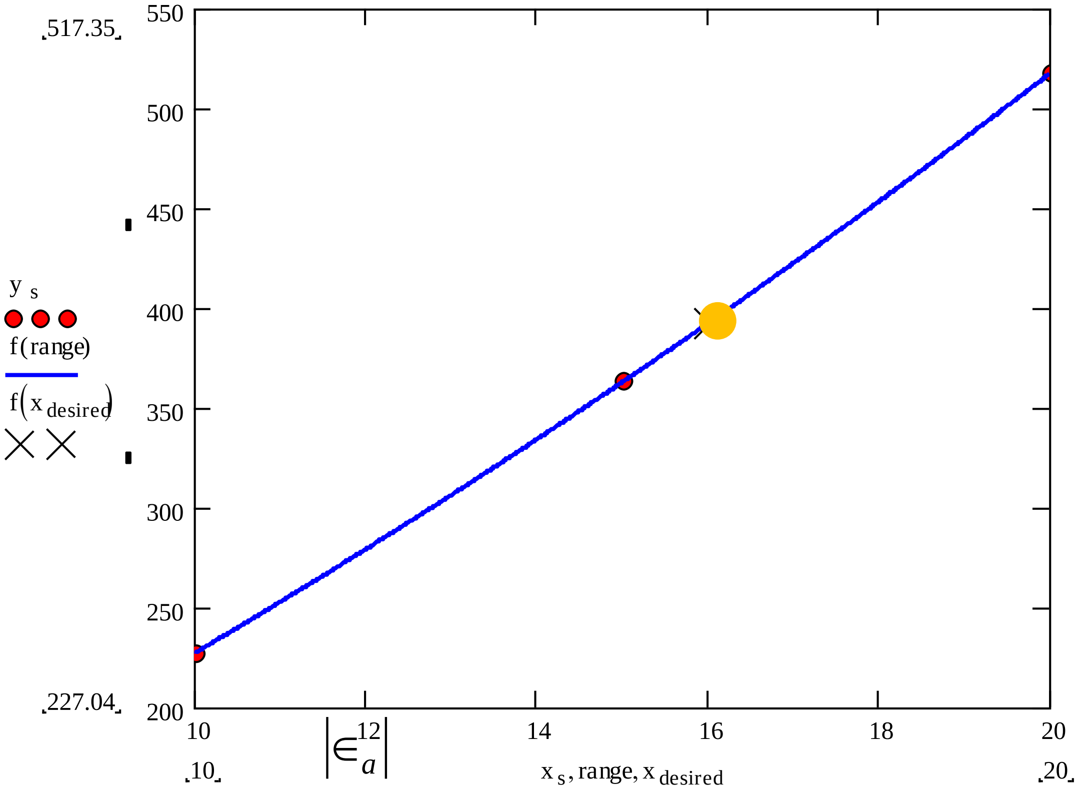
$$= 392.19 \text{ m/s}$$

Example 2. Direct Method for Quadratic Interpolation (2nd-Order Polynomial)

Absolute relative approximate error between the results from the 1st and 2nd order polynomial is

$$\left|\in_a\right| = \left|\frac{392.19 - 393.70}{392.19}\right| \times 100$$

$$= 0.38410\%$$



Cubic (Spline) Interpolation (3rd-Order Polynomial)

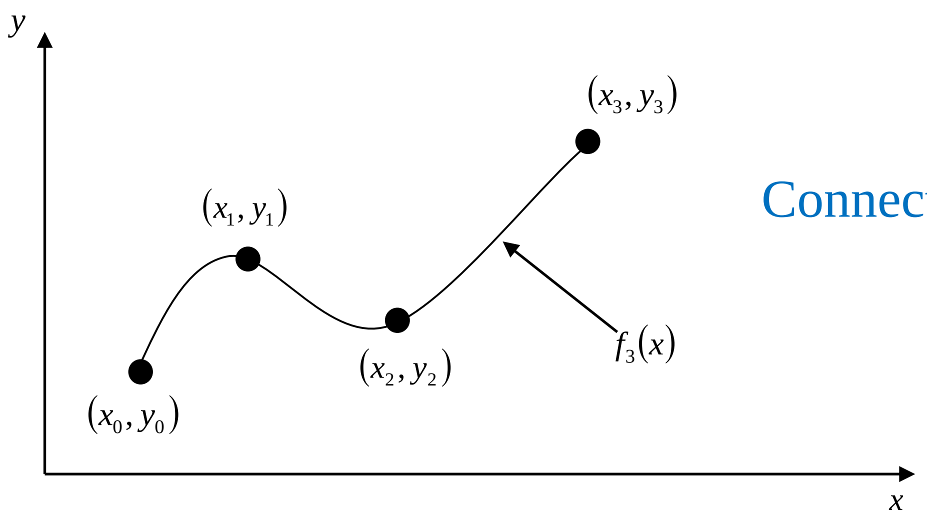
$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3$$

Example 3. Direct Method for Cubic (Spline) Interpolation (3rd-Order Polynomial)

Velocity as a function of time.

t, (s)	$v(t)$, (m/s)
0	0
10	227.04
15	362.78
20	517.35
22.5	602.97
30	901.67

Find the velocity at $t = 16$ seconds, using the given data points for upward velocity of a rocket is given as a function of time.



Connecting 4 points

Cubic (Spline) interpolation.

Estimate using 2nd degree polynomial

$$v(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3$$

Example 3. Direct Method for Cubic (Spline) Interpolation (3rd-Order Polynomial)

$$v(t) = a_0 + a_1 t$$

Step 1. Choose 4 data points that are closest to *t = 16 s* that also bracket *t =16 s*.

t, (s)	$v(t)$, (m/s)
0	0
10	227.04
15	362.78
20	517.35
22.5	602.97
30	901.67

Step 2. Evaluate function (velocity) at these 4 points (times) to find a_0, a_1 , a_2 and a_3

$$v(10) = 227.04 = a_0 + a_1(10) + a_2(10)^2 + a_3(10)^3$$

$$v(15) = 362.78 = a_0 + a_1(15) + a_2(15)^2 + a_3(15)^3$$

$$v(20) = 517.35 = a_0 + a_1(20) + a_2(20)^2 + a_3(20)^3$$

$$v(22.5) = 602.97 = a_0 + a_1(22.5) + a_2(22.5)^2 + a_3(22.5)^3$$

$$a_0 = -4.2540$$

$$a_1 = 21.266$$

$$a_2 = 0.13204$$

$$a_3 = 0.0054347$$

Step 3: Set equation of interpolant

$$v(t) = -4.2540 + 21.266t + 0.13204t^2 + 0.0054347t^3, \quad 10 \le t \le 22.5$$

Step 4. Evaluate velocity at t =16 s

$$v(16) = -4.2540 + 21.266(16) + 0.13204(16)^2 + 0.0054347(16)^3$$

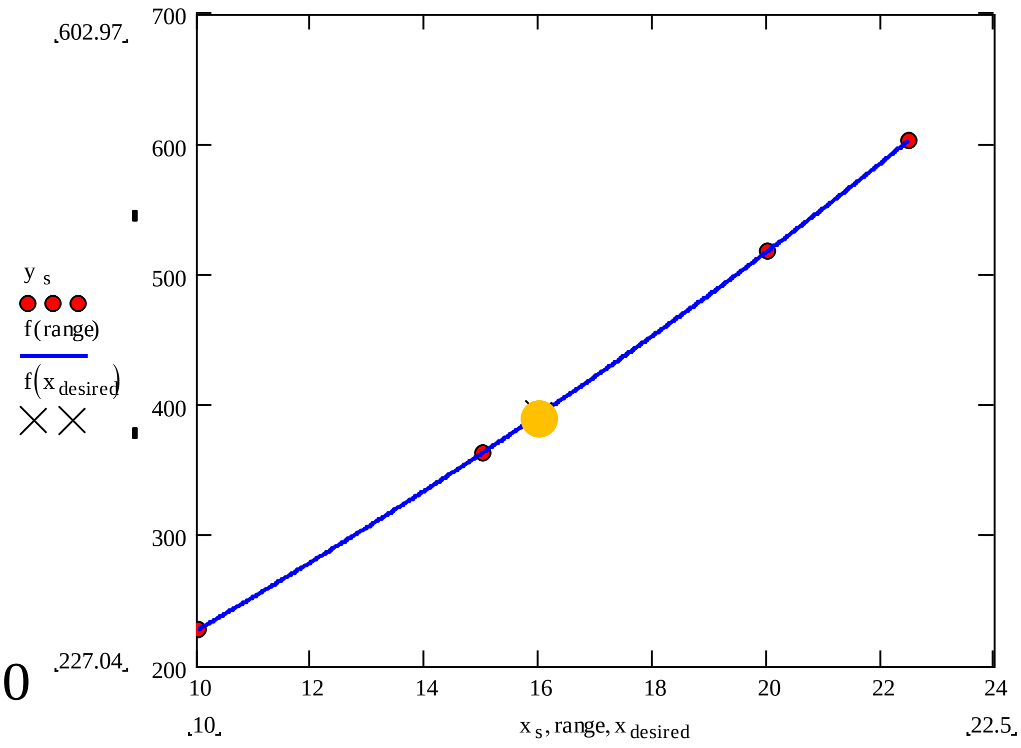
$$= 392.06 \text{ m/s}$$

Example 3. Direct Method for Cubic (Spline) Interpolation (3rd-Order Polynomial)

Absolute relative approximate error between the results from the 2nd- and 3rd-order polynomial is

$$\left|\in_a\right| = \left|\frac{392.06 - 392.19}{392.06}\right| \times 100$$

$$= 0.033269\%$$



Comparison of Different Orders of Polynomials

$$f(x) = a_0 + a_1 x$$

$$f(x) = a_0 + a_1 x + a_2 x^2$$

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3$$

Comparison of Different Orders of Polynomials

Order of Polynomial	1	2	3
$v(t = 16)$ m/s	393.7	392.19	392.06
Absolute Relative Approximate Error	----------	0.38410 %	0.033269 %

t(s)	v (m/s)
0	0
10	227.04
15	362.78
20	517.35
22.5	602.97
30	901.67

Notes

- Sometimes polynomials can lead to erroneous results because of round off error and overshoot.

- Apply lower-order polynomials to subsets of data points. Such connecting polynomials are called spline functions.