MODULE 12: LEAST-SQUARES

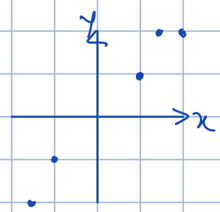
TOPIC 1: Least-Squares Problem

Least-Squares Solutions to a Linear System

Motivation: * A series of measurements are corrupted by random errors. How can the dominant trend be exacted from the measurements with random error?

Suppose we are given the data below.

x	-1.5	-1	1	1.5	2
y	-2	-1	1	2	2



Suppose we want to fit a straight line of the form $y = mx + b$ to our data. But how can we determine m and b?

Using the model $y = mx + b$,

$$\begin{aligned} m(-1.5) + b &= -2 \\ m(-1) + b &= -1 \\ m(1) + b &= 1 \\ m(1.5) + b &= 2 \\ m(2) + b &= 2 \end{aligned} \quad \Rightarrow \quad \begin{pmatrix} -1.5 & 1 \\ -1 & 1 \\ 1 & 1 \\ 1.5 & 1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} -2 \\ -1 \\ 1 \\ 2 \\ 2 \end{pmatrix}$$

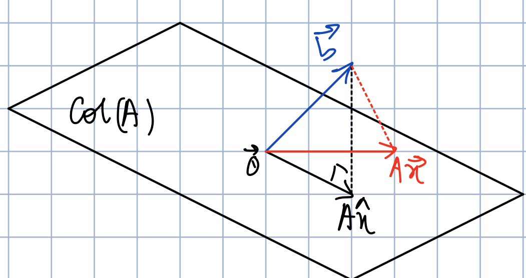
We have a system of the form $A\vec{x} = \vec{b}$. There is no solution to our system, because there is no line that passes through all data points. To find the line that best approximates our data we can use least-squares.

DEFINITION — Least-Squares Solution
↳ Let A be an $m \times n$ matrix. A least-squares solution to $A\vec{x} = \vec{b}$ is the solution $\hat{v}$ for which

$$\|\vec{b} - A\hat{x}\| \leq \|\vec{b} - A\vec{x}\| \ \forall \ \vec{x} \in \mathbb{R}^n.$$

In other words, we want to identify the $\vec{x}$ that minimizes $\|\vec{b} - A\vec{x}\|$, which we denote as $\hat{x}$.

A Geometric Interpretation:



The closest vector in Col A to $\vec{b}$ is $A\hat{x}$.

* If $\vec{b} \in \text{Col}(A)$, then $A\vec{x} = \vec{b}$ is consistent.
* In general, we seek $\hat{x}$ so that $A\hat{x}$ is as close to $\vec{b}$ as possible.

We will need a process to identify $\hat{x}$.

Exercise:

Five points in $\mathbb{R}^2$ with coordinates (x, y) are $(-1, 4)$, $(-0.5, 2)$, $(0, 1)$, $(0.5, 3)$ and $(1, -4)$. We want to construct a linear system to determine the coefficients c_1 and c_2 for the line

$$y = c_1 x + c_2.$$

that best fits the points.

For the first data point, $(-1, 4)$, we obtain the linear equation $-c_1 + c_2 = 4$.

For the second point, $(-0.5, 2)$, we can create an equation of the form $c_1 x + c_2 = y$. What value can we use for x that uses the second data point?

$$-0.5c_1 + c_2 = 2. \quad \therefore \underline{\underline{-0.5}}$$

For the third point, $(0, 1)$, we can create an equation of the form $c_1 x + c_2 = y$. What value can we use for y that uses the third data point?

$$0c_1 + c_2 = 1. \quad \therefore \underline{\underline{1}}$$

Now we want to construct a linear system to determine the coefficients c_1 and c_2 for the line

$$y = c_1 x + c_2$$

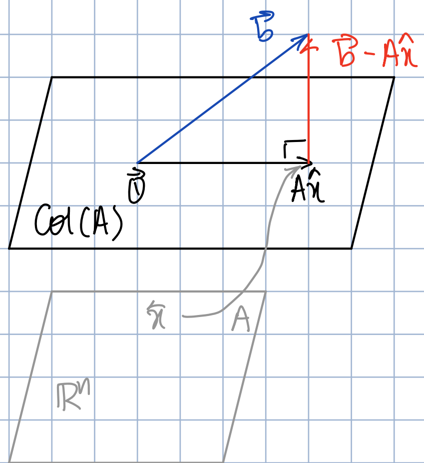
that best fits the points. Using data and our model $y = c_1 x + c_2$, we can construct a set of linear equations:

$$\begin{aligned} -c_1 + c_2 &= 4 \\ -\tfrac{1}{2}c_1 + c_2 &= 2 \\ 0c_1 + c_2 &= 1 \\ \tfrac{1}{2}c_1 + c_2 &= -3 \\ c_1 + c_2 &= -4. \end{aligned}$$

We can represent these equations as a linear system of the form $A\vec{x} = \vec{y}$. Find A.

$$A = \begin{pmatrix} -1 & 1 \\ -1/2 & 1 \\ 0 & 1 \\ 1/2 & 1 \\ 1 & 1 \end{pmatrix}$$

The Normal Equations



The least-squares solution $\hat{x}$ is in $\mathbb{R}^n$.

Recall orthogonal decomposition theorem:

If $A\hat{x}$ is the closest vector in Col A to $\vec{b}$, then $\vec{b} - A\hat{x}$ is orthogonal to Col A.

Also recall: $(\text{Col } A)^{\perp} = \text{Null } A^T$

$\therefore A^T(\vec{b} - A\hat{x}) = \vec{0} \Rightarrow A^T A\hat{x} = A^T\vec{b}.$

Theorem: Normal Equations for Least Squares
The least squares solutions to $A\vec{x}=\vec{b}$ coincide with the solutions to

$$A^TA\hat{x} = A^T\vec{b}$$

This linear system is referred to as the <u>Normal Equations</u>.

Theorem: Unique solutions for Least-Squares
Let A be any $m\times n$ matrix. These statements are equivalent.
↳ The columns of A are linearly independent.
↳ The matrix A^TA is invertible.
↳ The equation $A\vec{x}=\vec{b}$ has a unique least-squares solution for each $\vec{b}\in\mathbb{R}^m$.

If the above statements hold, the least squares solution is

$$\hat{x} = (A^TA)^{-1}A^T\vec{b}.$$

Example: Compute the least-squares solution to $A\vec{x}=\vec{b}$, where

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad \vec{b} = \begin{pmatrix} 2 \\ 2 \\ -1 \end{pmatrix}$$

$$A^TA = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}$$

$$A^T\vec{b} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \end{pmatrix}\begin{pmatrix} 2 \\ 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix}$$

The normal equations $A^TA\vec{x} = A^T\vec{b}$ are

$$\begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}\hat{x} = \begin{pmatrix} 3 \\ 3 \end{pmatrix}$$

$$\therefore \hat{x} = \frac{1}{6}\begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}\begin{pmatrix} 3 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 3/2 \end{pmatrix} \ast$$

Exercise:
Consider again the problem where we have a problem where we have five points in $\mathbb{R}^2$ with coordinates (x,y) are $(-1,4)$, $(-0.5,2)$, $(0,1)$, $(0.5,-3)$ and $(1,-4)$.

Now we want to solve a <u>linear</u> system to determine the coefficients c_1 and c_2 for the line

$$y = c_1x + c_2$$

that best fits the points. Using data and our model $y=c_1x+c_2$, we can construct a system of the form:

$$A\vec{x} = \vec{y}, \quad A = \begin{pmatrix} -1 & 1 \\ -1/2 & 1 \\ 0 & 1 \\ 1/2 & 1 \\ 1 & 1 \end{pmatrix}, \quad \vec{y} = \begin{pmatrix} 4 \\ 2 \\ 1 \\ -3 \\ -4 \end{pmatrix}$$

Find A^TA and $A^T\vec{y}$.

$$A^TA = \begin{pmatrix} -1 & -1/2 & 0 & 1/2 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 \\ -1/2 & 1 \\ 0 & 1 \\ 1/2 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 5/2 & 0 \\ 0 & 5 \end{pmatrix} *$$

$$A^T\vec{y} = \begin{pmatrix} -1 & -1/2 & 0 & 1/2 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 4 \\ 2 \\ 1 \\ -3 \\ -4 \end{pmatrix} = \begin{pmatrix} -4-1+0-3/2-4 \\ 4+2+1-3-4 \end{pmatrix} = \begin{pmatrix} -21/2 \\ 0 \end{pmatrix} *$$

Find c_1 and c_2.

Using $c_1x + c_2 = y$,

$$\begin{pmatrix} 5/2 & 0 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} -21/2 \\ 0 \end{pmatrix}.$$

$5/2\, c_1 = -21/2 \Rightarrow c_1 = -21/5 *$
$5c_2 = 0 \Rightarrow c_2 = 0 *$

QR and Least-Squares

Motivation:

The normal equations give us a method for calculating a least-squares solution, $\hat{x}$.

$$A^TA\hat{x} = A^T\vec{b}.$$

* These equations are often used in situations where A and $\vec{b}$ are constructed using measured data that contain errors.

Theorem: Least-Squares and QR
If $A \in \mathbb{R}^{m\times n}$ has linearly independent columns, then $A = QR$, and for every $\vec{b} \in \mathbb{R}^m$, $A\vec{x} = \vec{b}$ has the unique least-squares solution.

$$R\hat{x} = Q^T\vec{b}.$$

Note → R is invertible and upper triangular. $\Rightarrow R\hat{x} = Q^T\vec{b}$ may be solved with back-substitution.

Proof: If $A = QR$ and the normal equations are $A^TA\hat{x} = A^T\vec{b}$, then

$$(QR)^T QR\hat{x} = (QR)^T\vec{b}$$
$$R^TQ^TQR\hat{x} = R^TQ^T\vec{b}$$
$$R^TR\hat{x} = R^TQ^T\vec{b} \quad \because Q^TQ = I$$

Since R is invertible,

$$\therefore R\hat{x} = Q^T\vec{b} \quad \text{(QED)}$$

Process of Using QR to Solve a Least-Squares Problem

Given A and $\vec{b}$, we can use the following process to compute the least-squares solution, $\hat{x}$, with the QR decomposition.

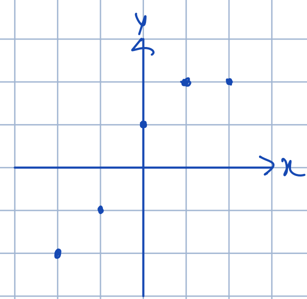
1. Construct QR decomposition of A
   (a) obtain orthonormal basis for Col A (can use Gram-Schmidt)
   (b) obtain Q from the orthonormal basis
   (c) obtain R using $R = Q^T A$.

2. Solve $R\hat{x} = Q^T\vec{b}$ to obtain $\hat{x}$.

Applying QR to Solve an Inconsistent System

Suppose we are given the data below.

x	-2	-1	0	1	2
y	-2	-1	1	2	2



Suppose we want to fit a straight line of the form $y = c_1 + c_2 x$ to our data and need to determine c_1 and c_2.

Using the model $y = c_1 + c_2 x$,

$$A\vec{x} = \begin{pmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} -2 \\ -1 \\ 1 \\ 2 \\ 2 \end{pmatrix} = \vec{b}$$

We have a system of the form $A\vec{x} = \vec{b}$. To apply the QR decomposition, we need matrices Q and R.

Example: Using QR to Solve Least Squares

Compute the least-squares solution to $A\vec{x} = \vec{b}$, where

$$A = \begin{pmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}, \quad \vec{b} = \begin{pmatrix} -2 \\ -1 \\ 1 \\ 2 \\ 2 \end{pmatrix}$$

$$R = Q^T A = \begin{pmatrix} 1/\sqrt{5} & 1/\sqrt{5} & 1/\sqrt{5} & 1/\sqrt{5} & 1/\sqrt{5} \\ -2/\sqrt{10} & -1/\sqrt{10} & 0 & 1/\sqrt{10} & 2/\sqrt{10} \end{pmatrix}\begin{pmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}$$

$$= \begin{pmatrix} 5/\sqrt{5} & 0 \\ 0 & 10/\sqrt{10} \end{pmatrix} = \begin{pmatrix} \sqrt{5} & 0 \\ 0 & \sqrt{10} \end{pmatrix}$$

Using Gram-Schmidt, the QR decomposition of A is

$$A = QR = \begin{pmatrix} 1/\sqrt{5} & -2/\sqrt{10} \\ 1/\sqrt{5} & -1/\sqrt{10} \\ 1/\sqrt{5} & 0 \\ 1/\sqrt{5} & 1/\sqrt{10} \\ 1/\sqrt{5} & 2/\sqrt{10} \end{pmatrix}\begin{pmatrix} \sqrt{5} & 0 \\ 0 & \sqrt{10} \end{pmatrix}$$

Next we need to compute $Q^T\vec{b}$.

$$Q^T\vec{b} = \begin{pmatrix} 1/\sqrt{5} & 1/\sqrt{5} & 1/\sqrt{5} & 1/\sqrt{5} & 1/\sqrt{5} \\ -2/\sqrt{10} & -1/\sqrt{10} & 0 & 1/\sqrt{10} & 2/\sqrt{10} \end{pmatrix}\begin{pmatrix}-2\\-1\\1\\2\\2\end{pmatrix} = \begin{pmatrix}2/\sqrt{5}\\ 11/\sqrt{10}\end{pmatrix}.$$

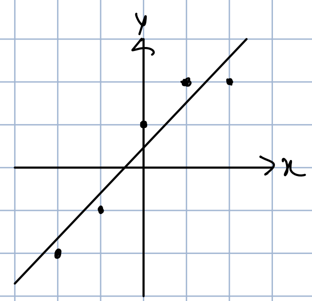
Finally, we solve $R\hat{x} = Q^T\vec{b}$.

$$\begin{pmatrix}\sqrt{5} & 0\\ 0 & \sqrt{10}\end{pmatrix}\hat{x} = \begin{pmatrix}2/\sqrt{5}\\ 11/\sqrt{10}\end{pmatrix}$$

$$\Rightarrow \hat{x} = \frac{1}{\sqrt{50}}\begin{pmatrix}\sqrt{10} & 0\\ 0 & \sqrt{5}\end{pmatrix}\begin{pmatrix}2/\sqrt{5}\\ 11/\sqrt{10}\end{pmatrix} = \frac{1}{\sqrt{50}}\begin{pmatrix}2\sqrt{2}\\ 11/\sqrt{2}\end{pmatrix} = \begin{pmatrix}2/\sqrt{25}\\ 11/\sqrt{100}\end{pmatrix}$$

$$\therefore \hat{x} = \begin{pmatrix}2/5\\ 11/10\end{pmatrix} \ast$$

Our linear model is $y = 2/5 + 11/10\, x$.



Exercise:

Consider again the problem where we have a problem where we have five points in $\mathbb{R}^2$ with coordinates (x, y) are $(-1, 4)$, $(-0.5, 2)$, $(0, 1)$, $(0.5, -3)$ and $(1, -4)$.

Now we want to solve a linear system using the QR factorization to determine the coefficients c_1 and c_2 for the line

$$y = c_1 x + c_2$$

that best fits the data. Using data and our model $y = c_1 x + c_2$, we can construct the linear system:

$$A\vec{x} = \vec{y}, \quad A = \begin{pmatrix} -1 & 1\\ -1/2 & 1\\ 0 & 1\\ 1/2 & 1\\ 1 & 1\end{pmatrix}, \quad \vec{y} = \begin{pmatrix}4\\2\\1\\-3\\-4\end{pmatrix}$$

$$(-1)^2 + (-1/2)^2 + 0^2 + (1/2)^2 + 1^2 = 1 + 1/4 + 0 + 1/4 + 1 = 5/2$$

$$\frac{1}{2}\sqrt{\frac{2}{5}} = \sqrt{\frac{2}{4\times 5}} = \sqrt{\frac{1}{10}}$$

Write down the QR factorization of A.

$$2\sqrt{2/5} + \sqrt{1/10}$$

$$Q = \begin{pmatrix} -\sqrt{2/5} & 1/\sqrt{5}\\ -1/2\sqrt{2/5} & 1/\sqrt{5}\\ 0 & 1/\sqrt{5}\\ 1/2\sqrt{2/5} & 1/\sqrt{5}\\ \sqrt{2/5} & 1/\sqrt{5}\end{pmatrix}, \quad Q^T = \begin{pmatrix} -\sqrt{2/5} & -\sqrt{1/10} & 0 & \sqrt{1/10} & \sqrt{2/5}\\ 1/\sqrt{5} & 1/\sqrt{5} & 1/\sqrt{5} & 1/\sqrt{5} & 1/\sqrt{5}\end{pmatrix}$$

$$R = Q^T A = \begin{pmatrix} -\sqrt{2/5} & -\sqrt{1/10} & 0 & \sqrt{1/10} & \sqrt{2/5}\\ 1/\sqrt{5} & 1/\sqrt{5} & 1/\sqrt{5} & 1/\sqrt{5} & 1/\sqrt{5}\end{pmatrix}\begin{pmatrix} -1 & 1\\ -1/2 & 1\\ 0 & 1\\ 1/2 & 1\\ 1 & 1\end{pmatrix} = \begin{pmatrix}\sqrt{5/2} & 0\\ 0 & 5/\sqrt{5}\end{pmatrix} = \begin{pmatrix}\sqrt{5/2} & 0\\ 0 & \sqrt{5}\end{pmatrix}$$

$$\therefore A = QR = \begin{pmatrix} -\sqrt{2/5} & 1/\sqrt{5} \\ -\sqrt{1/10} & 1/\sqrt{5} \\ 0 & 1/\sqrt{5} \\ \sqrt{1/10} & 1/\sqrt{5} \\ \sqrt{2/5} & 1/\sqrt{5} \end{pmatrix} \begin{pmatrix} \sqrt{5/2} & 0 \\ 0 & \sqrt{5} \end{pmatrix}$$

Hence or otherwise, find $\hat{x}$.

$$R\hat{x} = Q^T\vec{y}$$

$$\begin{pmatrix} \sqrt{5/2} & 0 \\ 0 & \sqrt{5} \end{pmatrix}\hat{x} = \begin{pmatrix} -\sqrt{2/5} & -\sqrt{1/10} & 0 & \sqrt{1/10} & \sqrt{2/5} \\ 1/\sqrt{5} & 1/\sqrt{5} & 1/\sqrt{5} & 1/\sqrt{5} & 1/\sqrt{5} \end{pmatrix} \begin{pmatrix} 4 \\ 2 \\ 1 \\ -3 \\ -4 \end{pmatrix}$$

$$\begin{pmatrix} \sqrt{5/2} & 0 \\ 0 & \sqrt{5} \end{pmatrix}\hat{x} = \begin{pmatrix} -4\sqrt{2/5} - 2\sqrt{1/10} + 0 - 3\sqrt{1/10} - 4\sqrt{2/5} \\ 1/\sqrt{5}(4+2+1-3-4) \end{pmatrix}$$

$$-8\sqrt{\tfrac{2}{5}} - 5\sqrt{\tfrac{1}{10}}$$
$$= -\sqrt{\frac{128}{5}} - \frac{5}{\sqrt{10}}$$
$$= -\sqrt{\frac{256}{10}} - \frac{5}{\sqrt{10}}$$
$$= -\frac{16}{\sqrt{10}} - \frac{5}{\sqrt{10}} = -\frac{21}{\sqrt{10}}$$

$$\begin{pmatrix} \sqrt{5/2} & 0 \\ 0 & \sqrt{5} \end{pmatrix}\hat{x} = \begin{pmatrix} -21/\sqrt{10} \\ 0 \end{pmatrix}$$

$$\hat{x} = \frac{1}{5/\sqrt{2}} \begin{pmatrix} \sqrt{5} & 0 \\ 0 & \sqrt{5/2} \end{pmatrix} \begin{pmatrix} -21/\sqrt{10} \\ 0 \end{pmatrix} = \frac{\sqrt{2}}{5} \begin{pmatrix} -21/\sqrt{2} \\ 0 \end{pmatrix}$$

$$\therefore \hat{x} = \begin{pmatrix} -21/5 \\ 0 \end{pmatrix} *$$

TOPIC 2: Applications to Linear Models

Residuals and Least-Squares

Motivation:

* linear models are commonly used in science and engineering model relationships between two or more quantities
  - ⇒ knowing more about this process is likely to be of use to many students.
* a better understanding of what the least-squares method is doing will aid in our understanding of how to interpret results from our algorithms, and determine which algorithms to use.
* introducing residual vector to characterize how our model fits the data

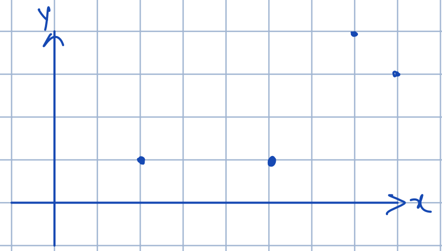
Example: Linear Model of the Form $y = c_0 + c_1 x$.

Suppose we are given the data below.

x_i	2	5	7	8
y_i	1	1	4	3

We wish to identify c_0 and c_1 so that $y = c_0 + c_1 x$ best fits the data.

The graph below shows our data.



Do you think the values of c_0 and c_1 will be positive or negative?

Compute the equation of the line $y = c_0 + c_1 x$ that best fits the data

x	2	5	7	8
y	1	1	4	3

To solve this problem, we might construct the system

$$\begin{pmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 4 \\ 3 \end{pmatrix}$$

This is a least-squares problem: $A\vec{x} = \vec{y}$. We can solve this using normal equations directly, or with a QR decomposition.

If we use the normal equations directly, we have

$$A^T = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{pmatrix}$$

$$A^TA = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{pmatrix} = \begin{pmatrix} 4 & 22 \\ 22 & 142 \end{pmatrix}, \quad A^T\vec{y} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 4 \\ 3 \end{pmatrix} = \begin{pmatrix} 9 \\ 59 \end{pmatrix}$$

The least-squares solution is given by solving

$$\begin{pmatrix} 4 & 22 \\ 22 & 142 \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \end{pmatrix} = \begin{pmatrix} 9 \\ 59 \end{pmatrix}$$

$$\begin{pmatrix} c_0 \\ c_1 \end{pmatrix} = \frac{1}{84} \begin{pmatrix} 142 & -22 \\ -22 & 4 \end{pmatrix} \begin{pmatrix} 9 \\ 59 \end{pmatrix} = \frac{1}{84} \begin{pmatrix} -20 \\ 38 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} c_0 \\ c_1 \end{pmatrix} = \begin{pmatrix} -5/21 \\ 19/42 \end{pmatrix}.$$

$y = c_0 + c_1 x = -\frac{5}{21} + \frac{19}{42}x$. $\therefore$ c_0 is negative, c_1 is positive.

Looking back at this process, we ask the following questions.

1. How well does our model approximate the given data?
2. How can we characterize the degree to which our model fit our data?

Model Fit

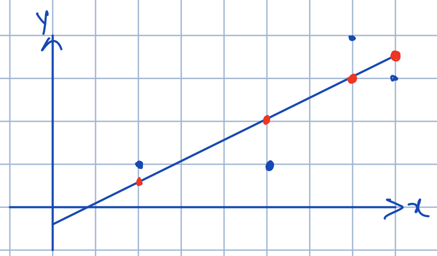
Using our model,

$$\hat{y}_i = c_0 + c_1 x_i = -\frac{5}{21} + \frac{19}{42} x_i$$

we obtain four estimates, $\hat{y}_i$.

x	2	5	7	8
y	1	1	4	3
$\hat{y}_i$	$2/3$	$85/42$	$41/14$	$71/21$

Our straight line, y_i, and $\hat{y}_i$ are shown below.



Can we use the distance between the data and our straight line to describe how well our model fit the data?

Residuals

By using the normal equations, we found the $\vec{x}$ that minimized

$$\|A\vec{x} - \vec{y}\|$$

over all possible $\vec{x} \in \mathbb{R}^n$. This is equivalent to minimizing

$$\|A\vec{x} - \vec{y}\|^2$$

over all possible $\vec{x} \in \mathbb{R}^n$. Now let $\vec{r} = A\vec{x} - \vec{y}$. Then,

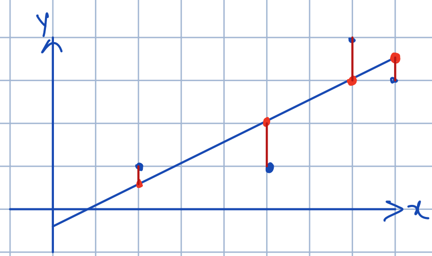
$$\|A\vec{x} - \vec{y}\|^2 = \|\vec{r}\|^2 = \sum_{i=1}^{n} r_i^2$$

where r_i are entries of $\vec{r}$.

Least-Squares Optimization

Our goal with least-squares is to minimize $\|r\|^2 = \sum_{i=1}^{n} r_i^2$.

Our previous example led to the linear model below.



The least-squares line minimizes the sum of squares of the residuals.

The distance between $\vec{y}$ and Col A is $\|\vec{r}\|$.
↳ As we are trying to minimize $\|\vec{r}\|^2$, note that $\|\vec{r}\|^2$ is the squared distance between $\vec{y}$ and Col A.
↳ In our example, this is:

$$\|\vec{r}\|^2 = \|\hat{y} - \vec{y}\|^2, \quad \vec{r} = \begin{pmatrix} 2/3 \\ 85/42 \\ 41/14 \\ 71/21 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ 4 \\ 3 \end{pmatrix} = \begin{pmatrix} -1/3 \\ 43/42 \\ -15/14 \\ 8/21 \end{pmatrix}$$

We can calculate that $\|r\|^2 = \frac{103}{42} \approx 2.45$. This is a way of describing how well our model fits our data.

Exercise:
Three points in $\mathbb{R}^2$ with coordinates (x, y) are $(1,3), (2,7), (3,5)$.

Suppose we were to construct a linear system to determine the coefficients c_1 and c_2 for the line $y = c_1 + c_2 x$ that best fits the data. After constructing the normal equations and solving them, we obtain the linear model below.

$$y = 3 + x.$$

Our line does not pass through all three data points.

Compute the residual vector, $\vec{r}$, and its squared length, $\|\vec{r}\|^2$. What is the squared length of the residual vector equal to?

The linear system representing this problem is $A\vec{x} = \vec{b}$, where

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}, \quad \vec{b} = \begin{pmatrix} 3 \\ 7 \\ 5 \end{pmatrix}, \quad \vec{x} = \begin{pmatrix} c_0 \\ c_1 \end{pmatrix}.$$

x	1	2	3
y	3	7	5
$\hat{y}_i$	4	5	6

Given $c_0 = 3$ and $c_1 = 1$,

$$\hat{b} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}$$

Note: can also use $\vec{b} - \hat{b}$

$$\Rightarrow \vec{r} = \hat{b} - \vec{b} = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} - \begin{pmatrix} 3 \\ 7 \\ 5 \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}. \quad \|\vec{r}\|^2 = 1^2 + (-2)^2 + 1^2 = 6$$

<u>Mean-Deviation Form</u>

* linear models are commonly used in science and engineering model relationships between two or more quantities ⇒ identifying ways to make that process more efficient is valuable.
* a common practice when using a model of the form $y = c_0 + c_1 x$ is to compute the average, $\bar{x}$, of the x-values, and introduce a new variable $x_* = x - \bar{x}$.
* the new data are in <u>mean-deviation</u> form.

Suppose we are given the data below.

x_i	2	5	7	8
y_i	1	1	4	3

If we were to fit a linear model to this data of the form $y = c_0 + c_1 x$, we could construct the system

$$\begin{pmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 4 \\ 3 \end{pmatrix}$$

This is a least-squares problem of the form $A\vec{x} = \vec{y}$.
If we used the normal equations directly, we have

$$A^T A = \begin{pmatrix} 4 & 22 \\ 22 & 142 \end{pmatrix}, \quad A^T \vec{y} = \begin{pmatrix} 9 \\ 59 \end{pmatrix}$$

The least-squares solution is given by solving

$$\begin{pmatrix} 4 & 22 \\ 22 & 142 \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \end{pmatrix} = \begin{pmatrix} 9 \\ 59 \end{pmatrix}$$

After solving this linear system, we obtain $c_0 = -\frac{5}{21}$ and $c_1 = \frac{19}{42}$.

$$y = c_0 + c_1 x = -\frac{5}{21} + \frac{19}{42} x.$$

If we were to use mean-deviation form, how would our approach be different?

The average value of x_i is $\bar{x} = 5.5$. Subtracting $\bar{x}$ from the x-values, our new linear model becomes

$$y = c_0 + c_1 x_*$$

and our linear system becomes

$$\begin{pmatrix} 1 & -3.5 \\ 1 & -0.5 \\ 1 & 1.5 \\ 1 & 2.5 \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \end{pmatrix} = A\vec{x} = \begin{pmatrix} 1 \\ 1 \\ 4 \\ 3 \end{pmatrix} = \vec{b}$$

What property do the columns of A now have?

The Normal Equations with Mean-Deviation Form

If we use the normal equations again, we now have $A^T A = \begin{pmatrix} 4 & 0 \\ 0 & 21 \end{pmatrix}$, $A^T \vec{y} = \begin{pmatrix} 9 \\ 9.5 \end{pmatrix}$.

$$A^T = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -3.5 & -0.5 & 1.5 & 2.5 \end{pmatrix}$$

$$A^T A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -3.5 & -0.5 & 1.5 & 2.5 \end{pmatrix} \begin{pmatrix} 1 & -3.5 \\ 1 & -0.5 \\ 1 & 1.5 \\ 1 & 2.5 \end{pmatrix} = \begin{pmatrix} 4 & 0 \\ 0 & 21 \end{pmatrix}.$$

$$A^T\vec{y} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -3.5 & -0.5 & 1.5 & 2.5 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 4 \\ 3 \end{pmatrix} = \begin{pmatrix} 9 \\ 9.5 \end{pmatrix}$$

By using mean-deviation form, A has orthogonal columns, so A^TA is diagonal. Upon solving the normal equations, we obtain (by inspection) $c_0 = 9/4$ and $c_1 = 19/42$.

$$y = c_0 + c_1 x_* = \frac{9}{4} + \frac{19}{42}(x - \bar{x}) = -\frac{5}{21} + \frac{19}{42}x.$$

Exercise:

Five points in $\mathbb{R}^2$ with coordinates (t, y) are given in the table below.

t_*	t	y
-2 $\leftarrow -3$	1	2
-1	2	0
0	3	1
1	4	-1
2	5	3

Let $\bar{t}$ be the average value of t. What is $\bar{t}$ equal to?

$$\bar{t} = \frac{1+2+3+4+5}{5}$$

Using mean-deviation form, we introduce the change of variable, $t_* = t - \bar{t}$ and the model:

$$c_1 + c_2 t_* = y.$$

We can set up an inconsistent system:

$$A\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \vec{b},$$

$$A = \begin{pmatrix} 1 & a_{12} \\ 1 & a_{22} \\ 1 & a_{32} \\ 1 & a_{42} \\ 1 & a_{52} \end{pmatrix}, \quad \vec{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{pmatrix}$$

If the first row of A corresponds to the first data point (1, 3), then what would we use for a_{12}? $1 - \bar{t} = 1 - 3 = \underline{\underline{-2}}$

What value would we use for b_1? $\underline{\underline{2}}$

$$A = \begin{pmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}, \quad \vec{b} = \begin{pmatrix} 2 \\ 0 \\ 1 \\ -1 \\ 3 \end{pmatrix}$$

Constructing our normal equations, we obtain the system: $A^TA\hat{x} = A^T\vec{b}$, where $\hat{x} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$.
Find A^TA and $A^T\vec{b}$.

$$A^TA = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \end{pmatrix}\begin{pmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 5 & 0 \\ 0 & 10 \end{pmatrix}. \quad A^T\vec{b} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \end{pmatrix}\begin{pmatrix} 2 \\ 0 \\ 1 \\ -1 \\ 3 \end{pmatrix} = \begin{pmatrix} 5 \\ 1 \end{pmatrix}$$

Use the normal equations to determine the coefficients c_1 and c_2 for the model

$$c_1 + c_2 t_* = y$$

What are the values of c_1 and c_2?

$$A^TA\hat{x} = A^T\vec{b}$$
$$\begin{pmatrix} 5 & 0 \\ 0 & 10 \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 5 \\ 1 \end{pmatrix}.$$
$$\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \frac{1}{50}\begin{pmatrix} 10 & 0 \\ 0 & 5 \end{pmatrix}\begin{pmatrix} 5 \\ 1 \end{pmatrix} = \frac{1}{50}\begin{pmatrix} 50 \\ 5 \end{pmatrix} = \begin{pmatrix} 1 \\ 1/10 \end{pmatrix}. \quad \therefore c_1 = 1,\ c_2 = \frac{1}{10}$$

The General Linear Model

Overview of General Linear Model:

* It can be helpful to fit data points with curves that are not straight lines.
* We can still use a least-squares approach in these situations, by constructing a system of the form

$$A\vec{x} = \vec{y}$$

which is (often) inconsistent, and then computing the least-squares solution with the normal equations,

$$A^TA\hat{x} = A^T\vec{y}.$$

* our approach will be to model the data using

$$\vec{y} = A\vec{x} + \vec{r}$$

and to minimize the length of $\vec{r}$.

Least Squares Fitting for Other Curves

We can consider least squares fitting for the form

$$y = c_0 + c_1 f_1(x) + c_2 f_2(x) + \cdots + c_k f_k(x)$$

If functions f_i are known, this is a linear problem in the c_i variables.

Example: Second Order Polynomial

Consider the data in the table below.

x	-1	0	0	1
y	2	1	0	6

Determine the coefficients of c_1 and c_2 for the curve $y = c_1 x + c_2 x^2$ that best fits the data.

Using the model $y = c_1 x + c_2 x^2$, we obtain the equations

$$\begin{aligned} -c_1 + c_2 &= 2 \\ 0c_1 + c_2 &= 1 \\ 0c_1 + 0c_2 &= 0 \\ c_1 + c_2 &= 6 \end{aligned}$$

We can use a matrix equation to represent our linear system.

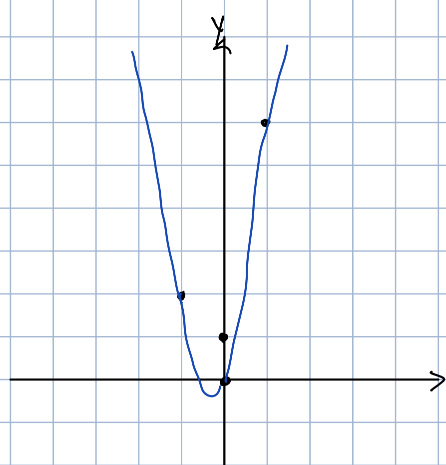
$$A\vec{x} = \begin{pmatrix} -1 & 1 \\ 0 & 0 \\ 0 & 0 \\ 1 & 1 \end{pmatrix} \vec{x} = \begin{pmatrix} 2 \\ 1 \\ 0 \\ 6 \end{pmatrix} = \vec{y}$$

$$A^T A = \begin{pmatrix} -1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 \\ 0 & 0 \\ 0 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}. \quad A^T \vec{y} = \begin{pmatrix} -1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ 0 \\ 6 \end{pmatrix} = \begin{pmatrix} 4 \\ 8 \end{pmatrix}$$

Constructing the normal equations, we obtain

$$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \hat{x} = \begin{pmatrix} 4 \\ 8 \end{pmatrix}$$

This leads to the solution $y = 2x + 4x^2$.



Multiple Regression

We can also consider least squares fitting for the form

$$z = c_0 + c_1 f_1(x, y) + c_2 f_2(x, y) + \dots + c_k f_k(x, y).$$

If functions f_i are known, this is another linear problem in the c_i variables.

Example: A Model of the Form $z = c_0 + c_1 x + c_2 y$

Consider the data in the table below.

x	-1	0	0	1
y	-1	-1	1	1
z	2	2	4	6

Determine the coefficients c_1 and c_2 for the surface $z = c_0 + c_1 x + c_2 y$ that best-fits the data.

Using the model $z = c_0 + c_1 x + c_2 y$, we obtain a matrix equation to represent our linear system.

$$A\vec{x} = \begin{pmatrix} 1 & -1 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} \vec{x} = \begin{pmatrix} 2 \\ 2 \\ 4 \\ 6 \end{pmatrix} = \vec{b}$$

$$A^TA = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 0 & 1 \\ -1 & -1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & -1 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix}.$$

$$A^T\vec{b} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 0 & 1 \\ -1 & -1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 2 \\ 2 \\ 4 \\ 6 \end{pmatrix} = \begin{pmatrix} 14 \\ 4 \\ 6 \end{pmatrix}$$

Constructing the normal equations, $A^TA\hat{x} = A^T\vec{b}$, we obtain $\begin{pmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix}\hat{x} = \begin{pmatrix} 14 \\ 4 \\ 6 \end{pmatrix}$.

This leads to the solution $\hat{x} = \begin{pmatrix} 7/2 \\ 2 \\ 3/2 \end{pmatrix}$.

$\therefore$ our model for this data is $z = 7/2 + 2x + 3/2\, y$.

Exercise:

Four points in $\mathbb{R}^3$ with coordinates (x, y, z) are given in the table below.

x	y	z
-1	2	5
0	-1	1
1	0	2
1	2	0

Determine the coefficients c_1 and c_2 for the plane $z = c_1 x + c_2 y$ that best fits the data.

$$A\vec{x} = \begin{pmatrix} -1 & 2 \\ 0 & -1 \\ 1 & 0 \\ 1 & 2 \end{pmatrix}\vec{x} = \begin{pmatrix} 5 \\ 1 \\ 2 \\ 0 \end{pmatrix} = \vec{b}$$

$$A^TA = \begin{pmatrix} -1 & 0 & 1 & 1 \\ 2 & -1 & 0 & 2 \end{pmatrix}\begin{pmatrix} -1 & 2 \\ 0 & -1 \\ 1 & 0 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 0 & 9 \end{pmatrix}, \quad A^T\vec{b} = \begin{pmatrix} -1 & 0 & 1 & 1 \\ 2 & -1 & 0 & 2 \end{pmatrix}\begin{pmatrix} 5 \\ 1 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} -3 \\ 9 \end{pmatrix}$$

$$A^TA\hat{x} = A^T\vec{b}.$$

$$\begin{pmatrix} 3 & 0 \\ 0 & 9 \end{pmatrix}\hat{x} = \begin{pmatrix} -3 \\ 9 \end{pmatrix} \Rightarrow \hat{x} = \frac{1}{27}\begin{pmatrix} 9 & 0 \\ 0 & 3 \end{pmatrix}\begin{pmatrix} -3 \\ 9 \end{pmatrix} = \frac{1}{27}\begin{pmatrix} -27 \\ 27 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

$\therefore c_1 = -1, \; c_2 = 1$ *