

Applied Calculus

Fall 2008

Module 3: Curve Fitting

Problem 1: Root Growth, a One Parameter Linear Model

Problem 2: Development Test Results, Two Parameter Linear Model

Problem 3: Concentration of Chemical, Two Parameter Non-linear Model

Introduction

In Module 1 we dealt with situation where a model could be fitted exactly to the data. It is always possible to find a model that fits the data exactly. This is not, however, the most appropriate way of modeling the data, especially if want to use the model to make predictions about future events. It is also inefficient, since there are as many parameters to calculate as there are data points. An efficient model is one with the fewest parameters that still manages to describe the behaviour of the data. The disadvantage of using a model with fewer parameters than data points is that it will not pass through all the data points, and so we find the model that is the 'best fit' for the data.

The first problem shows you have to fit the 'best' model of the form $y = mx$ to the data. Within this problem we look at various ways of selecting the 'best' model. In the second problem we fit the 'best' model of the form $y = mx + b$. In the final problem, and in the projects that your group will be assigned, you will first fit find an exact fit model, then a two parameter non-linear model and finally fit a model with more than two parameters. We will then compare the properties of these models.

In the process of completing these problems, we will see how to find minimum and maximum values of functions, and see how to find derivatives of functions in several variables.

Schedule for Module 3

Date	Lecture	Assignment
Mon 6 th Oct	Problem 1: Selecting and Fitting a Model	Homework 1: See Moodle for description. Prepare for the Module 2 Test.
Wed 8 th Oct	Problem 1: Optimizing the Model	Prepare for the Module 2 Test.
Fri 10 th Oct	Module 2 Test	
Mon 13 th Oct	Finding and classifying critical points	Homework 2: Sec 2.3: 1, 5, 7, 8* Sec 4.3: 27, 28* Supplimentary Problems: See page 6 below
Wed 15 th Oct	Problem 2: Finding the best $mx + b$ model for data	Homework 3: See page 11 below
Fri 17 th Oct	Problem 3: Selecting the Model and Objective Function	Read your project description, look up the research paper, read the abstract, and find the table from which your data was taken.
Wed 22 nd Oct	Problem 3: Optimizing the Model	Work on your projects
Fri 24 th Oct	Multiple Variable Functions and Partial Derivatives	Homework 4: Sec 7.2: 2*, 3, 5, 7, 12*, 13, 14*, 21, 22*
Mon 27 th Oct	Problem 4: Working with other multiple parameter models	Homework 5: See page 17 below.
Mon 3 rd Nov	Presentation of Module 3 Projects	
Wed 5 th Nov	Presentation of Module 3 Projects	

1 Problem 1

1.1 Statement of the Problem

A biologist is studying how fast the root of a particular species of plant grows. The results of one experiment are below. The biologist believes that the 'law' of root growth is

$$\text{Growth} = m \times \text{Days},$$

and has asked us to determine the value of m from the experimental data.

Day	Growth (mm)
0	0
1	1.2
2	3.8
3	4.2
4	6.3
5	7.2
6	7.3

Instructions

1. Plot these points on the graph paper provided.
2. Draw a straight line that goes through the origin, and is as close to the other points as possible.
3. Calculate the slope of the line

1.2 Calculate the Distance of the Line from the Points

Fill out the table below for your line:

x_i	y_i	mx_i	$y_i - mx_i$	$(y_i - mx_i)^2$
0	0	0	0	0
1	1.2			
2	3.8			
3	4.2			
4	6.3			
5	7.2			
6	7.3			

$$\sum_{i=0}^6 (y_i - mx_i)^2 =$$

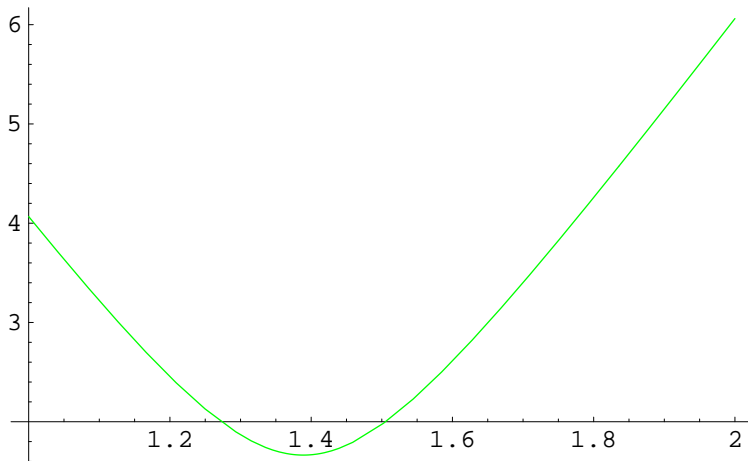
$$V_2(m) = \sqrt{\sum_{i=0}^6 (y_i - mx_i)^2} = \dots\dots\dots$$

Assignment (Task 1):

- Complete this calculation of $V_2(m)$.
- Type your answer into the subject line of an email to your professor. Your answers must be braced together, for example, if your value for m is 1.48 and the value for $V_2(m)$ is 1.88, type: {1.48, 1.88}
- Send the email. Do not type anything else.
- You will lose your homework grade if you do not follow these directions.

1.3 Solution

The best possible value of m is the value that minimizes the function $V_2(m)$. If we plot the function in Mathematica we get



From the graph of function we can see that this is the same as the value of m at which $\frac{dV_2}{dm} = 0$. We can use Mathematica to find the derivative of $V_2(m)$ by typing `D[V2[m],m]` The output is:

$$\frac{-12(7.3 - 6m) - 10(7.2 - 5m) - 8(6.3 - 4m) - 6(4.2 - 3m) - 4(3.8 - 2m) - 2(1.2 - m)}{2\sqrt{(7.3 - 6m)^2 + (7.2 - 5m)^2 + (6.3 - 4m)^2 + (4.2 - 3m)^2 + (3.8 - 2m)^2 + (1.2 - m)^2}}$$

Then to solve $\frac{dV_2}{dm} = 0$, we can type `Solve[D[V2[m],m]==0,m]` and if this fails to work we can use `FindRoot[D[V2[m],m]==0,{m,{1.3,1.5}}]`.

Homework 2

Find all the location of the critical points, within the given domain, of the following functions:

1. $f(x) = x \ln(x)$ for $x \geq 0$.
2. $f(x) = x^2 e^{-x}$ for all real x , i.e., $-\infty < x < \infty$.
3. $f(t) = 3t^5 - 18t^4 - 5t^2 + 7t - 15$ for all real t (**Hint:** use Mathematica).
4. $H(\theta) = \theta \cos(5\theta)$ for $\theta \in [-\pi, \pi]$.

Questions 1 and 4 will be graded.

2 Alternative Optimization Methods

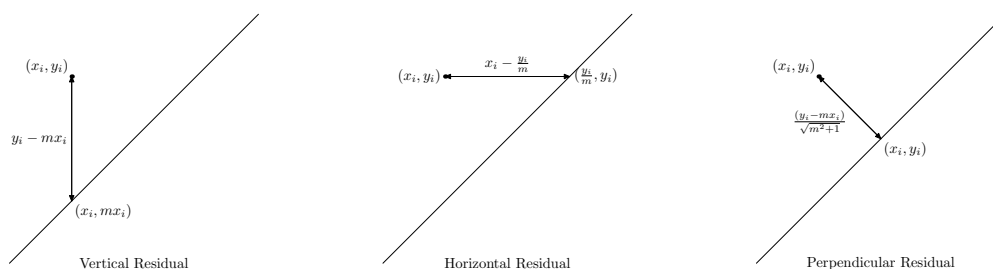
In Problem 1, to find the “best” fitting line of the form $y = mx$ we found the value of m that minimized the function

$$V_2(m) = \sqrt{(1.2 - m)^2 + (3.8 - 2m)^2 + (4.2 - 3m)^2 + (6.3 - 4m)^2 + (7.2 - 5m)^2 + (7.3 - 6m)^2}$$

It was claimed that $V_2(m)$ represented the total distance between the data points and the model (the $y = mx$ line), so the lower this distance the better the fit. There are, however, alternative ways to measure the distance between a data set and a model.

2.1 Alternative Residuals

A residual is a distance from a single data point and the curve of the model. The total distance function is a combination of all the residuals. When deriving the function V_2 , we used ‘vertical’ residuals (i.e., residuals which are parallel to the y -axis). We could, however, use a horizontal residual (i.e., one parallel to the x -axis), or a perpendicular residual (i.e., one at right-angles to the model line). These three alternatives are illustrated below:



To calculate the total distance functions from these three residuals, we can square each residual distance, sum all these, and find the square root of this sum. The values of the seven residuals for these three possibilities are given below:

Data Point	Vertical	Horizontal	Perpendicular
(0, 0)	0	0	0
(1, 1.2)	$1.2 - m$	$1 - \frac{1.2}{m}$	$\frac{1.2 - m}{\sqrt{m^2 + 1}}$
(2, 3.8)	$3.8 - 2m$	$2 - \frac{3.8}{m}$	$\frac{3.8 - 2m}{\sqrt{m^2 + 1}}$
(3, 4.2)	$4.2 - 3m$	$3 - \frac{4.2}{m}$	$\frac{4.2 - 3m}{\sqrt{m^2 + 1}}$
(4, 6.3)	$6.3 - 4m$	$4 - \frac{6.3}{m}$	$\frac{6.3 - 4m}{\sqrt{m^2 + 1}}$
(5, 7.2)	$7.2 - 5m$	$5 - \frac{7.2}{m}$	$\frac{7.2 - 5m}{\sqrt{m^2 + 1}}$
(6, 7.3)	$7.3 - 6m$	$6 - \frac{7.3}{m}$	$\frac{7.3 - 6m}{\sqrt{m^2 + 1}}$

From this table we can easily write down the different total distance functions:

$$V_2(m) = \sqrt{(1.2 - m)^2 + (3.8 - 2m)^2 + (4.2 - 3m)^2 + (6.3 - 4m)^2 + (7.2 - 5m)^2 + (7.3 - 6m)^2}$$

$$H_2(m) = \sqrt{\left(1 - \frac{1.2}{m}\right)^2 + \left(2 - \frac{3.8}{m}\right)^2 + \left(3 - \frac{4.2}{m}\right)^2 + \left(4 - \frac{6.3}{m}\right)^2 + \left(5 - \frac{7.2}{m}\right)^2 + \left(6 - \frac{7.3}{m}\right)^2}$$

$$P_2(m) = \sqrt{\left(\frac{1.2 - m}{\sqrt{m^2 + 1}}\right)^2 + \left(\frac{3.8 - 2m}{\sqrt{m^2 + 1}}\right)^2 + \dots + \left(\frac{7.2 - 5m}{\sqrt{m^2 + 1}}\right)^2 + \left(\frac{7.3 - 6m}{\sqrt{m^2 + 1}}\right)^2}$$

These functions are defined in the Mathematica file named `M3_P1_OtherOfs.nb`. We find the critical point of these functions, and use a test to confirm that the point is a local minimum. Note that each function gives a different optimal value for the slope of the of the straight line model.

2.2 Different ways of totaling the distance

Once the type of residual has been selected we have to have a scheme for combining the individual residuals into a ‘total distance’. The scheme we have been using is to square, add and square root the sum (we can call this as ‘pythagorean scheme’). Other schemes are:

- Just sum the residuals.
- Cube, sum, and cube root the sum.
- Take the fourth power, sum, and take the fourth root of the sum.
- Take the fifth power, sum, and take the fifth root of the sum.
- etc.

In other words we can define the total distance function, for this set of data, by any one of the formulae below, for any whole value of p :

$$V_p(m) = [(1.2 - m)^p + (3.8 - 2m)^p + \dots + (7.2 - 5m)^p + (7.3 - 6m)^p]^{\frac{1}{p}}$$

$$H_p(m) = \left[\left(1 - \frac{1.2}{m}\right)^p + \left(2 - \frac{3.8}{m}\right)^p + \dots + \left(5 - \frac{7.2}{m}\right)^p + \left(6 - \frac{7.3}{m}\right)^p \right]^{\frac{1}{p}}$$

$$P_p(m) = \left[\left(\frac{1.2 - m}{\sqrt{m^2 + 1}}\right)^p + \left(\frac{3.8 - 2m}{\sqrt{m^2 + 1}}\right)^p + \dots + \left(\frac{7.2 - 5m}{\sqrt{m^2 + 1}}\right)^p + \left(\frac{7.3 - 6m}{\sqrt{m^2 + 1}}\right)^p \right]^{\frac{1}{p}}$$

Note that for odd values of p the functions V_p , H_p and P_p do not have a minimum. They either increase or decrease. They do, however, pass through the x -axis, meaning that there is a value of m for which the ‘total distance’ between data and model is zero (this is when the positive residuals cancel out the negative residuals). So, for odd values of p we find the optimal value of m by finding the zero of the distance function.

3 Problem 2

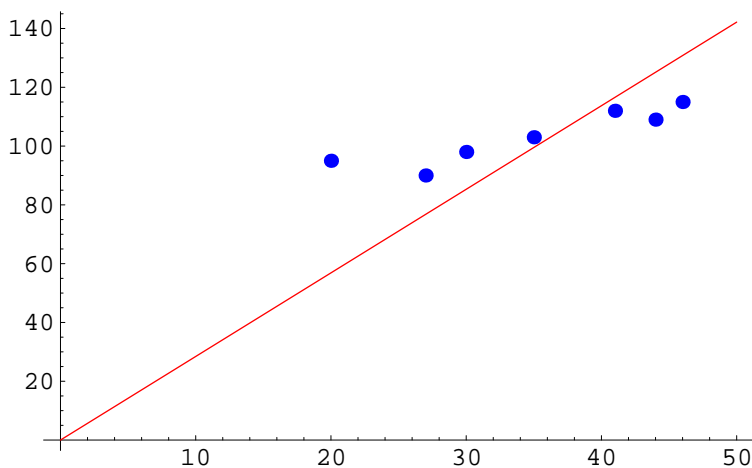
3.1 Statement of the Problem

The following data shows the age of seven children, in months, and their score on a particular developmental test.

Age (Months)	Score
20	95
27	90
30	98
35	103
41	112
44	109
46	115

3.2 Fitting $y = mx$

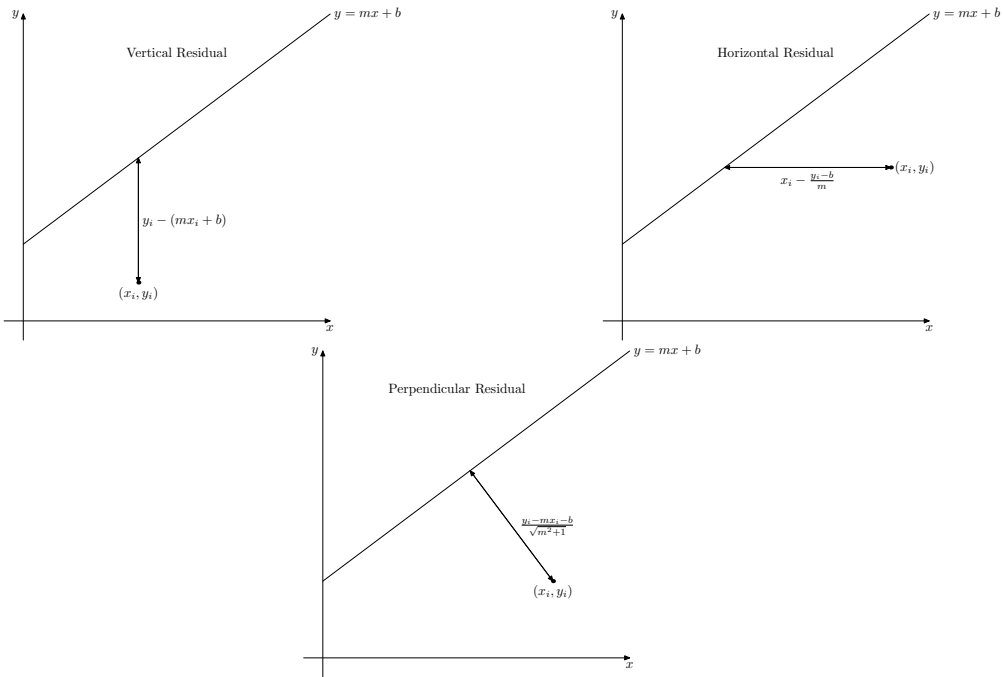
To try to fit $y = mx$ model to this data, we can set the objective function to be $V_2(m) = \sqrt{(95 - 20m)^2 + (90 - 27m)^2 + \dots + (109 - 44m)^2 + (115 - 46m)^2}$, finding the minimum of this function is attained when $m = \dots\dots\dots$ (fill in this answer from the lecture). When we plot the data together with $y = x$ the Mathematica output is



Clearly this is a poor fit, so all models of the form $y = mx$ are invalid. We need to use a different type of model.

3.3 Fitting $y = mx + b$

For this data no value of m can make $y = mx$ fit the data well. The next simplest model is one of the form $y = mx + b$. The figure below shows the vertical, horizontal and perpendicular residuals.



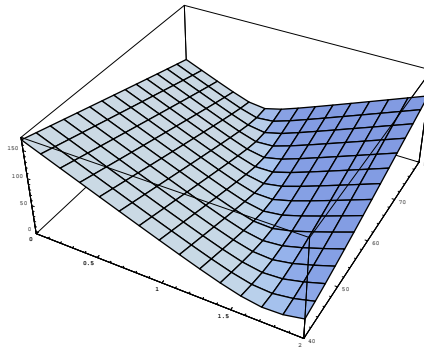
Hence, the list of objective functions, we have to choice from, is:

$$V_p(m, b) = [(95 - 20m - b)^p + (90 - 27m - b)^p + \dots + (109 - 44m - b)^p + (115 - 46m - b)^p]^{\frac{1}{p}}$$

$$H_p(m, b) = \left[\left(20 - \frac{95 - b}{m}\right)^p + \left(27 - \frac{90 - b}{m}\right)^p + \dots + \left(44 - \frac{109 - b}{m}\right)^p + \left(46 - \frac{115 - b}{m}\right)^p \right]^{\frac{1}{p}}$$

$$P_p(m, b) = \left[\left(\frac{95 - 20m - b}{\sqrt{m^2 + 1}}\right)^p + \left(\frac{90 - 27m - b}{\sqrt{m^2 + 1}}\right)^p + \dots + \left(\frac{109 - 44m - b}{\sqrt{m^2 + 1}}\right)^p + \left(\frac{115 - 46m - b}{\sqrt{m^2 + 1}}\right)^p \right]^{\frac{1}{p}}$$

Their values will change with both m and b . Their graphs are surfaces. Below is the graph of $V_2(m, b)$:



It is hard to see that this function does have a minimum value. However we can try to solve the two of simultaneous equations $\frac{\partial V_2}{\partial m} = 0$ and $\frac{\partial V_2}{\partial b} = 0$. This will give us

points, if any exist, where both rate of change in the m -direction is zero, and the rate of change in the b -direction is zero.

Homework 3

- Using the commands in the Mathematica workbook `M3_P2_AffineModel.nb`, answer the questions below.
 - The answers should be written in a word-document, named "your last name"_M3_HW3.doc
 - You may work together with your fellow group members if you wish, but should submit individual copies of your work.
1. Using the total distance function $H_4(m, b)$, find the optimal model for the test data, i.e.

Age (Months)	Score
20	95
27	90
30	98
35	103
41	112
44	109
46	115

Provide a graphic evidence that the model is valid.

2. Using the $P2$ distance function, fit the best model of the form $y = mx + b$ to the data about the growth of plant roots from Problem 1, i.e.

Day	Growth (mm)
0	0
1	1.2
2	3.8
3	4.2
4	6.3
5	7.2
6	7.3

Comment on whether a model of the form $y = mx$ or a model of the form $y = mx + b$ is most suitable for this data.

4 Problem 3

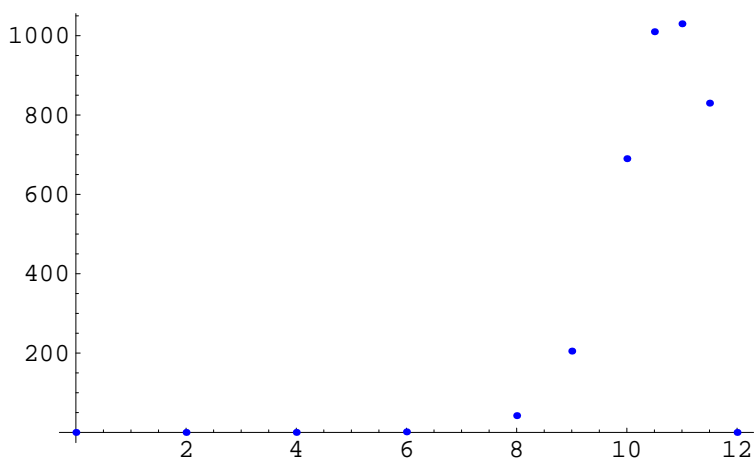
4.1 Statement of the Problem

A factory needs to use a certain toxic chemical in its processing. There is a steady flow of water through the machinery that washes the chemical into a settlement tank. The factory runs its process for eleven hours and then stops for one hour to allow the equipment to be washed clear of the chemical. The table below shows the concentration of chemical in the outlet pipe, at different times during a typical twelve hour shift.

Time (Hours)	Concentration (mg/L)
0	0.00
2	0.00
4	0.120
6	1.42
8	42.30
9	205.00
10	690.00
10.5	1010.00
11	1030.00
11.5	830.00
12	0.00

4.2 Selecting the Model

When plotting the data Mathematica gives a graphic of the form:



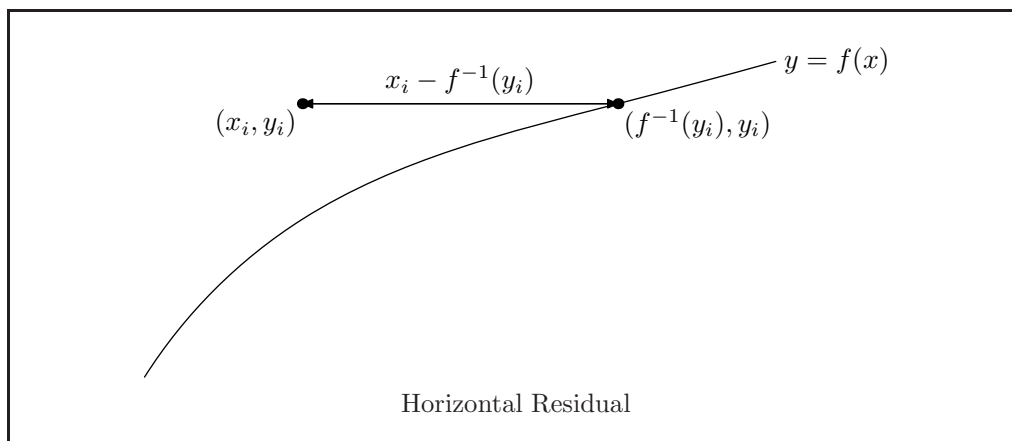
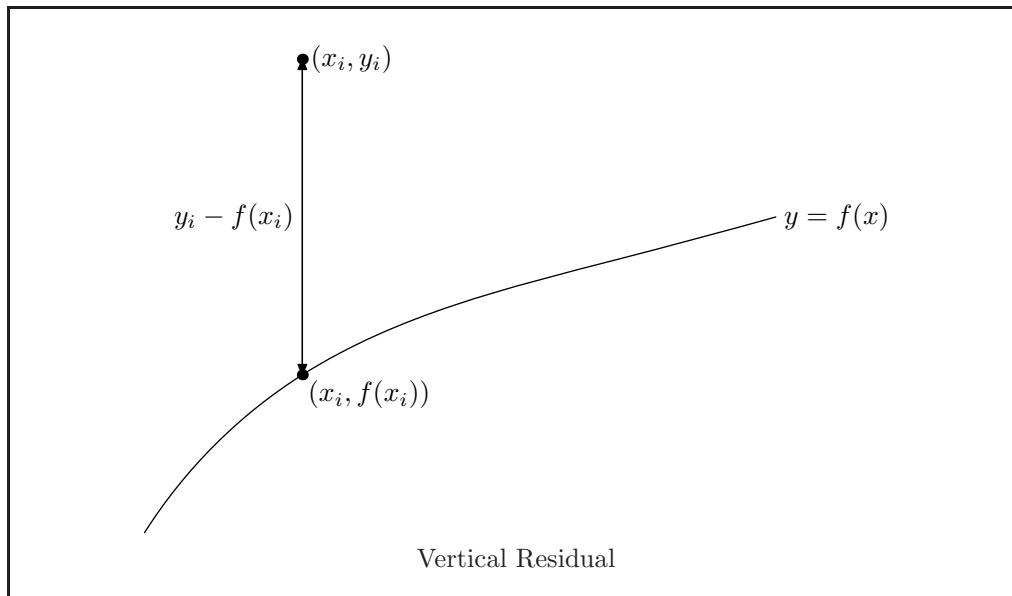
The model that is suggested by this data is one of the form

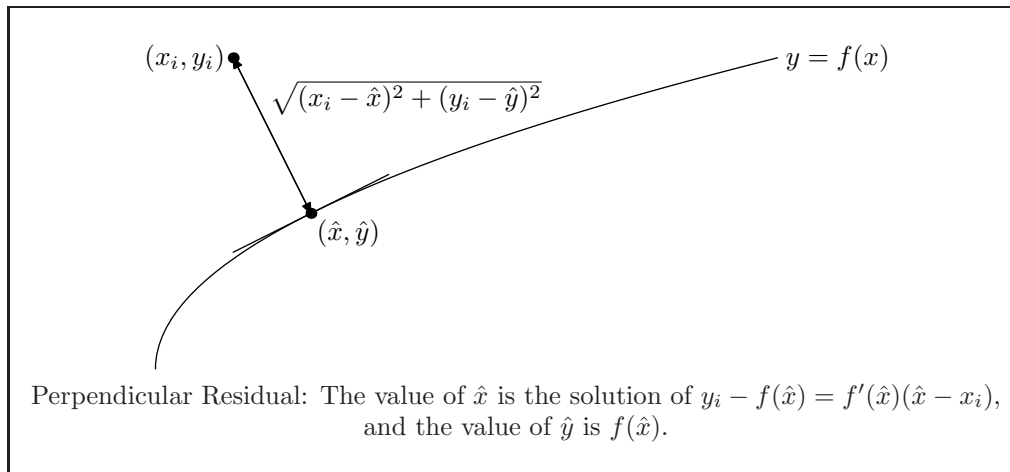
$$y = a(12 - x)^2 e^{-b(12-x)}$$

To find the 'best' values of a and b , we will set up a total distance function in a manner similar to how we proceeded for straight line models.

4.3 Definition of Residuals to curves

The next three diagrams show how the different residuals are defined

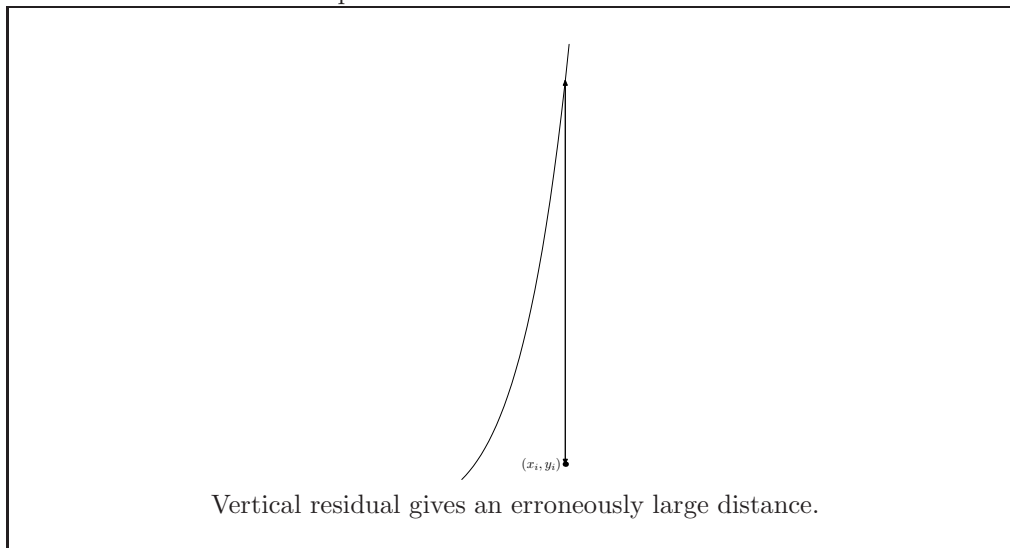


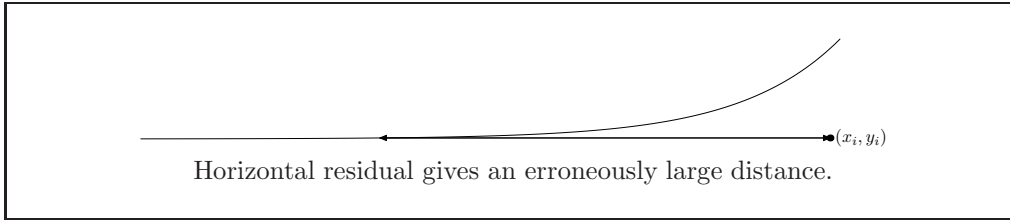


Notice that in order to calculate either horizontal or perpendicular residuals one must first solve an equation. Consequently, these often involve practical difficulties, and vertical residuals are often the only ones that can be calculated.

4.4 Misleading Residuals

There are situations where using horizontal or vertical residuals can give inaccurate values to ‘the distance from point to curve’. These are illustrated below.





Note that in both of these situations the perpendicular residual gives a more reasonable value to ‘distance from point to curve’, however finding the \hat{x} which satisfies $y_i - f(\hat{x}) = f'(\hat{x})(\hat{x} - x_i)$ could be difficult.

4.5 Useful Tricks for Minimizing Total Distance Functions

4.5.1 Simplifying the total distance function

The Pythagorean total distance for vertical residuals is the function:

$$V_2(a, b) = \sqrt{(y_1 - f_{(a,b)}(x_1))^2 + (y_2 - f_{(a,b)}(x_2))^2 + \cdots + (y_n - f_{(a,b)}(x_n))^2},$$

where a , and b are the parameters of the model. The values of a and b which minimize this function will also minimize the function:

$$V_2^2(a, b) = (y_1 - f_{(a,b)}(x_1))^2 + (y_2 - f_{(a,b)}(x_2))^2 + \cdots + (y_n - f_{(a,b)}(x_n))^2.$$

If this seems unlikely to you, just think about a list of positive numbers, (e.g. $\{5, 2, 8, 9, 8, 11, 4\}$), take note of the position of the minimum number, now square every number in the list (e.g. $\{25, 4, 64, 81, 121, 16\}$), and the minimum of this list is in the same position as the minimum of the original list. Notice that this trick only works when the function (or list of numbers) is positive.

4.5.2 Estimating the parameters by finding an ‘exact fit’

Often we can find reasonable estimates for the parameters by finding a model that is an exact fit to a few data points. If your model has 3 parameters, say, then you should pick three of the data points that are representative of the spread of the data, then find the parameter values that force the model to pass exactly through these points. These parameter values will not be the best fit for all the data but very often they are close to the optimal values. For this data and model, we would pick two points e.g. $(6, 1.42)$ and $(11, 1030)$ (avoid the points with y -value zero) and solve the following equations simultaneously:

$$\begin{aligned} 1.42 &= a \times 6^2 \times e^{-b \times 6} \\ 1030 &= a \times 1^2 \times e^{-b \times 1} \end{aligned}$$

5 Problem 4

5.1 Statement

The most critical stage of a 100m race is as the athlete accelerates out of the blocks. The athlete will be able to maintain the speed they manage to reach in this phase. Below are the values of an athlete's acceleration in the first 0.8 seconds of a practice run. What is the athlete's speed 0.8 seconds into the run?

Time (sec)	Acceleration (m/s ²)
0	20.4
0.1	20.3
0.2	19.4
0.3	17.3
0.4	14.0
0.5	9.8
0.6	5.3
0.7	1.6
0.8	0

5.2 Fitting a four parameter

Suppose we want to fit a cubic model to the data, ie. a model of the form $a(t) = at^3 + bt^2 + ct + d$, then we will follow the same procedure as before. We define the total distance function, and find where this function is a minimum. In this example if used the V_2 total distance function it would be defined as

$$V_2(a, b, c, d) = [(20.4 - a(0))^2 + (20.3 - a(0.1))^2 + (19.4 - a(0.2))^2 + (17.3 - a(0.3))^2 + (14.0 - a(0.4))^2 + (9.8 - a(0.5))^2 + (5.3 - a(0.6))^2 + (1.6 - a(0.7))^2 + (0 - a(0.8))^2]^{\frac{1}{2}}$$

To minimize this function we find all four partial derivatives of V_2 , and find the locations when all four are simultaneously equal to zero. That is solve:

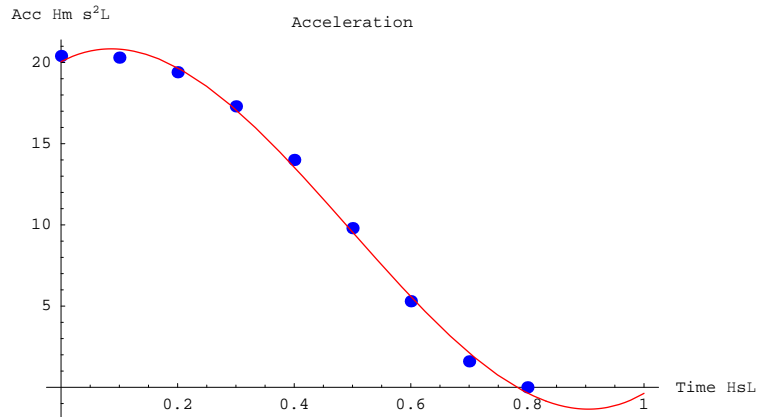
$$\frac{\partial V_2}{\partial a} == 0, \quad \frac{\partial V_2}{\partial b} == 0, \quad \frac{\partial V_2}{\partial c} == 0, \quad \frac{\partial V_2}{\partial d} == 0$$

We now need to run some type of test to classify the type of critical points we have found. Unfortunately the equivalent of the second derivative test is too complex for this course. We can't even examine the graph of V_2 to see if the location looks like a minimum, because the graph of V_2 lies in five dimensional space. The only check we can perform is to plot the data and the new model and see if it is a reasonable fit.

5.3 Solution

When each of the stages is done on Mathematica (using file: `M3_P4_acceleration.nb`) we find that the model that minimizes the V_2 function is $a(t) = 80.8923t^3 - 120.209t^2 +$

$18.8938t + 20.0525$ and if this plotted next to the data we find that it is a reasonable fit:



So we cross our fingers and hope that this really is a minimum and not some four dimensional equivalent of a saddle point.

Homework 5:

- Using the commands in the Mathematica workbook `M3_P4_Acceleration.nb`, answer the questions below.
 - The answers should be written in a word-document, named "your last name"_M3_HW4.doc
 - You may work together with your fellow group members if you wish, but should submit individual copies of your work.
1. Find the model of the form $at^4 + bt^3 + c$ that is a best fit, using V_2 , of the data in this problem. Include a graph that supports your claim that your model is the one that fits the data best.