Proceedings of the Third International DERIVE/TI-92 Conference

Curve Fitting with DERIVE

Steven Schonefeld
Tri-State University
Angola, IN 46703
schonefelds@alpha.tristate.edu

Introduction

In my numerical analysis class, we spend several days studying least squares curve fitting. We use DERIVE to implement these curve fits. The students are more interested in problems which come from real-life data. I found a treasure trove of data in ***THE WORLD ALMANAC AND BOOK OF FACTS***. The first page of interest contains a table of **Monthly Normal Temperature and Precipitation** for selected cities – mostly from the United States. The students work in two-person teams – one takes notes, the other operates the computer. Of course, these duties and the partnerships are permuted regularly during the school term. For several of our experiments, I have each team select a city from the table of **Normal Temperature and Precipitation**. They use the data for their selected city for several least square curve fitting experiments.

Fitting a straight line.

Suppose we wish to fit a linear function

$$y = ax + b$$

to data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. The ***Least Square Line*** minimizes the square error:

$$SQE(a, b) = \Sigma(ax_k + b - y_k)^2$$

over all possible pairs of numbers (a, b). This line may be found by solving a system of linear equations (***Normal Equations)*** for (a, b). Fortunately, DERIVE does all this dirty work for us.

Suppose that a team of students selected Baltimore, MD as their city for the experiments. The data for this city appears in the table as follows:

Jan.		Feb.		Mar.		..	Dec.	
T.	P.	R.	P.	T.	P.	…	T.	P.
33	3.0	35	3.0	43	3.7	…	37	3.4

Thus, in Baltimore, the January the mean temperature is 33^o F and the mean precipitation is 3.0 inches. For February, these numbers are 35^o F and 3.0 inches, etc. The first experiment involves fitting a least square line to the temperature versus precipitation. The students must define the data matrix:

m := [[33, 3.0], [35, 3.0], [43, 3.7], . . . [37, 3.4]]

and **A**uthor and **S**implify

FIT([x, ax + b], m)

to obtain the linear function

0.0204407·x + 2.36401

that will minimize the square error. This function is plotted together with the data points on the same coordinate axes to give a visual goodness of fit. See Figure 1 for this plot. Students also calculate the square error for this fit.

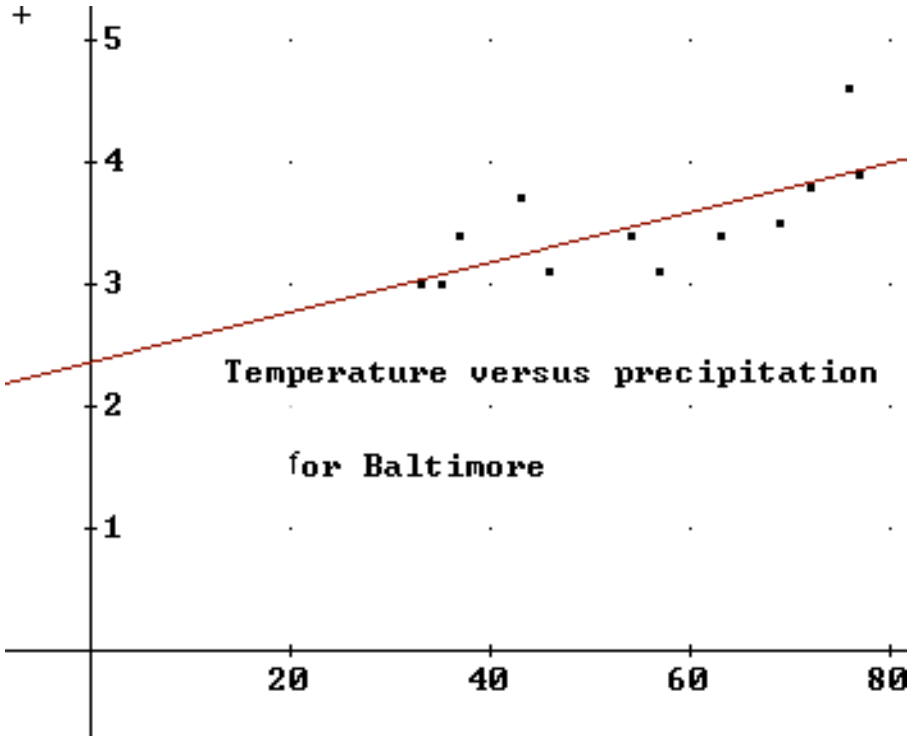

Figure 1

Fitting a sinusoidal curve.

For our next experiment, we replace each month by a number (Jan. = 1, Feb. = 2, Mar. = 3, . . .) and use this number as the first coordinate of data points. The second coordinate will be either temperature or precipitation taken from the table of **Normal Temperature and Precipitation**. For temperatures at Baltimore, we have a new matrix of data points:

m := [[1, 33], [2, 35], [3, 43], . . . [12, 37]].

We plot these data points and discover they appear to be sinusoidal in nature, with period, T = 12 (months.) We find the best fit over all such curves by **A**uthoring and **S**implifying:

FIT([t, a + b COS(πt/6) + c SIN(πt/6)], m)

to obtain the sinusoidal function

– 17.8045·COS(0.523598·t) – 12.9341·SIN(0.523598·t) + 55.1666.

Students again plot the data and graph on the same axes. See Figure 2 for this plot. If time permits, we can repeat this experiment with precipitation as the second coordinate. The resulting curves usually do not fit the data points as well.

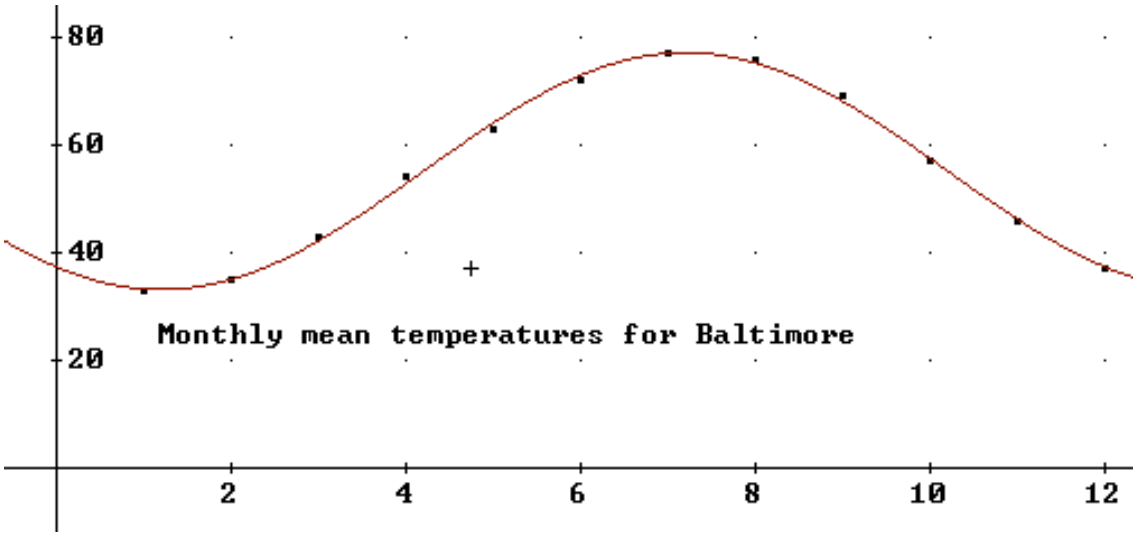

Figure 2

Fitting an exponential function to data.

Suppose we wish to fit a function of the form

$$u = \beta \, EXP(\alpha t) \qquad (*)$$

to data points $(t_1, u_1), (t_2, u_2), \ldots, (t_n, u_n)$. We must minimize the square error:

$$SQE(\alpha, \beta) = \Sigma \, (\beta \, EXP(\alpha t_k) - u_k)^2.$$

Taking first partials results in non-linear normal equations in α and β – which are quite difficult to solve. We can take logarithms of both sides of equation (*) to get

$$\ln(u) = \ln(\beta) + \alpha \, t.$$

By letting $y = \ln(u)$, $a = \alpha$, and $b = \ln(\beta)$, the above equation becomes

$$y = a\,t + b.$$

So, if data points $(t_1, u_1), (t_2, u_2), \ldots, (t_n, u_n)$ are ***close to*** an exponential function (*), then the data points

$$(t_1, \ln(u_1)), (t_2, \ln(u_2)), \ldots, (t_n, \ln(u_n))$$

should be ***close to*** a line

$$y = a\,t + b.$$

The Plan:

1. Transform data points: $\{(t_k, u_k)\}$ to $\{(t_k, \ln(u_k))\}$.
2. Fit a line $y = a\,t + b$ to the new points.
3. Recover α and β to get (*).

It should be noted that the resulting exponential function may not be the closest such function in the least squares sense. However, this exponential function will be close to the desired function. A similar plan (using a different transformation for each function) may be used to fit functions of several other forms, including**:**

$u = \beta t^{\alpha}$
$u = \alpha\,LN(t) + \beta$,
$u = \alpha/t + \beta$,
$u = \beta/(t + \alpha)$,
and **$u = \beta\,t/(t + \alpha)$**.

In all, there are seven such, **two-parameter curves**, counting a straight line. For a given set of data points, we fit all seven of these curves. We then calculate the square error for each function and select the one with the smallest square error. This usually gives a reasonable fit. I have developed a **.MTH** file which automates this process. As usual, the data and one or more of the functions may be plotted on the same graph. We again find data in ***THE WORLD ALMANAC AND BOOK OF FACTS***. This time we use a table of Population of the 100 Largest U.S. Cities. In order to keep the data manageable, we measure the population in thousands and use a linear translation

$$x = (\text{year})/100 - 18$$

on the year. So, $x = 1.0$ corresponds to the year 1900, $x = 1.5$ corresponds to 1950, etc. For Mesa, Arizona, the data points are:

m := [[1, 0.722], [1.5,16.79], [1.6, 33.772], . . ,[1.9, 288.104]]

The results of all the curve fittings are summarized in the following table.

Function	Square error
256.451*t – 313.565	2.68951*10^4
0.000865830*EXP(6.64839*t)	**846.758**
0.590078*t^(9.11083)	8008.64
326.127*LN(t) – 50.7826	3.19832*10^4
354.986 – 397.555/t	3.64929*10^4
2.08018*10^9/(4.59176*10^8 – 2.40164*10^8*t)	1.98202*10^5
1.08272*10^9*t/(3.95332*10^8 – 2.06616*10^8*t)	2.14008*10^5

Clearly, the exponential function has the smallest square error for this set of data. See Figure 3 for a plot of the data – together with this exponential function.

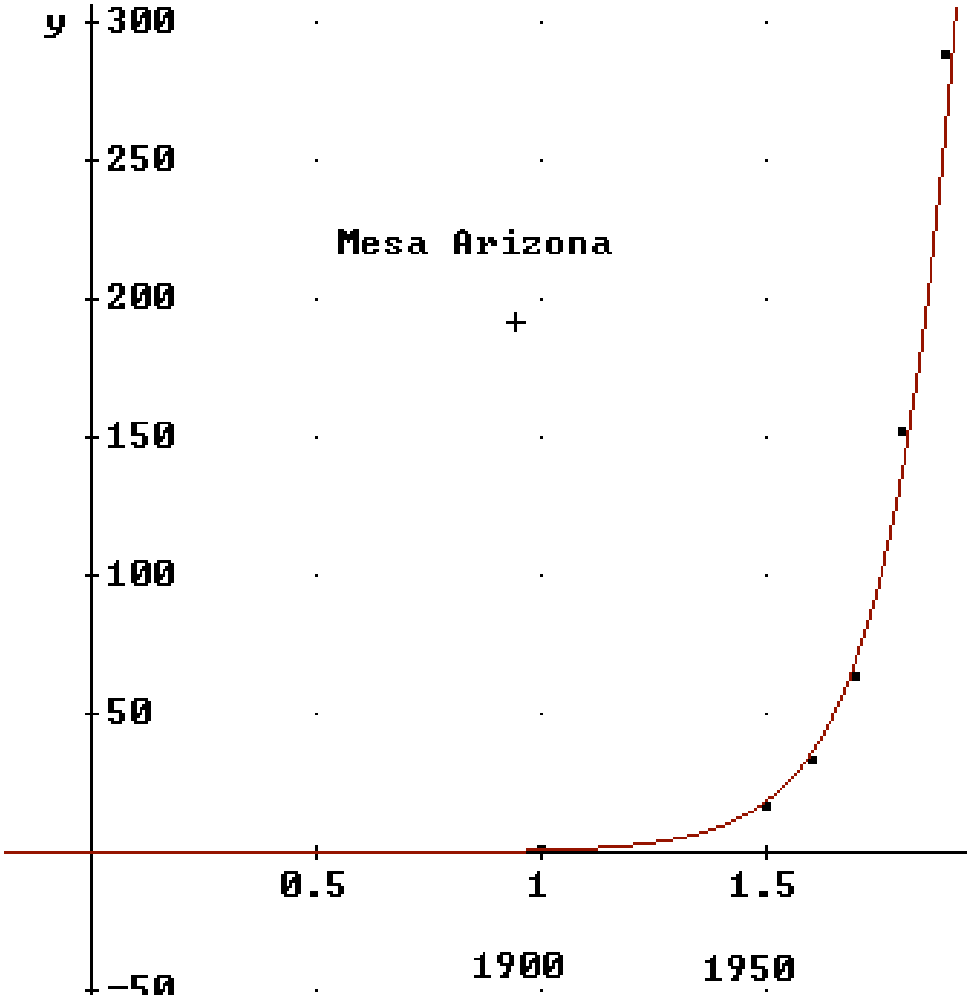

Figure 3

Concluding remarks.

If you are looking for experimental data, try ***THE WORLD ALMANAC AND BOOK OF FACTS***. It contains an ample supply of data to be used for curve fitting. Another advantage of using this source is that no copyrights apply to the material contained therein. If you are interested in the **.MTH** file for fitting the seven two-parameter curves described above, they are included on the disk which is distributed with my book: ***Numerical Analysis via DERIVE***, published by MathWare, 604 E. Mumford Dr., Urbana, IL 61801, phone: 1 (800) 255-2468. Alternately, send me a blank disk and I will send this **.MTH** file to you.