Mathcad Lecture #8 In-class Worksheet
Curve Fitting and Interpolation

At the end of this lecture, you will be able to:

- explain the difference between *curve fitting* and *interpolation*
- decide whether curve fitting *or* interpolation should be used for a particular application
- interpolate values between data points using linterp and interp with cspline.
- fit equations using line, regress, linfit, and genfit.

1. Interpolation and Curve Fitting Defined

Background

Thermophysical and transport property data are often needed in engineering calculations. Quantities such as he capacities, thermal conductivities, diffusivities, densities, and viscosities are measured experimentally and the results are found in handbooks and journal articles. Usually, the data are not published for every condition that may be experienced in real life problems. For example, vapor pressures might be published for at 310 K and 320 K, but you may need the value at 315 K. In such situations, two options are available: **interpolating** and **curve fitting**.

Interpolation is determining a new point between two existing points. For example, if I measured the heat capacity to be 10 J/mol/K at 300K and 20 J/mol/K at 310 K, I might reasonably estimate the heat capacity at 305 K to be 15 J/mol/K.

Curve Fitting is determining a function that closely matches the data. For example, the data may look like a straight line, so we can determine the slope and intercept that gives a line that matches the data the best. The Antoine Equation is a familiar function that "fits" the temperature dependence of vapor pressures well.

A few notes:

1. Interpolation can really be thought of as a special case of curve fitting where the function is forced to pass through every data point.
2. Interpolation is generally done linearly or with cubic splines. Cubic splines means a third-order polynomial is generated connecting the points rather than a straight line.
3. **Extrapolation** is defined as predicting values *beyond the range of the available data*. For example, vapor pressures were measured at 310 K and 320 K, I could interpolate to obtain a value for a value between 310 K and 320 K, but I would be *extrapolating* if I used the same points to obtain a value for 305 K or 325 K. Thus, the *uncertainties* are greater for extrapolated values compared to interpolated values.
4. <u>You generally use curve fitting when you know a function that matches the data well. When you don't know a function, you interpolate. (Functions for fitting can come from simple observation, experience, trial and error, or theory.</u>

2. Interpolating

Demonstration

Consider the following time/temperature measured in an experiment data. I would like to know the values of the temperature at all times, not just the whole minutes.

$$\mathrm{Time} := \begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{pmatrix} \mathrm{min} \qquad \mathrm{Temp} := \begin{pmatrix} 298 \\ 300 \\ 300 \\ 305 \\ 305 \\ 310 \\ 312 \\ 318 \\ 317 \\ 322 \end{pmatrix} \cdot \mathrm{K}$$

Linear Interpolation

Linear interpolation is done by using the linterp function which has three arguments. It takes the form: $y_{new} := \text{linterp}(xdata,ydata,x_{new})$.

For single values.

$temp_{new} := \text{linterp}(Time, Temp, 4.5min)$ $temp_{new} = 307.5\ K$

Can create functions

$Temp\_linterp(t) := \text{linterp}(Time, Temp, t)$ $Temp\_linterp(4.5min) = 307.5\ K$

Key Points:
1. The last argument of linterp is the value of x at which you want to estimate y.
2. The last argument can be a variable if a function is defined. This allows you to use the interpolation over and over which is useful for graphing or integrating data.

Cubic Spline Interpolation

Instead of using a line to interpolate between data points, a cubic polynomial may be used to connect the points. This is done by using the interp and cspline functions.

For single values.

$temp_{new} := \text{interp}(\text{cspline}(Time, Temp), Time, Temp, 4.5min)$ $temp_{new} = 307.335\ K$

Can create functions

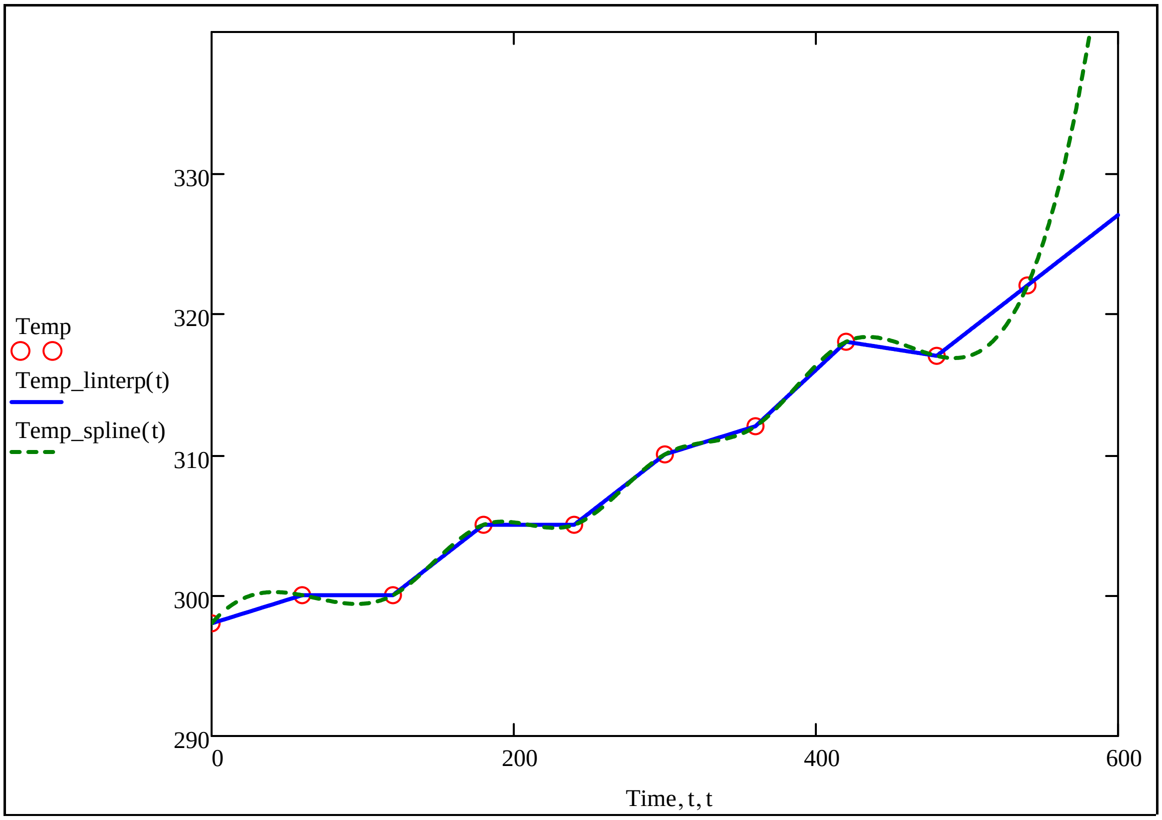
$Temp\_spline(t) := \text{interp}(\text{cspline}(Time, Temp), Time, Temp, t)$ $Temp\_spline(4.5min) = 307.335\ K$

Alternatively, you can first calculate cspline and then feed the results to interp.

$vs := \text{cspline}(Time, Temp)$

$Temp\_spline2(t) := \text{interp}(vs, Time, Temp, t)$ $Temp\_spline2(4.5min) = 307.335\ K$

Visualization



3. Curve Fitting

Background

Mathcad has several utilities to fit data to curves. These are:

- line (or slope and intercept)
- regress
- linfit
- genfit

Each utility can fit certain types of equations. line will return the slope and intercept of the linear relationship that best fit the data. regress is used to return the polynomial that best represents the data. linfit may be used to fit data to expressions which are *linear combinations* of functions. genfit may be used to fit data to functions of any type.

Each fitting method can fit the functions of those above it in the list. For example, regress can do the same thing that line can because a line is just a 1st order polynomial. Since a line is also a linear combination of functions, linfit can also be used. Genfit can be used in any case.

3.1 Fitting Data to a Line

Syntax

line(xdata, ydata)

Demonstration

Determine the best fit line to the time/temperature data used above.

$$\text{Time} = \begin{array}{|c|r|} \hline & 0 \\ \hline 0 & 0 \\ \hline 1 & 60 \\ \hline 2 & 120 \\ \hline 3 & 180 \\ \hline 4 & 240 \\ \hline 5 & 300 \\ \hline 6 & 360 \\ \hline 7 & 420 \\ \hline 8 & 480 \\ \hline 9 & 540 \\ \hline \end{array}\ \text{s} \qquad \text{Temp} = \begin{array}{|c|r|} \hline & 0 \\ \hline 0 & 298 \\ \hline 1 & 300 \\ \hline 2 & 300 \\ \hline 3 & 305 \\ \hline 4 & 305 \\ \hline 5 & 310 \\ \hline 6 & 312 \\ \hline 7 & 318 \\ \hline 8 & 317 \\ \hline 9 & 322 \\ \hline \end{array}\ \text{K}$$

Use the line function to get the slope and intercept.

$$\text{line}\left(\frac{\text{Time}}{\text{min}}, \frac{\text{Temp}}{\text{K}}\right) = \begin{pmatrix} 296.4 \\ 2.733 \end{pmatrix}$$

intercept K

Slope K/min

Question: What are the units on the intercept and the slope?

Key Points:

1. The arguments to the line function must be dimensionless.
2. You must determine the units yourself.

Alternate Solution: Use slope and intercept functions.

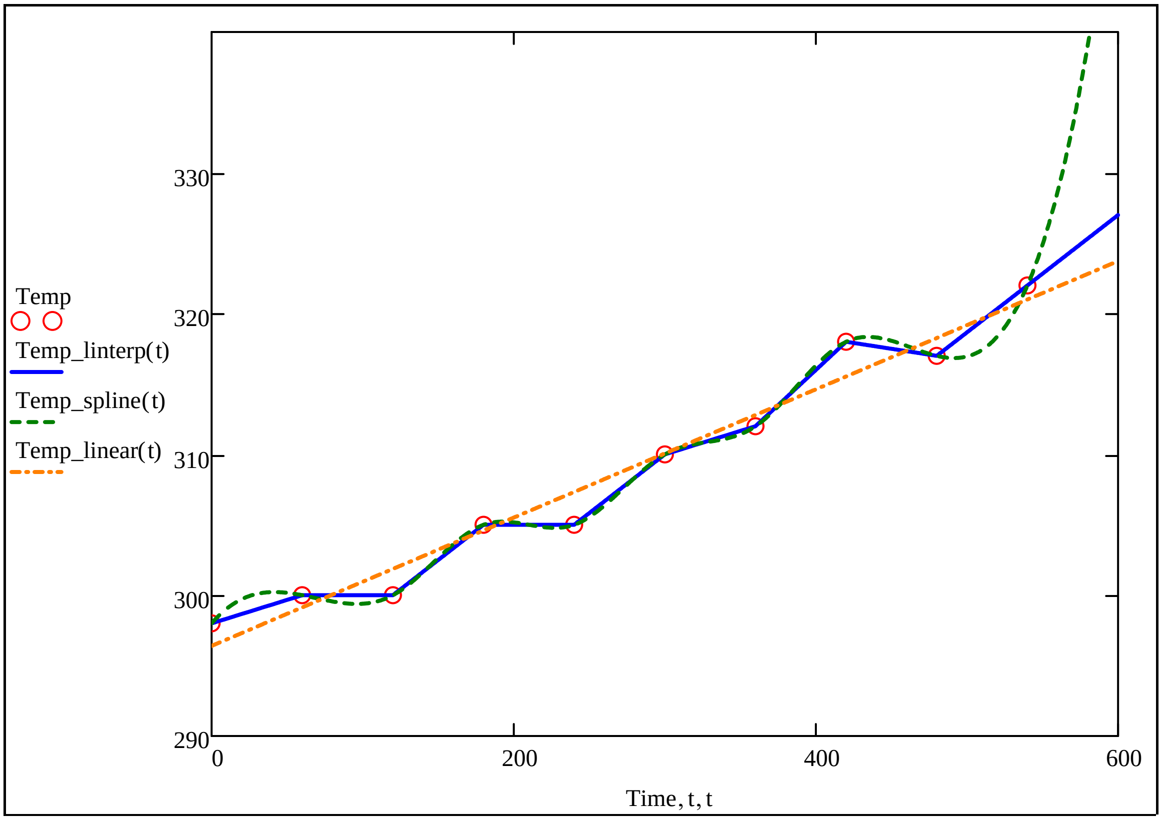
$$\text{slope}(\text{Time}, \text{Temp}) = 2.733 \cdot \frac{\text{K}}{\text{min}} \qquad \text{intercept}(\text{Time}, \text{Temp}) = 296.4\ \text{K}$$

Key Point: You don't need to remove the units on the input data if you use slope and intercept instead of line.

Can also create a function to use later.

$$\text{Temp\_linear}(t) := \text{slope}(\text{Time}, \text{Temp}) \cdot t + \text{intercept}(\text{Time}, \text{Temp})$$

Visualization



3.2 Fitting Data to a Polynomial

Syntax

vs := regress(xdata,ydata,order_of_polynomial)   interp(vs,xdata,ydata,xnew)

Demonstration

The heat capacity of water as a function of temperatures has been measured in a calorimeter. The data are found in the matrices below. Fit the data to a fourth order polynomial.

D :=

	0	1
0	273.15	$7.617\cdot10^4$
1	293.15	$7.536\cdot10^4$
2	313.15	$7.534\cdot10^4$
3	333.15	$7.536\cdot10^4$
4	353.15	$7.556\cdot10^4$
5	373.15	$7.595\cdot10^4$
6	393.15	$7.653\cdot10^4$
7	413.15	$7.725\cdot10^4$
8	433.15	$7.813\cdot10^4$
9	453.15	$7.928\cdot10^4$

kmol := 1000mol

$Temp_{exp} := D^{\langle 0\rangle}\cdot K \quad Cp_{exp} := D^{\langle 1\rangle}\cdot\frac{J}{kmol\cdot K}$

The regress function will give the coefficients of each term in the polynomial.

$$vs := \text{regress}\left(\frac{Temp_{exp}}{K}, \frac{Cp_{exp}}{\frac{J}{mol \cdot K}}, 4\right)$$

$$vs = \begin{pmatrix} 3 \\ 3 \\ 4 \\ 276.37 \\ -2.09 \\ 8.125 \times 10^{-3} \\ -1.412 \times 10^{-5} \\ 9.37 \times 10^{-9} \end{pmatrix}$$

Key Point: The output of regress is a matrix. The first three rows are NOT the coefficients, they are values needed if interp is going to be used. The fourth row is the coefficient on the x^0 term (the constant) The fifth row is the coefficient on the x^1 term. The last row is the coefficient on the x^n term.

Usually, you feed the output of regress to the interp function.

$$Cp_{regress}(t) := \text{interp}\left(vs, \frac{Temp_{exp}}{K}, \frac{Cp_{exp}}{\frac{J}{mol \cdot K}}, t\right)$$

$$Cp_{regress}(350) = 75.528$$

Question: What are the units on the temperature and heat capacity?

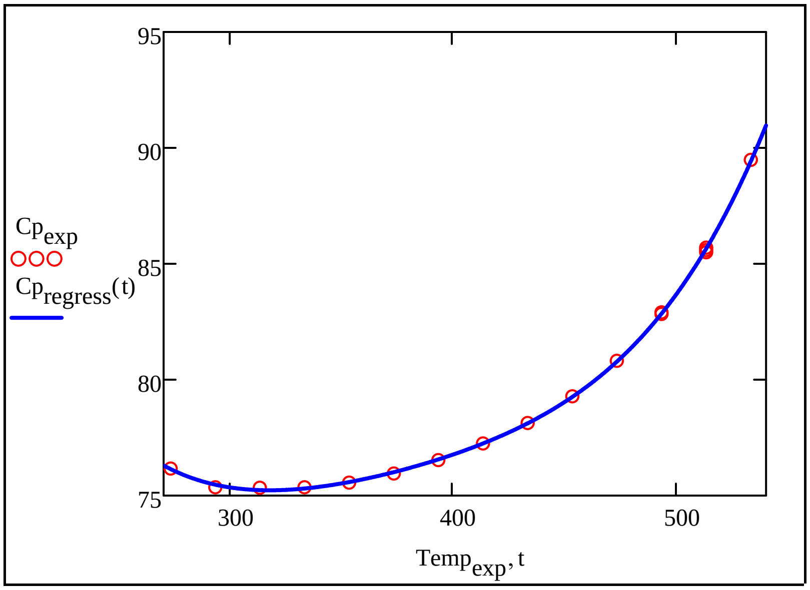
Key Point: Using regress with interp means you cannot include units in the function.

Can create your own function to use units.

$$Cp_{regress}(t) := \left[vs_3 + vs_4 \cdot \frac{t}{K} + vs_5 \cdot \left(\frac{t}{K}\right)^2 + vs_6 \cdot \left(\frac{t}{K}\right)^3 + vs_7 \cdot \left(\frac{t}{K}\right)^4\right] \cdot \frac{J}{mol \cdot K}$$

$$Cp_{regress}(350K) = 75.528 \cdot \frac{J}{mol \cdot K}$$

Check the Results



3.3 Fitting Data to a Linear Combination of Functions

Often, the expression we wish to use to fit the data is not a polynomial but a linear combination of functions. A linear combination of functions is one where each term only has one fitting parameter. For example,

$$y = A + B\cdot\sin(x) + C\cdot\exp(x)$$

is a linear combination of functions. Only one fitting parameter is found in each term and they are multiplying the term. The following are NOT linear combinations of functions.

$$y = A + \sin(B\cdot x) + C\cdot\exp(x) \qquad y = A + B\cdot\sin(x + C) + D\cdot\exp\left(\frac{x}{E}\right)$$

In the first expression, B is inside the sine function. Thus, the B dependence is not linear. In the second equation, multiple fitting parameters appear in the last two terms.

Syntax

$$F(x) = \begin{pmatrix} \text{term1} \\ \text{term2} \\ \vdots \\ \text{term\_n} \end{pmatrix} \qquad \beta = \text{linfit}(\text{xdata}, \text{ydata}, F)$$

The terms in the F(x) matrix are the functional forms of each term without the coefficient. For example, if the function to fits is

$$y = A + B\cdot\sin(x) + C\cdot\exp(x)$$

term1 would be 1, term2 would be $\sin(x)$ and term3 would be $\exp(x)$.

Demonstration

Experimental data for the liquid density of acetone is found in the matrices below. Fit this data to the following expression.

$$Y = A + B\cdot\left(1 - \frac{T}{T_C}\right)^{0.35} + C\cdot\left(1 - \frac{T}{T_C}\right)^{\frac{2}{3}} + D\cdot\left(1 - \frac{T}{T_C}\right) + E\cdot\left(1 - \frac{T}{T_C}\right)^{\frac{4}{3}}$$

The critical temperature of acetone, Tc, is 508.2 K.

D :=

	0	1
0	183.15	15.679
1	198.15	15.406
2	216.97	15.062
3	223.15	14.943
4	233.15	14.752
5	234.79	14.736
6	243.15	14.561
7	253.15	14.37
8	255.7	14.345
9	263.15	14.18

$T_C := 508.2$

$Temp_{exp} := D^{\langle 0 \rangle}\cdot K$ $\qquad \rho_{exp} := D^{\langle 1 \rangle}\cdot\frac{kmol}{m^3}$

Obtain the coefficients to the equation using linfit.

$$F(x) := \begin{bmatrix} 1 \\ \left(1 - \frac{x}{T_C}\right)^{0.35} \\ \left(1 - \frac{x}{T_C}\right)^{\frac{2}{3}} \\ \left(1 - \frac{x}{T_C}\right) \\ \left(1 - \frac{x}{T_C}\right)^{\frac{4}{3}} \end{bmatrix}$$

$$\beta := \mathrm{linfit}\left(\frac{Temp_{exp}}{K}, \frac{\rho_{exp}}{\frac{kmol}{m^3}}, F\right)$$

$$\beta = \begin{pmatrix} 5.671 \\ -1.316 \\ 36.436 \\ -44.882 \\ 23.221 \end{pmatrix}$$

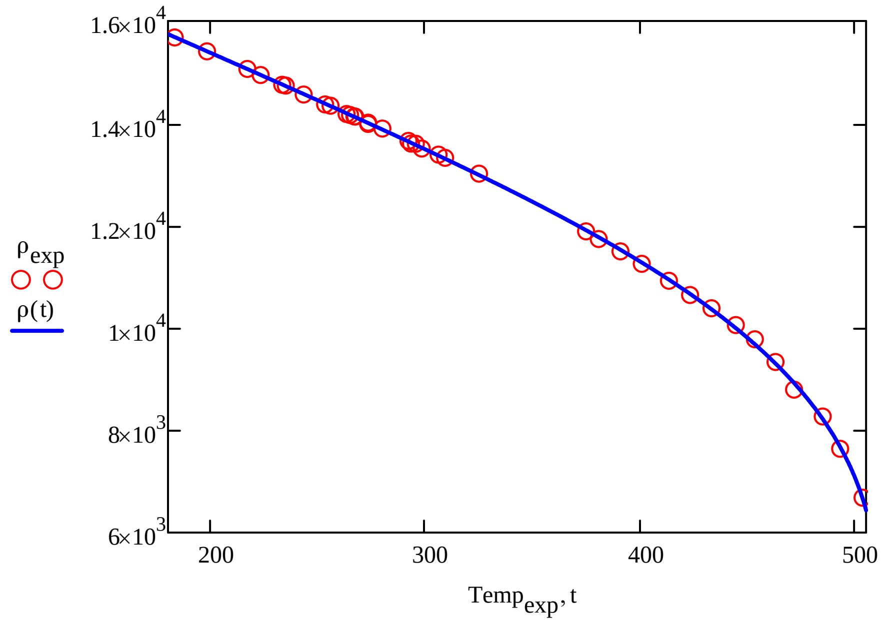
Key Point: Linfit does not like units on its input data so divide them out.

Create a function with the results of linfit.

$$\rho(t) := \left[\beta_0 + \beta_1\cdot\left(1 - \frac{t}{T_C\cdot K}\right)^{0.35} + \beta_2\cdot\left(1 - \frac{t}{T_C\cdot K}\right)^{\frac{2}{3}} + \beta_3\cdot\left(1 - \frac{t}{T_C\cdot K}\right) + \beta_4\cdot\left(1 - \frac{t}{T_C\cdot K}\right)^{\frac{4}{3}}\right]\cdot\frac{kmol}{m^3}$$

Key Point: Put in the units explicitly when you define the function.

Check the Results



3.4 Fitting Data to an Expression of Any Form Using Genfit

When the function to fit is not a line, polynomial, or linear combination, you must use genfit.

Key Point: Only use genfit if you have to, genfit requires good guesses and it is easier to make errors with genfit.

Syntax

$$g(x,\gamma) = \begin{pmatrix} y(x) \\ \frac{\delta y(x)}{\delta\gamma_0} \\ \frac{\delta y(x)}{\delta\gamma_1} \\ \vdots \\ \frac{\delta y(x)}{\delta\gamma_n} \end{pmatrix} \qquad \gamma_{guess} = \begin{pmatrix} guess_{A0} \\ guess_{A1} \\ \vdots \\ guess_{An} \end{pmatrix} \qquad \gamma = genfit(xdata, ydata, \gamma_{guess}, g)$$

Demonstration

Experimental data for the vapor pressure of acetone is contained in the matrices below. Fit the data to the Antoine Equation:

$$\ln(Psat) = A - \frac{B}{T + C}$$

D :=

	0	1
0	203.65	58.928
1	212.45	119.72
2	234.05	797.13
3	243.25	$1.581 \cdot 10^3$
4	253.34	$3.48 \cdot 10^3$
5	259.18	$4.267 \cdot 10^3$
6	262.12	$5.182 \cdot 10^3$
7	282.23	$1.484 \cdot 10^4$
8	265.04	$6.005 \cdot 10^3$
9	267.05	$6.853 \cdot 10^3$

$Temp_{exp} := D^{\langle 0 \rangle} \cdot K \qquad Psat_{exp} := D^{\langle 1 \rangle} \cdot Pa$

The Antoine Equation must be fit using genfit. The first thing needed is the partial derivatives of the function with respect to each parameter (constant). You can do the derivatives in your head or using Mathcad.

RHS of Antoine Equation

$$a_{zero} - \frac{a_{one}}{t + a_{two}}$$

Partial Derivatives of RHS of Antoine Equation.

$$\frac{d}{da_{zero}}\left(a_{zero} - \frac{a_{one}}{t + a_{two}}\right) \rightarrow 1$$

$$\frac{d}{da_{one}}\left(a_{zero} - \frac{a_{one}}{t + a_{two}}\right) \rightarrow -\frac{1}{a_{two} + t}$$

$$\frac{d}{da_{two}}\left(a_{zero} - \frac{a_{one}}{t + a_{two}}\right) \rightarrow \frac{a_{one}}{\left(a_{two} + t\right)^2}$$

Key Point: If you use Mathcad to do the derivatives, you cannot use [notation for the subscripts.

Create the g(x,a) matrix. (Note: this doesn't have to be called g.)

$$g(t,a) := \begin{bmatrix} a_{zero} - \dfrac{a_{one}}{t + a_{two}} \\ 1 \\ -\dfrac{1}{a_{two} + t} \\ \dfrac{a_{one}}{(a_{two} + t)^2} \end{bmatrix}$$

Change the subscripts to [notation.

$$g(t,a) := \begin{bmatrix} a_0 - \dfrac{a_1}{t + a_2} \\ 1 \\ -\dfrac{1}{a_2 + t} \\ \dfrac{a_1}{(a_2 + t)^2} \end{bmatrix}$$

Key Points:
1. Genfit requires a the constants be a matrix variable. Use [notation.
2. One common mistake is to interchange the t and the a order when defining g(t,a). This must not be done.

Provide some guess values for a.

$$a_g := \begin{pmatrix} 50 \\ 3000 \\ 50 \end{pmatrix}$$

Key Point: You should have a guess value for each constant in your equation.

Use genfit to get the correct constants.

$$a := \text{genfit}\left(\frac{Temp_{exp}}{K}, \ln\left(\frac{Psat_{exp}}{Pa}\right), a_g, g\right)$$

$$a = \begin{pmatrix} 21.662 \\ 2.997 \times 10^3 \\ -33.723 \end{pmatrix}$$

Key Points:
1. If genfit doesn't converge, try a new guess value.
2. Notice that the y data is ln(Psat) as required by the original equation.
3. You cannot have units on the data so divide them out.
4. The last argument in genfit is the name of the matrix containing the partial derivatives. Do not put g(t,a); it is just g.

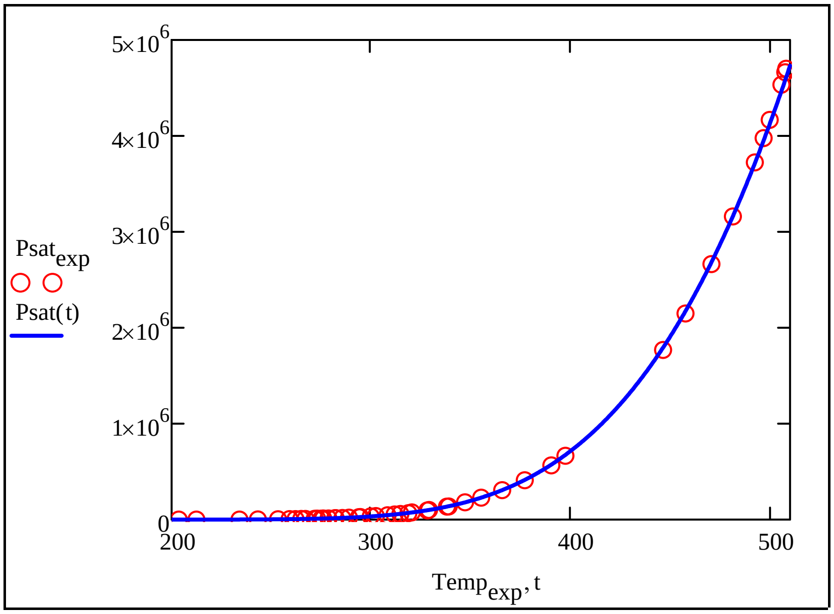
Define the function with the correct parameters.

$$Psat(t) := \exp\left(a_0 - \frac{a_1}{\frac{t}{K} + a_2}\right)\cdot Pa$$

$$Psat(300K) = 3.311 \times 10^4\ Pa$$

Check the fit by plotting

Key Point: You must always check to make sure the fit is good by plotting. If it doesn't fit well, change your guess and ensure you did not make a mistake.



Extra Practice

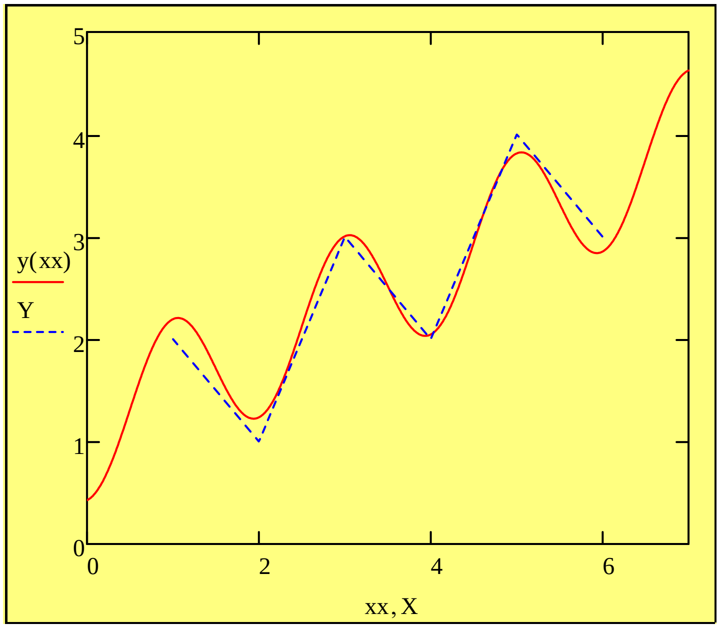
Fit the data given in the matrices to the right to the following expression.

$$y(x) = \gamma_0\cdot\cos(\gamma_1\cdot x) + \gamma_2\cdot x + \gamma_3$$

$$X := \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{pmatrix} \qquad Y := \begin{pmatrix} 2 \\ 1 \\ 3 \\ 2 \\ 4 \\ 3 \end{pmatrix}$$

$$g(x,\gamma) := \begin{pmatrix} \gamma_0\cdot\cos(\gamma_1\cdot x) + \gamma_2\cdot x + \gamma_3 \\ \cos(\gamma_1\cdot x) \\ -\gamma_0\cdot x\cdot\sin(\gamma_1\cdot x) \\ \gamma_2 \\ 1 \end{pmatrix} \qquad \gamma_g := \begin{pmatrix} -.2 \\ 2.5 \\ .40 \\ .10 \end{pmatrix} \qquad \gamma := \mathrm{genfit}(X, Y, \gamma_g, g) \qquad \gamma = \begin{pmatrix} -0.682 \\ 3.146 \\ 0.405 \\ 1.109 \end{pmatrix}$$

$$y(x) := \gamma_0\cdot\cos(\gamma_1\cdot x) + \gamma_2\cdot x + \gamma_3$$



5
4
3
2
1
0
0
2
4
6
y(xx)
Y
xx, X