EECS 16A Designing Information Devices and Systems I
Spring 2022 Homework 13

This homework is due April 29, 2022, at 23:59.
Self-grades are due May 2, 2022, at 23:59.

Submission Format
Your homework submission should consist of **one** file.

- `hw13.pdf`: A single PDF file that contains all of your answers (any handwritten answers should be scanned).

Submit the file to the appropriate assignment on Gradescope.

1. Course Evaluation

Please fill out the course evaluation for EECS 16A. If at least 70% of the class fills out the course evaluation, everyone will receive one point of extra credit on the final, and if at least 80% fills it out, then everyone will receive two extra points.

Also please fill out this study group feedback survey found here: https://docs.google.com/forms/d/e/1FAIpQLSf_rgvJmmUZCwprNsvcT-ggqh1xTWIyk6flbh9FiBzQShVHig/viewform?usp=sf_link

2. Reading Assignment

For this homework, please read Note 22 (Trilateration and Correlation) and Note 23 (Least Squares).

3. Audio File Matching

Learning Goal: *This problem motivates the application of correlation for pattern matching applications such as Shazam.*

Many audio processing applications rely on representing audio files as vectors, referred to as audio *signals*. Every component of the vector determines the sound we hear at a given time. We can use inner products to determine if a particular audio clip is part of a longer song, similar to an application like *Shazam*.

Let us consider a very simplified model for an audio signal, $\vec{x}$. At each timestep k, the audio signal can be either $x[k] = -1$ or $x[k] = 1$.

(a) Say we want to compare two audio files of the same length N to decide how similar they are. First, consider two vectors that are exactly identical, namely $\vec{x}_1 = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T$ and $\vec{x}_2 = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T$. What is the inner product of these two vectors? What if $\vec{x}_1 = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T$ but $\vec{x}_2$ oscillates between 1 and -1? Assume that N, the length of the two vectors, is an even number.
Use this to suggest a method for comparing the similarity between a generic pair of length-N vectors.
Solution:
The inner product of $\vec{x}_1 = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T$ and $\vec{x}_2 = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T$ is $\vec{x}_1 \cdot \vec{x}_2 = N$. The inner product of $\vec{x}_1 = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T$ and $\vec{x}_2 = \begin{bmatrix} 1 & -1 & 1 & -1 & \cdots & 1 & -1 \end{bmatrix}^T$ is $\vec{x}_1 \cdot \vec{x}_2 = 0$ when the vector length

is even. To see this, take the sum of the first two terms of each vector.

$$\vec{x}_{1,1} \cdot \vec{x}_{2,1} + \vec{x}_{1,2} \cdot \vec{x}_{2,2} = 1 \cdot 1 + 1 \cdot -1 = 0$$

This yields zero, and repeats multiple times, leading to a total sum of 0. To compare two vectors of length N composed of 1 and -1, we take the inner product of the two vectors, a large inner product means the vectors have a similar direction.

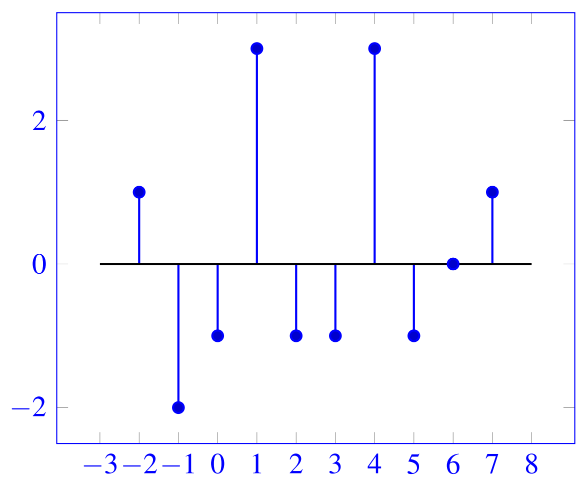
In many circumstances, an inner product with a very large negative value would mean the vectors are very different, but it turns out that humans are unable to perceive the sign of sound, so two sounds vectors $\vec{x}$ and $-\vec{x}$ sound exactly the same. As a result, for this problem we are interested in is the **absolute value** of the dot product, but in many other problems, we will interpret a large negative dot product as very different vectors. Don't take off points in parts (a), (b), or (c) if you didn't mention the absolute value.

(b) Next, suppose we want to find a short audio clip in a longer one. We might want to do this for an application like *Shazam*, which is able to identify a song from a short clip. Consider the vector of length 8, $\vec{x} = \begin{bmatrix} -1 & 1 & 1 & -1 & 1 & 1 & -1 & 1 \end{bmatrix}^T$.

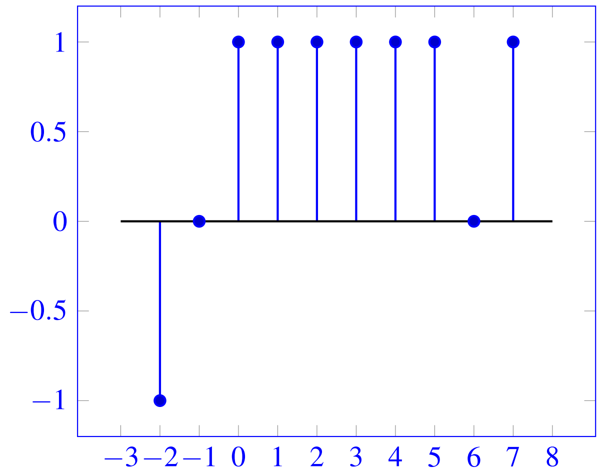
We want to find the short segment $\vec{y} := \begin{bmatrix} y[0] & y[1] & y[2] \end{bmatrix}^T = \begin{bmatrix} 1 & 1 & -1 \end{bmatrix}^T$ in the longer vector. To do this, perform the linear cross correlation between these two finite length sequences and identify at what shift(s) the linear cross correlation is maximized. Apply the same technique to identify what shift(s) gives the best match for $\vec{y} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T$.

(If you wish, you may use iPython to do this part of the question, but you do not have to.)

Solution:



The above plot is $\text{corr}_{\vec{x}}(\vec{y})[k]$ where $\vec{y} = \begin{bmatrix} 1 & 1 & -1 \end{bmatrix}^T$. At shifts 1 and 4 the cross correlation is its *maximum possible value*, 3. These are both good matches.



The above plot is $\text{corr}_{\vec{x}}(\vec{y})[k]$ where $\vec{y} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T$. At shifts 0 through 5 the cross correlation is only 1. There is not a really good match like before.

(c) Now suppose our audio vector is represented using integers beyond simply just 1 and -1. Find the short audio clip $\vec{y} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}^T$ in the song given by $\vec{x} = \begin{bmatrix} 1 & 2 & 3 & 1 & 2 & 2 & 3 & 10 \end{bmatrix}^T$. Where do you expect to see the peak in the correlation of the two signals? Is the peak where you want it to be, i.e. does it pull out the clip of the song that you intended? Why?

(If you wish, you may use iPython to do this part of the question, but you do not have to.)

Solution:

Applying the technique in part (b), we get the best match to be $\begin{bmatrix} 2 & 3 & 10 \end{bmatrix}^T$ as this has the largest dot product with $\vec{y} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}^T$. This is not where we expect to see the peak, as we observe the short audio clip $\vec{y}$ appears at the beginning of the song.

This happens because the volume at the end of the song is louder than the beginning of the song. Despite the angle not matching as well, the louder volume causes the linear cross correlation to be larger.

(d) Let us think about how to get around the issue in the previous part. We applied cross-correlation to compare segments of $\vec{x}$ of length 3 (which is the length of $\vec{y}$) with $\vec{y}$. Instead of directly taking the cross correlation, we want to normalize each inner product computed at each shift by the magnitudes of both segments, i.e. we want to consider the inner product $< \frac{\vec{x}_k}{\|\vec{x}_k\|}, \frac{\vec{y}}{\|\vec{y}\|} >$, where $\vec{x}_k$ is the length 3 segment starting from the k-th index of $\vec{x}$. This is referred to as normalized cross correlation. Using this procedure, now which segment matches the short audio clip best?

Solution: Using the normalized cross correlation procedure, the best match for the short audio clip is at the 0^{th} shift and it perfectly matches the clip.

(e) We can use this on a more ‘realistic’ audio signal – refer to the IPython notebook, where we use normalized cross-correlation on a real song. Run the cells to listen to the song we are searching through, and add a simple comparison function `vector_compare` to find where in the song the clip comes from. Running this may take a couple minutes on your machine, but note that this computation can be highly optimized and run super fast in the real world! Also note that this is not exactly how Shazam works, but it draws heavily on some of these basic ideas.

Solution:

See `sol13.ipynb`.

4. Mechanical Projections

Learning Goal: *The objective of this problem is to practice calculating projection of a vector and the corresponding squared error.*

(a) Find the projection of $\vec{b} = \begin{bmatrix} 3 \\ 2 \\ -1 \end{bmatrix}$ onto $\vec{a} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$. What is the squared error between the projection and $\vec{b}$, i.e. $\|e\|^2 = \|\text{proj}_{\vec{a}}(\vec{b}) - \vec{b}\|^2$?

Solution:

$$\text{proj}_{\vec{a}}(\vec{b}) = \frac{\langle \vec{b}, \vec{a} \rangle}{\|\vec{a}\|^2} \vec{a} = \frac{\vec{b}^T \vec{a}}{\|\vec{a}\|^2} \vec{a} \tag{1}$$

First, compute $\|\vec{a}\|^2 = \langle \vec{a}, \vec{a} \rangle = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = 2$.

Second, compute $\langle \vec{b}, \vec{a} \rangle = \begin{bmatrix} 3 & 2 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = 2$.

Plugging in, $\text{proj}_{\vec{a}}(\vec{b}) = \frac{2\vec{a}}{2} = \vec{a}$.

The squared error between $\vec{b}$ and its projection onto $\vec{a}$ is $\|e\|^2 = \|\vec{a} - \vec{b}\|^2 = 12$.

(b) Find the projection of $\vec{b} = \begin{bmatrix} 1 \\ 4 \\ -5 \end{bmatrix}$ onto the column space of $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$. What is the squared error between the projection and $\vec{b}$, i.e. $\|e\|^2 = \|\text{proj}_{\text{Col}(\mathbf{A})}(\vec{b}) - \vec{b}\|^2$?

Solution: Let $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$ and $\vec{x} \in \mathbb{R}^2$ such that the projection of $\vec{b}$ onto the column space of $\mathbf{A}$ is $\mathbf{A}\vec{x}$.

We will compute $\hat{\vec{x}}$ by solving the following least squares problem,

$$\min_{\vec{x}} \|\mathbf{A}\vec{x} - \vec{b}\|^2 \tag{2}$$

The solution yields,

$$\hat{\vec{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\vec{b} \tag{3}$$

$$= \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}^T \begin{bmatrix} 1 \\ 4 \\ -5 \end{bmatrix} \tag{4}$$

$$= \begin{bmatrix} 0.5 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -4 \\ 4 \end{bmatrix} \tag{5}$$

$$= \begin{bmatrix} -2 \\ 4 \end{bmatrix} \tag{6}$$

Plugging in, the projection of $\vec{b}$ onto the column space of $\mathbf{A}$ is $\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -2 \\ 4 \end{bmatrix} = \begin{bmatrix} -2 \\ 4 \\ -2 \end{bmatrix}$.

The squared error between the projection and $\vec{b}$ is $\|\vec{e}\|^2 = \left\| \begin{bmatrix} -2 \\ 4 \\ -2 \end{bmatrix} - \begin{bmatrix} 1 \\ 4 \\ -5 \end{bmatrix} \right\|^2 = 18$.

5. Mechanical Trilateration

Learning Goal: *The objective of this problem is to practice using trilateration to find the position based on the distance measurements and known beacon locations.*

Trilateration is the problem of finding one's coordinates given distances from known beacon locations. For each of the following trilateration problems, you are given 3 beacon locations ($\vec{s}_1, \vec{s}_2, \vec{s}_3$) and the corresponding distance (d_1, d_2, d_3) from each beacon to your location.

(a) $\vec{s}_1 = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$, $d_1 = 5$, $\vec{s}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$, $d_2 = 2$, $\vec{s}_3 = \begin{bmatrix} -11 \\ 6 \end{bmatrix}$, $d_3 = 13$. First, use any graphing calculator or ipython to graph the set of constraints given by ($\vec{s}_1, \vec{s}_2, \vec{s}_3$) and ($d_1, d_2, d_3$), and take note of the number of solutions, or possible locations that you could be. Then use trilateration to find your location or possible locations. If a solution does not exist, state that it does not.

Solution: From graphing these equations we can see there is a single point of intersection, or in other words, a single possible solution to our location.

Now, we show a general approach to the trilateration problem, so that we can immediately write the linear system of equations for all three parts and solve for our solution algebraically. However, if you solved directly using concrete values, give yourself full credit.

$$\|\vec{x} - \vec{s}_1\|^2 = d_1^2$$
$$\|\vec{x} - \vec{s}_2\|^2 = d_2^2$$
$$\|\vec{x} - \vec{s}_3\|^2 = d_3^2$$

Now, let's show this algebraically with trilateration. We can expand each left hand side out in terms of the definition of the norm:

$$\|\vec{x} - \vec{s}_i\|^2 = \langle \vec{x} - \vec{s}_i, \vec{x} - \vec{s}_i \rangle = (\vec{x} - \vec{s}_i)^T (\vec{x} - \vec{s}_i)$$

$$\vec{x}^T\vec{x} - 2\vec{x}^T\vec{s}_1 + \vec{s}_1^T\vec{s}_1 = d_1^2$$
$$\vec{x}^T\vec{x} - 2\vec{x}^T\vec{s}_2 + \vec{s}_2^T\vec{s}_2 = d_2^2$$
$$\vec{x}^T\vec{x} - 2\vec{x}^T\vec{s}_3 + \vec{s}_3^T\vec{s}_3 = d_3^2$$

Finally, take one equation and subtract it from the other two to get a system of linear equations in $\vec{x}$:

$$2\vec{x}^T\vec{s}_3 - 2\vec{x}^T\vec{s}_1 = d_1^2 - d_3^2 + \vec{s}_3^T\vec{s}_3 - \vec{s}_1^T\vec{s}_1$$
$$2\vec{x}^T\vec{s}_3 - 2\vec{x}^T\vec{s}_2 = d_2^2 - d_3^2 + \vec{s}_3^T\vec{s}_3 - \vec{s}_2^T\vec{s}_2$$

We can express as a matrix equation in $\vec{x}$:

$$\begin{bmatrix} 2(\vec{s}_3 - \vec{s}_1)^T \\ 2(\vec{s}_3 - \vec{s}_2)^T \end{bmatrix} \vec{x} = \begin{bmatrix} d_1^2 - d_3^2 + \|\vec{s}_3\|^2 - \|\vec{s}_1\|^2 \\ d_2^2 - d_3^2 + \|\vec{s}_3\|^2 - \|\vec{s}_2\|^2 \end{bmatrix}$$

We have that:

$$2(\vec{s}_3 - \vec{s}_1) = \begin{bmatrix} -30 \\ 2 \end{bmatrix}$$
$$2(\vec{s}_3 - \vec{s}_2) = \begin{bmatrix} -24 \\ 14 \end{bmatrix}$$
$$d_1^2 - d_3^2 + \|\vec{s}_3\|^2 - \|\vec{s}_1\|^2 = 25 - 169 + 157 - 41 = -28$$
$$d_2^2 - d_3^2 + \|\vec{s}_3\|^2 - \|\vec{s}_2\|^2 = 4 - 169 + 157 - 2 = -10$$

Which gives us the system $\begin{bmatrix} -30 & 2 \\ -24 & 14 \end{bmatrix} \vec{x} = \begin{bmatrix} -28 \\ -10 \end{bmatrix}$ with solution $\vec{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

A solution existing for this system of linear equations does not necessarily guarantee consistency of the system of nonlinear equations, but we can validate:

$$\left\| \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 4 \\ 5 \end{bmatrix} \right\|^2 = \left\| \begin{bmatrix} -3 \\ -4 \end{bmatrix} \right\|^2 = 25 = d_1^2$$
$$\left\| \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\|^2 = \left\| \begin{bmatrix} 0 \\ 2 \end{bmatrix} \right\|^2 = 4 = d_2^2$$
$$\left\| \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} -11 \\ 6 \end{bmatrix} \right\|^2 = \left\| \begin{bmatrix} 12 \\ -5 \end{bmatrix} \right\|^2 = 169 = d_3^2$$

(b) $\vec{s}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $d_1 = 5\sqrt{2}$, $\vec{s}_2 = \begin{bmatrix} 10 \\ 0 \end{bmatrix}$, $d_2 = 5\sqrt{2}$, $\vec{s}_3 = \begin{bmatrix} 5 \\ 0 \end{bmatrix}$, $d_3 = 5$. First, use any graphing calculator or ipython to graph the set of constraints given by $(\vec{s}_1, \vec{s}_2, \vec{s}_3)$ and (d_1, d_2, d_3), and take note of the number of solutions, or possible locations that you could be. Then use trilateration to find your location or possible locations. Why can't we precisely determine our location, even though we have the same number of measurements as part (a)? Can we use our original constraints to narrow down our set of possible solutions we got from trilateration?

Solution: Graphing our constraints gives us two points of intersection.

Now, let's try to algebraically solve for these points using trilateration. Using the linearization approach from part (a) we get:

$$2(\vec{s}_3 - \vec{s}_1) = \begin{bmatrix} 10 \\ 0 \end{bmatrix}$$
$$2(\vec{s}_3 - \vec{s}_2) = \begin{bmatrix} -10 \\ 0 \end{bmatrix}$$
$$d_1^2 - d_3^2 + \|\vec{s}_3\|^2 - \|\vec{s}_1\|^2 = 50 - 25 + 25 - 0 = 50$$
$$d_2^2 - d_3^2 + \|\vec{s}_3\|^2 - \|\vec{s}_2\|^2 = 50 - 25 + 25 - 100 = -50$$

Which gives us the system $\begin{bmatrix} 10 & 0 \\ -10 & 0 \end{bmatrix} \vec{x} = \begin{bmatrix} 50 \\ -50 \end{bmatrix}$ with solution $\vec{x} = \begin{bmatrix} 5 \\ \alpha \end{bmatrix}$. We can see that by having collinear beacons, we may not be able to precisely determine our location (short exercise: how does this relate to span and vector spaces?)

However, from the graph we know that not all values of α are valid, so we can plug our solution back into the third distance equation:

$$\left\| \begin{bmatrix} 5 \\ \alpha \end{bmatrix} - \begin{bmatrix} 5 \\ 0 \end{bmatrix} \right\|^2 = 5^2 \implies \alpha^2 = 25 \implies \alpha = \pm 5$$

The system of nonlinear equations is consistent with this solution. We do not have enough information to uniquely determine our location, but we know we are at either $\vec{x} = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$ or $\vec{x} = \begin{bmatrix} 5 \\ -5 \end{bmatrix}$.

(c) $\vec{s}_1 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$, $d_1 = 5$, $\vec{s}_2 = \begin{bmatrix} 0 \\ -2 \end{bmatrix}$, $d_2 = 2$, $\vec{s}_3 = \begin{bmatrix} -12 \\ 5 \end{bmatrix}$, $d_3 = 12$. First, use any graphing calculator or ipython to graph the set of constraints given by $(\vec{s}_1, \vec{s}_2, \vec{s}_3)$ and (d_1, d_2, d_3), and take note of the number of solutions, or possible locations that you could be. Then use trilateration to find your location or possible locations. If a solution does not exist, state that it does not.

Solution: Graphing our equations gives us no points of intersection, meaning that there will be no solutions.

Now, let's show this algebraically with trilateration. Using again what was shown in part (a) we have that:

$$2(\vec{s}_3 - \vec{s}_1) = \begin{bmatrix} -30 \\ 2 \end{bmatrix}$$

$$2(\vec{s}_3 - \vec{s}_2) = \begin{bmatrix} -24 \\ 14 \end{bmatrix}$$

$$d_1^2 - d_3^2 + \|\vec{s}_3\|^2 - \|\vec{s}_1\|^2 = 25 - 144 + 169 - 25 = 25$$

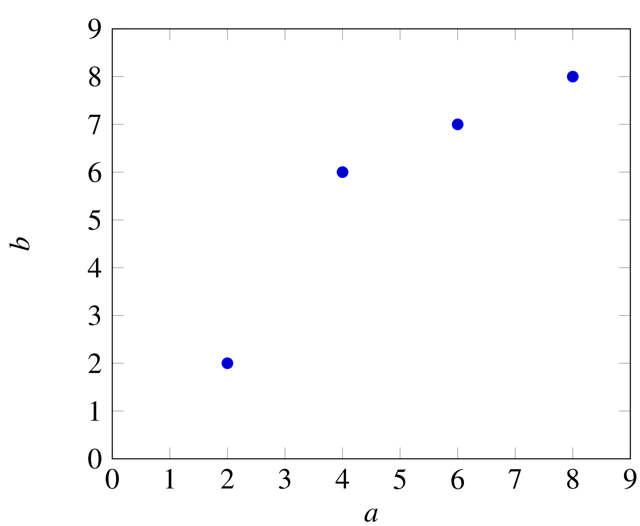
$$d_2^2 - d_3^2 + \|\vec{s}_3\|^2 - \|\vec{s}_2\|^2 = 4 - 144 + 169 - 4 = 25$$

Which gives us the system $\begin{bmatrix} -30 & 2 \\ -24 & 14 \end{bmatrix} \vec{x} = \begin{bmatrix} 25 \\ 25 \end{bmatrix}$. While a solution, $\vec{x} = \begin{bmatrix} -\frac{75}{93} \\ \frac{75}{186} \end{bmatrix}$, for this system of linear equations exists, it will yield inconsistent distances when substituted back into the nonlinear equations.

$$\|\vec{s}_1 - x\|^2 = (3 + \frac{75}{93})^2 + (4 - \frac{75}{186})^2 \neq d_1^2$$

Therefore there is no solution.

6. Mechanical: Linear Least Squares



a	2	4	6	8
b	2	6	7	8

(a) Consider the above data points. Find the linear model of the form

$$\vec{b} = \vec{a}x$$

that best fits the data, i.e. find the scalar value of x that minimizes the squared error

$$\|\vec{e}\|^2 = \left\| \begin{bmatrix} b_1 \\ \vdots \\ b_4 \end{bmatrix} - \begin{bmatrix} a_1 \\ \vdots \\ a_4 \end{bmatrix} x \right\|^2 = \|\vec{b} - \vec{a}x\|^2. \tag{7}$$

Note: by using this linear model, we are implicitly forcing the fit equation to go through the origin, i.e. $0 = x0$ for all x.

Do not use IPython for this calculation and show your work. Once you've computed x, compute the squared error between your model's prediction and the actual $\vec{b}$ values as shown in Equation 7. Plot the best fit line along with the data points to examine the quality of the fit. (It is okay if your plot of $\vec{b} = \vec{a}x$ is approximate.)

Reminder: $\hat{x} = (\vec{a}^T\vec{a})^{-1}\vec{a}^T\vec{b}$

Solution:

Define $\vec{a} = \begin{bmatrix} 2 & 4 & 6 & 8 \end{bmatrix}^T$ and $\vec{b} = \begin{bmatrix} 2 & 6 & 7 & 8 \end{bmatrix}^T$. Applying the linear least squares formula, we get

$$\begin{aligned} \hat{x} &= (\vec{a}^T\vec{a})^{-1}\vec{a}^T\vec{b} \\ &= \left(\begin{bmatrix} 2 \\ 4 \\ 6 \\ 8 \end{bmatrix}^T \begin{bmatrix} 2 \\ 4 \\ 6 \\ 8 \end{bmatrix} \right)^{-1} \begin{bmatrix} 2 \\ 4 \\ 6 \\ 8 \end{bmatrix}^T \begin{bmatrix} 2 \\ 6 \\ 7 \\ 8 \end{bmatrix} \\ &= (120)^{-1}(134) = 1.1167 \end{aligned}$$

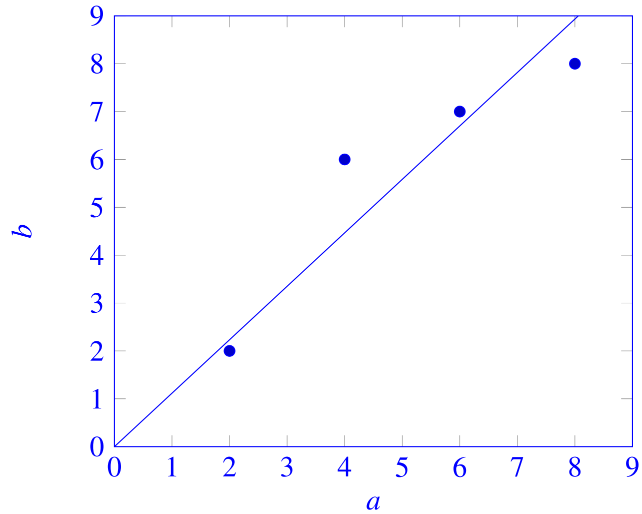
The error between the model's prediction and actual b values is

$$\begin{aligned} \vec{e} &= \vec{b} - \vec{\hat{b}} = \vec{b} - \hat{x}\vec{a} \\ &= \begin{bmatrix} 2 \\ 6 \\ 7 \\ 8 \end{bmatrix} - 1.1167 \begin{bmatrix} 2 \\ 4 \\ 6 \\ 8 \end{bmatrix} = \begin{bmatrix} -0.234 \\ 1.534 \\ 0.3 \\ -0.934 \end{bmatrix} \end{aligned}$$

and the sum of squared errors is

$$\vec{e}^T\vec{e} = 3.367$$

All Rights Reserved. This may not be publicly shared without explicit permission.



(b) You will notice from your graph that you can get a better fit by adding a b-intercept. That is we can get a better fit for the data by assuming a linear model of the form

$$\vec{b} = x_1\vec{a} + x_2.$$

In order to do this, we need to augment our $\mathbf{A}$ matrix for the least squares calculation with a column of 1's (do you see why?), so that it has the form

$$\mathbf{A} = \begin{bmatrix} a_1 & 1 \\ \vdots & \vdots \\ a_4 & 1 \end{bmatrix}.$$

Find x_1 and x_2 that minimize the squared error

$$\|\vec{e}\|^2 = \left\| \begin{bmatrix} b_1 \\ \vdots \\ b_4 \end{bmatrix} - \begin{bmatrix} a_1 & 1 \\ \vdots & \vdots \\ a_4 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right\|^2. \tag{8}$$

Do not use IPython for this calculation and show your work.

Reminder: $\hat{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\vec{b}$

Reminder: $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

Compute the squared error between your model's prediction and the actual $\vec{b}$ values as shown in Equation 8. Plot your new linear model. Is it a better fit for the data?

Solution:

Let $\vec{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^T$. Using the linear least squares formula with the new augmented $\mathbf{A}$ matrix, we calculate the optimal approximation of $\vec{x}$ as

$$\begin{aligned}
\vec{\hat{x}} &= \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\vec{b} \\
&= \left(\begin{bmatrix}2 & 1\\4 & 1\\6 & 1\\8 & 1\end{bmatrix}^T \begin{bmatrix}2 & 1\\4 & 1\\6 & 1\\8 & 1\end{bmatrix}\right)^{-1} \begin{bmatrix}2 & 1\\4 & 1\\6 & 1\\8 & 1\end{bmatrix}^T \begin{bmatrix}2\\6\\7\\8\end{bmatrix} \\
&= \begin{bmatrix}120 & 20\\20 & 4\end{bmatrix}^{-1} \begin{bmatrix}2 & 4 & 6 & 8\\1 & 1 & 1 & 1\end{bmatrix} \begin{bmatrix}2\\6\\7\\8\end{bmatrix} \\
&= \frac{1}{120(4)-20(20)} \begin{bmatrix}4 & -20\\-20 & 120\end{bmatrix} \begin{bmatrix}134\\23\end{bmatrix}
\end{aligned}$$

$$\vec{\hat{x}} = \begin{bmatrix}\hat{x}_1\\\hat{x}_2\end{bmatrix} = \begin{bmatrix}0.95\\1\end{bmatrix}$$

The linear model's prediction of $\vec{b}$ is given by

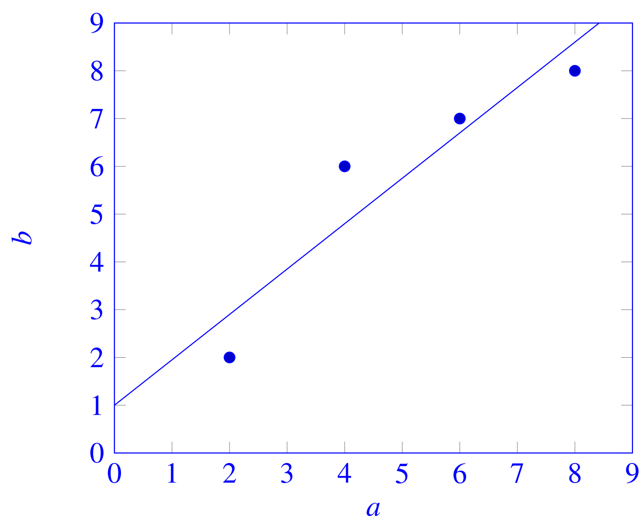
$$\vec{\hat{b}} = \mathbf{A}\vec{\hat{x}} = \begin{bmatrix}2 & 1\\4 & 1\\6 & 1\\8 & 1\end{bmatrix} \begin{bmatrix}0.95\\1\end{bmatrix} = \begin{bmatrix}2.9\\4.8\\6.7\\8.6\end{bmatrix}$$

and the error is given by

$$\vec{e} = \vec{b} - \vec{\hat{b}} = \begin{bmatrix}-0.9 & 1.2 & 0.3 & -0.6\end{bmatrix}^T$$

The summed squared error is

$$\vec{e}^T\vec{e} = 2.7$$



We can see both qualitatively from the plots and quantitatively from the sum of the squared errors that the fit is better with the b-intercept.

7. Proof: Least Squares

Let $\vec{\hat{x}}$ be the solution to a linear least squares problem.

$$\vec{\hat{x}} = \underset{\vec{x}}{\operatorname{argmin}} \left\| \vec{b} - \mathbf{A}\vec{x} \right\|^2$$

Show that the error vector $\vec{b} - \mathbf{A}\vec{\hat{x}}$ is orthogonal to the columns of $\mathbf{A}$ by direct manipulation (i.e. plug the formula for the linear least squares estimate into the error vector and then check if $\mathbf{A}^T$ times the vector is the zero vector.)

Solution:

We want to show that the error in the linear least squares estimate is orthogonal to the columns of the $\mathbf{A}$, i.e., we want to show that $\mathbf{A}^T(\vec{b} - \mathbf{A}\vec{\hat{x}})$ is the zero vector. Plugging in the linear least squares formula for $\vec{\hat{x}}$, we get

$$\begin{aligned}
\mathbf{A}^T\left(\vec{b} - \mathbf{A}\vec{\hat{x}}\right) &= \mathbf{A}^T\left(\vec{b} - \mathbf{A}\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\vec{b}\right) \\
&= \mathbf{A}^T\vec{b} - \mathbf{A}^T\mathbf{A}\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\vec{b} \\
&= \mathbf{A}^T\vec{b} - \mathbf{I}\mathbf{A}^T\vec{b} \\
&= \mathbf{A}^T\vec{b} - \mathbf{A}^T\vec{b} = \vec{0}
\end{aligned}$$

8. Homework Process and Study Group

Who did you work with on this homework? List names and student ID's. (In case you met people at homework party or in office hours, you can also just describe the group.) How did you work on this homework? If you worked in your study group, explain what role each student played for the meetings this week.

Solution:

I first worked by myself for 2 hours, but got stuck on problem 5. Then I met with my study group.

XYZ played the role of facilitator ... etc. We were still stuck on problem 5 so we went to office hours to talk about the problem.

Then I went to homework party for a few hours, where I finished the homework.