Is there a line which fits the data in some optimal sense?
Slope = ?
y-intercept = ?
When are the points and the line “close” in some sense? We could try to minimize the “projected” distance of each point to the line.
We could minimize the vertical distance to the line

**Question**: What is the slope and y-intercept of the line that minimizes this vertical distance?
What does our model look like?

Our independent variable is $x$ and the dependent variable is $y$, and the model we are trying to fit is

$$y = mx + b$$

Our goal is to find $m$ and $b$
Over-determined system

One approach is to attempt to “solve” the system of equations given by

\[ mx_1 + b = y_1 \]
\[ mx_2 + b = y_2 \]
\[ mx_3 + b = y_3, \]
\[ \vdots \]
\[ mx_N + b = y_N \]

An exact solution does not exist in general.
Over-determined system

The linear system is given by

\[
\begin{bmatrix}
x_1 & 1 \\
x_2 & 1 \\
x_3 & 1 \\
\vdots & \\
x_N & 1 \\
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3 \\
\vdots \\
x_N \\
\end{bmatrix}
= 
\begin{bmatrix}
ym \\
b \\
\end{bmatrix}
= 
\begin{bmatrix}
y_1 \\
y_2 \\
y_3 \\
\vdots \\
y_N \\
\end{bmatrix}
\]

Not invertible!

Solve the normal equations

\[
A^T A \hat{x} = A^T b
\]
Normal equation solution

Normal equations

\[ A^T A \hat{x} = A^T b \]

\[ \hat{x} = (A^T A)^{-1} A^T b \]

to get \( m = \hat{x}_1 \) and \( b = \hat{x}_2 \)

This system is invertible as long as \( A \) has “full column rank”
The vertical distance from a point \((x_i, y_i)\) to the line is given by ...

Vertical distance to the line: 
\[ d_i = |y_i - (mx_i + b)| \]
Minimize sum of squares

The normal equation solution minimises the sum of the squared distances:

\[ D(m, b) = \frac{1}{2} \sum_{i=1}^{N} (y_i - (mx_i + b))^2 \]

From Calculus III, we know that at a minimum of a multivariable function, we have:

\[ \frac{\partial D}{\partial m} = 0 \]
\[ \frac{\partial D}{\partial b} = 0 \]
Minimize sum of squares

\[
\frac{\partial D}{\partial m} = \sum_{i=1}^{N} (y_i - (mx_i + b))(-x_i) = 0
\]

\[
\frac{\partial D}{\partial b} = \sum_{i=1}^{N} (y_i - (mx_i + b)) = 0
\]

These equations are exactly satisfied at the \( \hat{x} \), the solution to the normal equations.

\[
\begin{bmatrix}
\sum x_i^2 & \sum x_i \\
\sum x_i & N
\end{bmatrix}
\begin{bmatrix}
m \\
b
\end{bmatrix} =
\begin{bmatrix}
\sum x_i y_i \\
\sum y_i
\end{bmatrix}
\]

\[
A^TA \quad \hat{x} \quad A^Tb
\]
Another way to view what least squares is doing:

\[ \hat{x} \text{ minimizes } \frac{1}{2} \| A\hat{x} - b \|^2 \]

We are minimizing the squared length of the vector \( A\hat{x} - b \).

**A linear algebra view:**

In linear algebra terms, \( b \) is not in general in the column space \( C(A) \) of \( A \). But we can ”project” \( b \) onto \( C(A) \) and get the vector \( A\hat{x} \) in \( C(A) \) that most closely resembles \( b \).