EECS 16A Designing Information Devices and Systems I
Spring 2018 Discussion 12B

1. Ohm's Law With Noise

We are trying to measure the resistance of a black box. We apply various i_{test} currents and measure the ouput voltage v_{test}. Sometimes, we are quite fortunate to get nice numbers. Oftentimes, our measurement tools are a little bit noisy, and the values we get out of them are not accurate. However, if the noise is completely random, then the effect of it can be averaged out over many samples. Say that we repeat our test many times.

Test	i_{test} (mA)	v_{test} (V)
1	10	21
2	3	7
3	-1	-2
4	5	8
5	-8	-15
6	-5	-11

(a) Plot the measured voltage as a function of the current.

Answer:

Notice that these points *do not* lie on a line!

(b) Suppose we stack the currents and voltages to get $\vec{I} = \begin{bmatrix} 10 \\ 3 \\ -1 \\ 5 \\ -8 \\ -5 \end{bmatrix}$ and $\vec{V} = \begin{bmatrix} 21 \\ 7 \\ -2 \\ 8 \\ -15 \\ -11 \end{bmatrix}$. Can you solve for R?

What conditions must $\vec{I}$ and $\vec{V}$ satisfy in order for us to solve for R?

Answer:

We cannot solve for R because $\vec{V}$ is not a scalar multiple of $\vec{I}$. In general, we need $\vec{V}$ to be a scalar multiple of $\vec{I}$ to be able to solve for R exactly (another linear algebraic way of saying this is that $\vec{V}$ is in the span of $\vec{I}$).

We know that the *physical* reason we are not able to solve for R is that we have imperfect observations of the voltage across the terminals, $\vec{V}$. Although these observations are imperfect, they are observations, all the same. Therefore, now that we know we cannot solve for R, a very pertinent goal would be to find a value of R that *approximates* the relationship between $\vec{I}$ and $\vec{V}$ as closely as possible.

Let's move on and see how we do this.

(c) Ideally, we would like to find R such that $\vec{V} = \vec{I}R$. If we cannot do this, we'd like to find a value of R that is the *best* solution possible, in the sense that $\vec{I}R$ is as "close" to $\vec{V}$ as possible. The idea of a best solution is subjective and dependent on the cost function we are using. One way of expressing this cost function in terms of R is to quantify the difference between each component of $\vec{V}$ (V_j) and each component of $\vec{I}R$ (I_jR) and add these "differences" up as follows:

$$\text{cost}(R) = \sum_{j=1}^{6}(V_j - I_jR)^2$$

Do you think this is a good cost function? Why or why not?

Answer:

For each point (I_j, V_j), we want $|V_j - I_jR|$ to be as small as possible. We can call this term the individual error term for this point.

One way of looking at the aggregate "error" in our fit is to add up the squares of the individual errors, so that all errors add up. If we did not square the differences, then a positive difference and a negative difference cancel each other out. This is precisely what we've done in the cost function.

(d) Show that you can also express the above cost function in vector form, that is,

$$\text{cost}(R) = \left\langle (\vec{V} - \vec{I}R), (\vec{V} - \vec{I}R) \right\rangle$$

Answer:

Let's define the error vector as

$$\vec{e} = \vec{V} - \vec{I}R.$$

Then, we observe that $e_j = V_j - I_jR$.

Therefore,

$$\begin{aligned}
\text{cost}(R) &= \sum_{j=1}^{6}(V_j - I_jR)^2 \\
&= \sum_{j=1}^{6} e_j^2 \\
&= \|\vec{e}\|_2^2 \\
&= \langle \vec{e}, \vec{e} \rangle \\
&= \left\langle (\vec{V} - \vec{I}R), (\vec{V} - \vec{I}R) \right\rangle
\end{aligned}$$

(e) Find $\hat{R}$, the optimal R that minimizes $\text{cost}(R)$.

Hint: Use calculus and minimize the expression in part (c).

Answer:

First, note that

$$\frac{d\text{cost}(R)}{dR} = -2\sum_{j=1}^{6} I_j(V_j - I_jR)$$

For $R = \hat{R}$, we will have $\frac{d\text{cost}(R)}{dR} = 0$. This means that

$$-2\sum_{j=1}^{6} I_j(V_j - I_j\hat{R}) = 0,$$

which will ultimately give us

$$\hat{R} = \frac{\sum_{j=1}^{6} I_jV_j}{\sum_{j=1}^{6} I_j^2} = \frac{\left\langle \vec{I}, \vec{V} \right\rangle}{\|\vec{I}\|_2^2}$$

In our particular example, $\left\langle \vec{I}, \vec{V} \right\rangle = 448$ and $\|\vec{I}\|_2^2 = 224$. Therefore, we will get $\hat{R} = 2$ again!

(f) On your original IV plot, also plot the line $v = \hat{R}i$. Can you visually see why this line "fits" the data well? What if we had guessed $R = 3$? How well would we have done? What about $R = 1$?
Calculate the cost functions for each of these choices of R to validate your answer.

Answer:

When $V = 2I$, we have

$$\begin{aligned}\text{cost}(2) &= (21-2\cdot 10)^2 + (7-2\cdot 3)^2 + (-2-2\cdot(-1))^2 + \\ &\quad (8-2\cdot 5)^2 + (-15-2\cdot(-8))^2 + (-11-2\cdot(-5))^2 \\ &= 8.\end{aligned}$$

When $V = 3I$, we have

$$\begin{aligned}\text{cost}(3) &= (21-3\cdot 10)^2 + (7-3\cdot 3)^2 + (-2-3\cdot(-1))^2 + \\ &\quad (8-3\cdot 5)^2 + (-15-3\cdot(-8))^2 + (-11-3\cdot(-5))^2 \\ &= 232.\end{aligned}$$

When $V = I$, we have

$$\begin{aligned}\text{cost}(1) &= (21-1\cdot 10)^2 + (7-1\cdot 3)^2 + (-2-1\cdot(-1))^2 + \\ &\quad (8-1\cdot 5)^2 + (-15-1\cdot(-8))^2 + (-11-1\cdot(-5))^2 \\ &= 232.\end{aligned}$$

(g) Now, suppose that we add a new data point: $i_7 = 2\,\text{mA}$, $v_7 = 4\,\text{V}$. Will $\hat{R}$ increase, decrease, or remain the same? Why? What does that say about the line $v = \hat{R}i$?

Answer:

We can qualitatively see that $\hat{R}$ will remain 2. This is because we already obtained $\hat{R}$ to fit our previous data in the best way. Now, you should notice that this new piece of data (I_7, V_7) also lies exactly on the line $V = \hat{R}I$! Therefore, you have no reason to change $\hat{R}$. It is the best fit for the old data and will fit the new data anyway.

(h) Let's add another data point: $i_8 = 4\,\text{mA}, v_8 = 11\,\text{V}$. Will $\hat{R}$ increase, decrease, or remain the same? Why? What does that say about the line $v = \hat{R}i$?

Answer:

We can qualitatively see that $\hat{R}$ should be something greater than or equal to 2. This is because you have already obtained $\hat{R}$ to fit your previous data in the best way. Now, you notice that this new piece of data (I_8, V_8) also lies *above* the line $V = \hat{R}I$! Therefore, if you decreased $\hat{R}$, it would be a worse fit for the old data and the new data. You would increase $\hat{R}$ to find a better fit.

(i) Now your mischievous friend has hidden the black box. You want to know what the output voltage would be across the terminals if you applied 5.5 mA through the black box. What would your best guess be? (This is an example of estimation from machine learning! You have *learned* what is going on inside the black box by making observations, and now you're using what you learned to make estimates.)

Answer:

Hopefully, by now, it makes sense to the class that you will estimate $\hat{V} = 5.5\,\text{mA}\cdot\hat{R} = 5.5\,\text{mA}\cdot 2\,\text{k}\Omega = 11\,\text{V}$. This is an example of estimation from machine learning! You have learned what is going on inside the black box, that is, $\hat{R}$, by making observations of $\vec{I}$ and $\vec{V}$. Now, you are using what you have learned, $\hat{R}$, to estimate $\hat{V}$ for new values of I.

2. Polynomial Fitting

Notice that least squares can only be applied to linear systems. Suppose that we have a vector $\vec{x}$ and a vector $\vec{y}$, and $\vec{y}[n] = f(\vec{x}[n])$. We would like to approximate f using least squares, where f is not necessarily a linear function.

(a) Suppose that $y = ax + b$. Set this up as a least squares problem. What are the elements in the matrix **A**?

Answer:

$$\begin{bmatrix} \vec{x} & \vec{1} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \vec{y}$$

Let's try an example. Say we know that the output, y, is a quartic polynomial in x. This means that we know that y and x are related as follows:

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4$$

We're also given the following observations:

x	y
0.0	24.0
0.5	6.61
1.0	0.0
1.5	−0.95
2.0	0.07
2.5	0.73
3.0	−0.12
3.5	−0.83
4.0	−0.04
4.5	6.42

(a) What are the unknowns in this question? What are we trying to solve for?

Answer:

The unknowns are a_0, a_1, a_2, a_3, and a_4. They are also what we are trying to solve for.

(b) Can you write an equation corresponding to the first observation (x_0, y_0), in terms of a_0, a_1, a_2, a_3, and a_4? What does this equation look like? Is it linear?

Answer:

Plugging (x_0, y_0) into the expression for y in terms of x, we get

$$24 = a_0 + a_1 \cdot 0 + a_2 \cdot 0^2 + a_3 \cdot 0^3 + a_4 \cdot 0^4$$

You can see that this equation is linear in a_0, a_1, a_2, a_3, and a_4.

(c) Now, write a system of equations in terms of a_0, a_1, a_2, a_3, and a_4 using *all of the observations*.

Answer:

Write the next equation using the second observation. You will now get:

$$6.61 = a_0 + a_1 \cdot (0.5) + a_2 \cdot (0.5)^2 + a_3 \cdot (0.5)^3 + a_4 \cdot (0.5)^4$$

And for the third:

$$0.0 = a_0 + a_1 \cdot (1) + a_2 \cdot 1^2 + a_3 \cdot 1^3 + a_4 \cdot 1^4$$

Do you see a pattern? Let's write the entire system of equations in terms of a matrix now.

$$\begin{bmatrix} 1 & 0 & 0^2 & 0^3 & 0^4 \\ 1 & 0.5 & (0.5)^2 & (0.5)^3 & (0.5)^4 \\ 1 & 1 & 1^2 & 1^3 & 1^4 \\ 1 & 1.5 & (1.5)^2 & (1.5)^3 & (1.5)^4 \\ 1 & 2 & 2^2 & 2^3 & 2^4 \\ 1 & 2.5 & (2.5)^2 & (2.5)^3 & (2.5)^4 \\ 1 & 3 & 3^2 & 3^3 & 3^4 \\ 1 & 3.5 & (3.5)^2 & (3.5)^3 & (3.5)^4 \\ 1 & 4 & 4^2 & 4^3 & 4^4 \\ 1 & 4.5 & (4.5)^2 & (4.5)^3 & (4.5)^4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} = \begin{bmatrix} 24 \\ 6.61 \\ 0.0 \\ -0.95 \\ 0.07 \\ 0.73 \\ -0.12 \\ -0.83 \\ -0.04 \\ 6.42 \end{bmatrix}$$

(d) Finally, solve for a_0, a_1, a_2, a_3, and a_4 using IPython. You have now found the quartic polynomial that best fits the data!

Answer:

Let **D** be the big matrix from the previous part.

$$\vec{a} = (\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T\vec{y} = \begin{bmatrix} 24.00958042 \\ -49.99515152 \\ 35.0039627 \\ -9.99561772 \\ 0.99841492 \end{bmatrix}$$

It turns out that the actual parameters for the polynomial equation were:

$$\vec{a} = \begin{bmatrix} 24 \\ -50 \\ 35 \\ -10 \\ 1 \end{bmatrix}$$

(Remember that our observations were noisy.)

Therefore, we have actually done pretty well with the least squares estimate!