

Curve Fitting

Justin Wyss-Gallifent

July 21, 2021

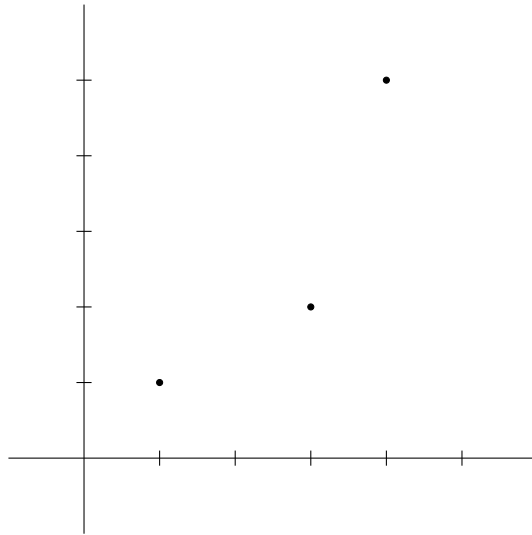
4.1	Straight Line Fitting	2
4.1.1	Introductory Example	2
4.1.2	Least Squares Line	4
4.2	More General Curve Fitting	4
4.3	More General Surface Fitting	6
4.4	Real World Modeling and Predictions	7
4.4.1	Choosing a Function	7
4.4.2	Predicting	10
4.5	Matlab	10
4.6	Exercises	11

4.1 Straight Line Fitting

4.1.1 Introductory Example

A classic application of the method of least squares is illustrated by the following example:

Example 4.1. Consider the three points $(1, 1)$, $(3, 2)$ and $(4, 5)$. As we can see these do not lie on a straight line:



But suppose we want to find a line that's really close to the points, whatever that might mean. How can we apply the above method to do this?

Let's look at the problem. We're trying (and failing) to find a line $y = mx + b$ such that all three points line on it. This means that we want the following to be true:

$$1 = m(1) + b$$

$$2 = m(3) + b$$

$$5 = m(4) + b$$

Or, as a matrix equation:

$$\begin{bmatrix} 1 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix}$$

Since we can't solve this (the points don't lie on a line) let's see what the least-squares solution is:

$$\begin{bmatrix} \widehat{m} \\ \widehat{b} \end{bmatrix} = (A^T A)^{-1} A^T \bar{b} = \begin{bmatrix} 17/14 \\ -4/7 \end{bmatrix}$$

This means $y = \frac{17}{14}x - \frac{4}{7}$ is somehow the best line. What does this mean?

Looking back at our matrix equation the vector

$$\begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix}$$

contained the y -values that we wanted to get but could not. Instead the y -values that we did get, those contained in the vector

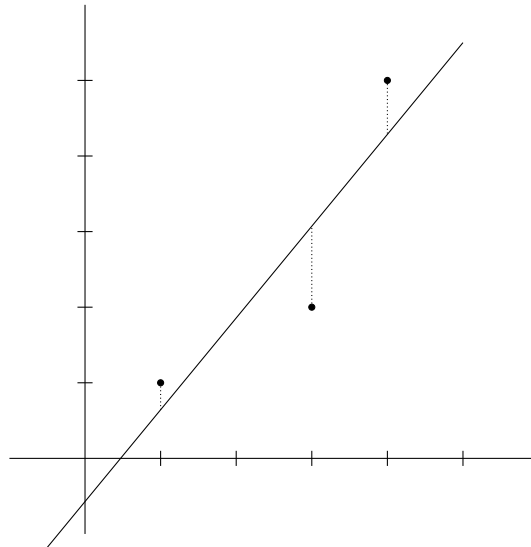
$$\begin{bmatrix} 1 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} \widehat{m} \\ \widehat{b} \end{bmatrix}$$

are those that minimize

$$\left\| \begin{bmatrix} 1 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} \widehat{m} \\ \widehat{b} \end{bmatrix} - \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix} \right\|$$

meaning we're minimizing the sum of the squares of the differences between the y -values we wanted and the y -values we obtained.

This can be nicely illustrated by the following picture where we've minimized the sum of the squares of the dotted distances shown:



An interesting note about the previous example is that there are two things going on at once. First, we're finding a best-fit line where "best-fit" means that the sum of the squares of the vertical distances from the points to the line is minimum. Second, we're attempting a matrix equation which is really a three-dimensional problem with no actual solution but with a least-squares solution.

4.1.2 Least Squares Line

We can summarize this as a definition and theorem:

Theorem 4.1.2.1. Given a set of points $(x_1, y_1), \dots, (x_n, y_n)$ with not all of the x_i equal, the *least squares line* is the line obtained by finding the least squares solution to

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

This line minimizes the sum of the squares of the distances between the y -values of the points and the y -values on the line.

4.2 More General Curve Fitting

Least squares doesn't only work for finding a straight line but it can work for finding any function in which the function is linear in the unknown variables.

What this means is as long as the function you're trying to fit has the form:

$$f(x) = a_1 f_1(x) + a_2 f_2(x) + \dots + a_n f_n(x)$$

Where the $f_i(x)$ are known, then least squares may be used to find the a_i .

Example 4.2. Consider the points $(-1, 2)$, $(0, 0)$, $(1, 2)$ and $(2, 3)$. These almost follow a parabola. Suppose we want to find a function $f(x) = ax^2 + bx + c$ (a parabola) which does a good job of fitting these four points.

Ideally we'd like the function to actually pass through these points, meaning:

$$\begin{aligned}
a(-1)^2 + b(-1) + c &= 2 \\
a(0)^2 + b(0) + c &= 0 \\
a(1)^2 + b(1) + c &= 2 \\
a(2)^2 + b(2) + c &= 3
\end{aligned}$$

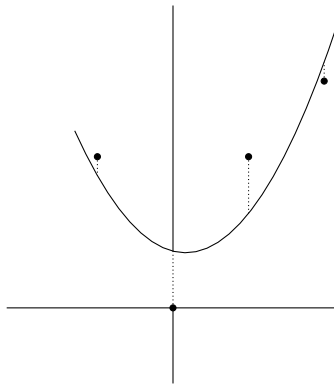
As a matrix equation:

$$\begin{bmatrix} 1 & -1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 2 \\ 3 \end{bmatrix}$$

This has no solution but we can find a least-squares solution:

$$\begin{aligned}
\begin{bmatrix} 1 & -1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \end{bmatrix}^T \begin{bmatrix} 1 & -1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \end{bmatrix} \widehat{\begin{bmatrix} a \\ b \\ c \end{bmatrix}} &= \begin{bmatrix} 1 & -1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \end{bmatrix}^T \begin{bmatrix} 2 \\ 0 \\ 2 \\ 3 \end{bmatrix} \\
\begin{bmatrix} 18 & 8 & 6 \\ 8 & 6 & 2 \\ 6 & 2 & 4 \end{bmatrix} \widehat{\begin{bmatrix} a \\ b \\ c \end{bmatrix}} &= \begin{bmatrix} 16 \\ 6 \\ 7 \end{bmatrix} \\
\widehat{\begin{bmatrix} a \\ b \\ c \end{bmatrix}} &= \begin{bmatrix} 0.75 \\ -0.25 \\ 0.75 \end{bmatrix}
\end{aligned}$$

So that the parabola $f(x) = 0.75x^2 - 0.25x + 0.75$ provides the best fit, where best means minimizes the sum of the squares of the vertical distances.



Just to really understand what we cannot do, and why:

Example 4.3. Consider the points $(-1, 1)$, $(1, 3)$ and $(2, 10)$. Suppose we believe these points follow the function $f(x) = ax + \sin(bx)$. This is all well and good except the corresponding system of equations is:

$$\begin{aligned}a(-1) + \sin(-b) &= 1 \\a(1) + \sin(b) &= 3 \\a(2) + \sin(2b) &= 10\end{aligned}$$

Unfortunately this cannot be written as a matrix equation and so the method of least squares cannot be applied.

4.3 More General Surface Fitting

Least squares doesn't just work when the function is of one variable. The only requirement is that the function be linear in the variables we wish to find, so again if the function has the form:

$$f(x_1, \dots, x_k) = a_1 f_1(x_1, \dots, x_k) + a_2 f_2(x_1, \dots, x_k) + \dots + a_n f_n(x_1, \dots, x_k)$$

Where the $f_i(x_1, \dots, x_n)$ are known.

Example 4.4. Consider the points $(1, 1, 6)$, $(3, 1, 22)$, $(5, 4, 95)$, $(-2, 0, 10)$. Suppose we believe these points follow an elliptical paraboloid of the form $f(x, y) = ax^2 + by^2 + c$. This would mean that we have:

$$\begin{aligned}a(1)^2 + b(1)^2 + c &= 6 \\a(3)^2 + b(1)^2 + c &= 22 \\a(5)^2 + b(4)^2 + c &= 95 \\a(-2)^2 + b(0)^2 + c &= 10\end{aligned}$$

As a matrix equation:

$$\begin{bmatrix} 1 & 1 & 1 \\ 9 & 1 & 1 \\ 25 & 16 & 1 \\ 4 & 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 6 \\ 22 \\ 95 \\ 10 \end{bmatrix}$$

This has no solution but we can find a least-squares solution:

$$\begin{aligned} \begin{bmatrix} 1 & 1 & 1 \\ 9 & 1 & 1 \\ 25 & 16 & 1 \\ 4 & 0 & 1 \end{bmatrix}^T \begin{bmatrix} 1 & 1 & 1 \\ 9 & 1 & 1 \\ 25 & 16 & 1 \\ 4 & 0 & 1 \end{bmatrix} \widehat{\begin{bmatrix} a \\ b \\ c \end{bmatrix}} &= \begin{bmatrix} 1 & 1 & 1 \\ 9 & 1 & 1 \\ 25 & 16 & 1 \\ 4 & 0 & 1 \end{bmatrix}^T \begin{bmatrix} 6 \\ 22 \\ 95 \\ 10 \end{bmatrix} \\ \begin{bmatrix} 723 & 410 & 39 \\ 410 & 258 & 18 \\ 39 & 18 & 4 \end{bmatrix} \widehat{\begin{bmatrix} a \\ b \\ c \end{bmatrix}} &= \begin{bmatrix} 2619 \\ 1548 \\ 133 \end{bmatrix} \\ \widehat{\begin{bmatrix} a \\ b \\ c \end{bmatrix}} &\approx \begin{bmatrix} 2.0048 \\ 2.7083 \\ 1.5155 \end{bmatrix} \end{aligned}$$

So that the elliptical paraboloid $f(x, y) = 2.0048x^2 + 2.7083y^2 + 1.5155$ provides the best fit, where best means minimizes the sum of the squares of the vertical distances.

4.4 Real World Modeling and Predictions

4.4.1 Choosing a Function

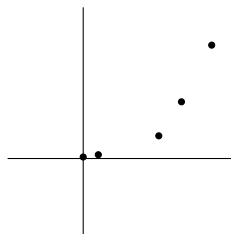
In real-world data modeling if you're using least-squares, especially with two variables, the first step would be to come up with a best-guess as to what the function might be. One thing to note is that the creation of a function (a model for the data) does not mean finding out what function the data follows. The data is just data, it doesn't necessarily obey any model at all. Whether a model is good or not is simply based upon whether it delivers on whatever we need it to do. Generally you would get data, build a model using some method for some reason, test it (on more data, or in the field) and then adjust accordingly.

Here's an example illustrating how you might start:

Example 4.5. Suppose you are analyzing average health insurance costs as a function of time. You collect the following sparse data in the form (t, d) where t is years after 2000 and d is average yearly cost taken over all individuals. In reality you'd have a LOT more data than this:

$$(0, 2), (2, 5), (10, 30), (13, 75), (15, 150)$$

The first thing you'd do is plot this, and you get the following where the axes ratio is not to scale:



There are any number of functions that might fit this data. The two that might leap out at you are exponential functions and quadratic functions. Might as well try both.

For an exponential function you might suggest $f(t) = a + be^t$ but then you note that e^{15} is really big, much bigger than 150, so perhaps a different base is better. For a guess you might want a base b so that $b^{15} \approx 150$ so $b \approx \sqrt[15]{150} \approx 1.3966$ so you figure that you'll give 1.4 a try and so you go with $f(t) = a + b(1.4)^t$.

If this were the case then you'd have:

$$\begin{aligned} f(0) &= a + b(1.4)^0 = 2 \\ f(2) &= a + b(1.4)^2 = 5 \\ f(10) &= a + b(1.4)^{10} = 30 \\ f(13) &= a + b(1.4)^{13} = 75 \\ f(15) &= a + b(1.4)^{15} = 150 \end{aligned}$$

And the corresponding matrix equation would be:

$$\begin{bmatrix} 1 & (1.4)^0 \\ 1 & (1.4)^2 \\ 1 & (1.4)^{10} \\ 1 & (1.4)^{13} \\ 1 & (1.4)^{15} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \\ 30 \\ 75 \\ 150 \end{bmatrix}$$

The least-squares solution to this is

$$\widehat{\begin{bmatrix} a \\ b \end{bmatrix}} = \begin{bmatrix} 1.8982 \\ 0.9463 \end{bmatrix}$$

which gives you the function:

$$f(t) = 1.8982 + 0.9463(1.4)^t$$

which has the least-squares error:

$$\left\| \begin{bmatrix} 1 & (1.4)^0 \\ 1 & (1.4)^2 \\ 1 & (1.4)^{10} \\ 1 & (1.4)^{13} \\ 1 & (1.4)^{15} \end{bmatrix} \begin{bmatrix} 1.8982 \\ 0.9463 \end{bmatrix} - \begin{bmatrix} 2 \\ 5 \\ 30 \\ 75 \\ 150 \end{bmatrix} \right\| \approx 2.7601$$

For a quadratic function you might suggest $f(t) = at^2 + bt + c$. If this were the case then you'd have:

$$\begin{aligned} f(0) &= a(0)^2 + b(0) + c = 2 \\ f(2) &= a(2)^2 + b(2) + c = 5 \\ f(10) &= a(10)^2 + b(10) + c = 30 \\ f(13) &= a(13)^2 + b(13) + c = 75 \\ f(15) &= a(15)^2 + b(15) + c = 150 \end{aligned}$$

And the corresponding matrix equation would be:

$$\begin{bmatrix} 0 & 0 & 1 \\ 4 & 2 & 1 \\ 100 & 10 & 1 \\ 169 & 13 & 1 \\ 225 & 15 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \\ 30 \\ 75 \\ 150 \end{bmatrix}$$

The least-squares solution to this is

$$\widehat{\begin{bmatrix} a \\ b \\ c \end{bmatrix}} = \begin{bmatrix} 1.3577 \\ -11.7172 \\ 10.9079 \end{bmatrix}$$

which gives you the function:

$$f(t) = 1.3577t^2 - 11.7172t + 10.9079$$

which has the least-squares error:

$$\left\| \begin{bmatrix} 0 & 0 & 1 \\ 4 & 2 & 1 \\ 100 & 10 & 1 \\ 169 & 13 & 1 \\ 225 & 15 & 1 \end{bmatrix} \begin{bmatrix} 1.3577 \\ -11.7172 \\ 10.9079 \end{bmatrix} - \begin{bmatrix} 2 \\ 5 \\ 30 \\ 75 \\ 150 \end{bmatrix} \right\| \approx 21.9904$$

You see that the least-squares error for the exponential is smaller and so your exponential function does a better job of modeling the data than your quadratic function does.

Note that this doesn't mean it's correct (that is, the data is not necessarily exponential), it just means out of two models, one fits the data better than the other.

4.4.2 Predicting

Once we have obtained our least squares function we can then use it to make predictions.

Example 4.6. In the previous health care example if you wanted to predict what the health care costs might be in 2020 you could use your quadratic model:

$$f(20) = 1.3577(20)^2 - 11.7172(20) + 10.9079 \approx 319.6439$$

If you're interested in when the costs might hit 400, you can solve $f(t) = 400$. Doing so yields both positive and negative solutions. We ignore the negative because we're focusing on the future with cost increasing as t increases. The positive solution is approximately 21.7851.

4.5 Matlab

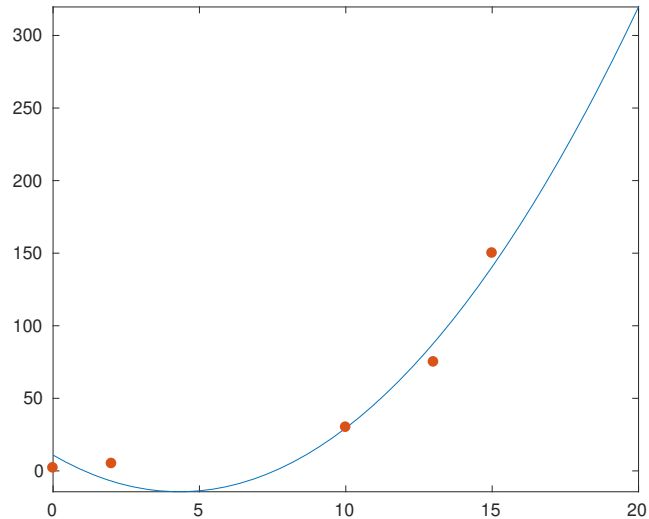
When you've found a function if you want to play with it in Matlab you can do so easily:

```
>> f(t) = 1.3577*t^2-11.7172*t+10.9079;
>> vpa(f(27),4)
ans =
684.3
>> vpa(solve(f(t)==500),4)
ans =
-15.15
23.78
```

It can be useful to plot a resulting function along with a line. The easiest way to do this is with the `hold on` command, which holds on to the current figure when you draw the next figure. The following code:

```
>> syms f(t)
>> f(t) = 1.3577*t^2-11.7172*t+10.9079;
>> pts = transpose([0,2;2,5;10,30;13,75;15,150]);
>> fplot(f,[0,20])
>> hold on
>> scatter(pts(1,:),pts(2:,:),'filled')
```

Shows the following:



A few things to note: The `fplot` command does a function plot and we've given it `[0,20]` for the domain. The `scatter` command does a scatter plot and requires all the x -coordinates followed by the y -coordinates. The matrix `pts` contains all the points with each point being a column, so then `pts(1,:)` pulls out the first row (the `:` gets the entire column) which consists of the x -coordinates and `pts(2,:)` pulls out the second row which consists of the y -coordinates.

4.6 Exercises

Exercise 4.1. Given the points $(-1, 0)$, $(1, 2)$, $(2, 2)$

(a) Find the least-squares line $f(x) = mx + b$ for the points.

- (b) Plot the points and the line and explain using your picture what exactly has been minimized.

Exercise 4.2. Given the points:

$$(1, 1), (5, 2), (6, 2), (8, 3)$$

- (a) Find the least-squares line $f(x) = mx + b$ for the points.
(b) Plot $f(x)$ along with the points.
(c) Use $f(x)$ to estimate y when $x = 20$.

Exercise 4.3. Given the points:

$$(-1, 3), (0, 1), (1, 2), (3, 9)$$

- (a) Find the least-squares parabola $f(x) = ax^2 + bx + c$ for the points.
(b) Plot $f(x)$ along with the points.
(c) Use $f(x)$ to estimate all x so that $f(x) = 10$.

Exercise 4.4. Given the points:

$$(-2, 0), (0, 3), (4, 4)$$

- (a) Find the least-squares exponential $f(x) = ae^x + b$ for the points.
(b) Plot $f(x)$ along with the points.
(c) Use $f(x)$ to estimate $f'(1)$.

Exercise 4.5. Given the points:

$$(-2, 6.3), (3, 1.2), (5, 7.1), (8, -2.8), (9, -0.05)$$

- (a) Find the least-squares $f(x) = a + b \sin x$ for the points.
(b) Plot $f(x)$ along with the points.
(c) Use $f(x)$ to estimate $f(\pi/2)$.

Exercise 4.6. Given the points:

$$(-3, -2, 45), (2, -2, 30), (0, 1, 6), (-2, 3, 55), (6, 5, 230)$$

- (a) Find the least-squares paraboloid $f(x, y) = ax^2 + by^2$ for the points.
(b) Use $f(x, y)$ to estimate $f(3, 5)$.

Exercise 4.7. Here is an interesting question - given the points $(0, 0)$, $(0, 1)$, $(1, 1)$ if we're looking for a best-fit line it's possible to look both for $y = mx + b$ and for $x = ny + c$. Neither has an exact solution but both have least-squares solution. Find each of these. Show that these don't yield the same line. Plot the points and both lines. From a geometric perspective of minimizing distance from the line, what is going on here?

Exercise 4.8. Suppose you would like to estimate the orbit of a certain object around the origin. Observations are made of both an angle and a distance. You collect five observations as follows where the first value is degrees and the second is in millions of miles:

$$(23^\circ, 152), (50^\circ, 135), (100^\circ, 102), (110^\circ, 110), (152^\circ, 137)$$

The equation of an ellipse in polar coordinates can be given by the following for some A and B :

$$Ar^2 \cos^2 \theta + Br^2 \sin^2 \theta = 1$$

- Find the least-squares best-fit ellipse.
- Use this to predict the distance of the object when $\theta = 225^\circ$.
- What is the furthest that the object ever gets from the origin?

Exercise 4.9. Repeating data points has an impact on the method of least squares. To visualize this, imagine we're trying to best-fit a straight line to a set of points. If a point appears more than once then the square of the distance to the line is being counted more than once and hence carries more weight in the method. To test this out find the least-squares line which best fits each of the following sets of points. Which line is closer to the point $(3, 2)$?

- The points $(1, 1)$, $(2, 1)$, $(3, 2)$
- The points $(1, 1)$, $(2, 1)$, $(3, 2)$, $(3, 2)$

Exercise 4.10. Consider the set of $n + 2$ points:

$$(1, 1), (2, 1), \underbrace{(3, 2), (3, 2), \dots, (3, 2)}_{n \text{ times}}$$

Suppose you wish to best-fit these to a line $y = mx + b$ using least-squares.

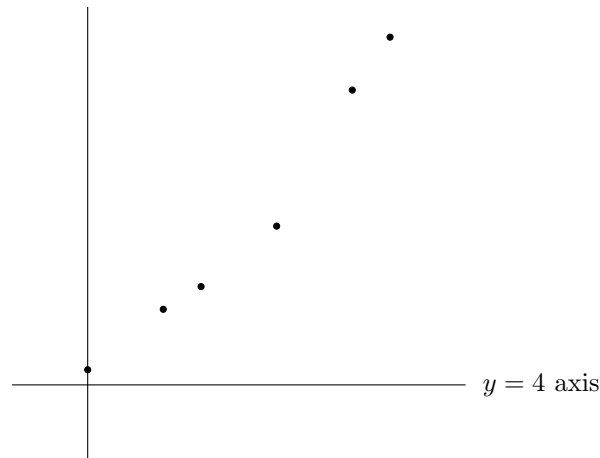
- Write down the corresponding matrix equation.
- Solve for \hat{x}_n using the method of least squares. Make sure you simplify; the answer should not be complicated.
- Find $\lim_{n \rightarrow \infty} \hat{x}_n$.

- (d) The line corresponding to your answer in (c) passes through $(3, 2)$. Why does this make sense?

Exercise 4.11. This problem loosely follows the data modeling example from class. Suppose you collect the following data points:

$$(0, 4.2), (2, 5), (3, 5.3), (5, 6.1), (7, 7.9), (8, 8.6)$$

When you plot these you see:



- Use least-squares to fit the function $f(x) = mx + b$.
- Use least-squares to fit the function $f(x) = ax^2 + bx + c$.
- If the data were to fit the function $f(x) = a + bc^x$, make an educated guess for c and then use least-squares to fit the function. Hint: One idea might be to ignore a and b and suggest that $f(x) \approx c^x$ especially for bigger x , but there are other options for guessing c .
- Calculating the least-squares error for each, which seems to provide the best fit?
- Use that function to predict $f(10)$.
- Use that function to predict which x would yield $f(x) = 50$.

Exercise 4.12. The following sets of points each approximately follows a familiar function which is linear in some unknowns. First plot the points. Then make an educated sensible guess as to the form of the function. Finally use the method of least squares to find a best-fit function and estimate the y -value corresponding to the given x value. The problems work from easier to more difficult.

(a) Two unknowns, $x = 10$.

$$(-1, 8.5), (1, 2.5), (2, 0.53), (3, -1.5), (6, -7.4), (7, -9.5)$$

(b) Two unknowns, $x = 20$.

$$(-3, 20), (-1, 3.5), (1, 3.6), (2, 9.5), (5, 52), (7, 100)$$

(c) Three unknowns, $x = 2$.

$$(-1, -5.4), (0, -1.9), (1, 0.51), (3, 2.5), (4, 2.0), (6, -1.9)$$

(d) Two unknowns, $x = -10$.

$$(-4, 6.1), (-1, 2.9), (0, 4.5), (2, 6.3), (6, 4.0), (7, 5.8), (8, 6.6), (10, 3.5), (11, 2.6)$$

(e) Three unknowns, $x = 20$.

$$(-2, -0.69), (-1, 1.5), (0, 2.5), (1, 1.8), (2, 0.08), (4, 0), (5, 2.2), (6, 3.7), (8, 1.9), (9, 0.49)$$

Exercise 4.13. For which of the following function templates will the method of least squares work and for which will it not. Explain. For one of the ones for which it will not work cite an example and show in detail what goes wrong. Your answer to this second part should touch on the issue of linear vs nonlinear systems.

(a) $f(x) = ax^2 + bx^c$

(b) $f(x) = ae^x + bx$

(c) $f(x) = e^{ax} + bx$

(d) $f(x) = a \sin(x) + b \cos(x) + c$

(e) $f(x) = a \sin(bx) + c$