

Activity 11.9

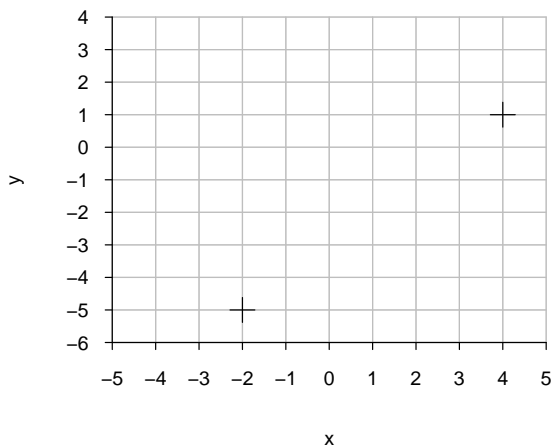
[V/V109]

Consider the following very small data set:

y	x
-5	-2
1	4

Part 1

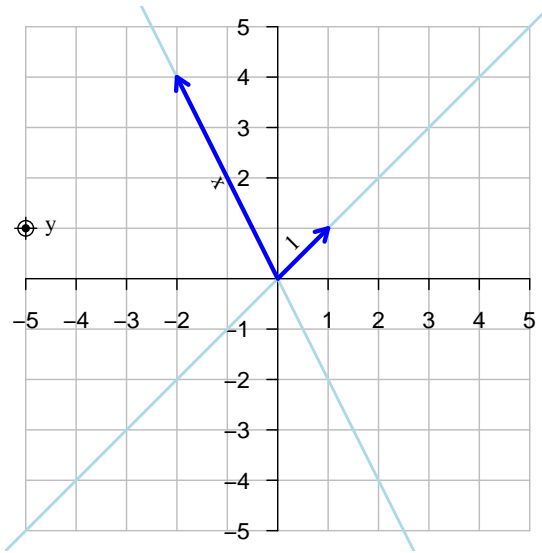
Here is a graph of y versus x for the $N = 2$ cases.



Find, by eye, the best fitting model of the form $y \sim x + 1$. It should be very easy to do this, since the model form describes a simple linear relationship and a line can be drawn that goes exactly through the $N = 2$ points. From your graph of the fitted model, find the intercept and slope coefficients.

Part 2

Now we will perform the same fitting process in the data space. The graph shows the goal point, $\begin{bmatrix} 2 \\ 4 \end{bmatrix}$, along with the two model terms: $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ (labelled simply "1") and $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$ (labelled "x").



By taking a linear combination of these two vectors we can reach the goal point y exactly. Find the multipliers of the “1” and x vector that bring you to y . Compare these multipliers to the coefficients you found by drawing the conventional slope-intercept-format line through the two points.

Part 3

The usual situation in model fitting is that we cannot reach the goal directly with the available model vectors.

Let’s consider the very simple model $y \sim 1$ which models all the y values as being the same.

In the y versus x graph, this model would have the form of a horizontal line. Draw in the horizontal line that is closest to the two data points. The intercept of this line will be the coefficient of the model. What is it?

Now determine the coefficient in this model using the data space graph. You can do this by finding the model point, that is, the point on the subspace defined by vector “1” that is closest to the goal y . The coefficient is the scalar multiplier on “1” that will bring you to this model point.

Find the residual vector from this model and compare it to the residuals you find from the y versus x graph.

Part 4

A less commonly encountered model is $y \sim x - 1$. This is an ordinary line, but one whose intercept is restricted to be zero. This is a line that passes through the origin.

In the y versus x graph, find, by eye, the line that goes through the origin and that comes as close as possible to both of the data points. What is the slope of this line? What are the residuals?

Now, in the data space graph, find the model coefficient by finding the point on the x subspace that is as close as possible to the goal point y . The model coefficient will be the multiplier on x that brings you to this point.

Also, find the residuals from the fitted model and compare these to the residuals you found in the y versus x graph.