Least-Squares Problems and Best-Fit Lines

Linear Algebra

This work was initially funded by an Institutional OER Grant from the Colorado Department of Higher Education (CDHE). For similar OER materials in other courses funded by this project in the Department of Mathematical and Statistical Sciences at the University of Colorado Denver, visit
https://github.com/CU-Denver-MathStats-OER

Inconsistent Linear Systems

Consider solving $A\mathbf{x} = \mathbf{b}$.

- If the system is consistent, then $\mathbf{b}$ is in Col A.
- If the system is inconsistent, then $\mathbf{b}$ is not in Col A.

Inconsistent Linear Systems

Consider solving $A\mathbf{x} = \mathbf{b}$.

- If the system is consistent, then $\mathbf{b}$ is in Col A.
- If the system is inconsistent, then $\mathbf{b}$ is not in Col A.

As a consolation prize, let's find $\hat{\mathbf{x}}$ such that $A\hat{\mathbf{x}}$ is the closest vector in Col A to $\mathbf{b}$; that is,

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\| \text{ for all } \mathbf{x} \text{ in } \mathbb{R}^n.$$

Such a sol'n $\hat{\mathbf{x}}$ is called a least-squares solution since $\hat{\mathbf{x}}$ minimizes $\|\mathbf{b} - A\mathbf{x}\| = \|\mathbf{z}\| = \sqrt{\sum \mathbf{z}_i^2}$.

Inconsistent Linear Systems

Consider solving $A\mathbf{x} = \mathbf{b}$.

- If the system is consistent, then $\mathbf{b}$ is in Col A.
- If the system is inconsistent, then $\mathbf{b}$ is not in Col A.

As a consolation prize, let's find $\hat{\mathbf{x}}$ such that $A\hat{\mathbf{x}}$ is the closest vector in Col A to $\mathbf{b}$; that is,

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\| \text{ for all } \mathbf{x} \text{ in } \mathbb{R}^n.$$

Such a sol'n $\hat{\mathbf{x}}$ is called a least-squares solution since $\hat{\mathbf{x}}$ minimizes $\|\mathbf{b} - A\mathbf{x}\| = \|\mathbf{z}\| = \sqrt{\sum \mathbf{z}_i^2}$.

Let $\hat{\mathbf{b}} = \text{proj}_{\text{Col } A}\mathbf{b}$. Then $\hat{\mathbf{b}}$ is the closest vector in Col A to $\mathbf{b}$, and $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$ is consistent.

Inconsistent Linear Systems

Consider solving $A\mathbf{x} = \mathbf{b}$.

- If the system is consistent, then $\mathbf{b}$ is in Col A.
- If the system is inconsistent, then $\mathbf{b}$ is not in Col A.

As a consolation prize, let's find $\hat{\mathbf{x}}$ such that $A\hat{\mathbf{x}}$ is the closest vector in Col A to $\mathbf{b}$; that is,

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\| \text{ for all } \mathbf{x} \text{ in } \mathbb{R}^n.$$

Such a sol'n $\hat{\mathbf{x}}$ is called a least-squares solution since $\hat{\mathbf{x}}$ minimizes $\|\mathbf{b} - A\mathbf{x}\| = \|\mathbf{z}\| = \sqrt{\sum \mathbf{z}_i^2}$.

Let $\hat{\mathbf{b}} = \text{proj}_{\text{Col } A}\mathbf{b}$. Then $\hat{\mathbf{b}}$ is the closest vector in Col A to $\mathbf{b}$, and $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$ is consistent.

We know how to do each of these steps:

- Compute $\hat{\mathbf{b}} = \text{proj}_{\text{Col } A}\mathbf{b}$. (How?)
- Solve $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$.

Inconsistent Linear Systems

Consider solving $A\mathbf{x} = \mathbf{b}$.

- If the system is consistent, then $\mathbf{b}$ is in Col A.
- If the system is inconsistent, then $\mathbf{b}$ is not in Col A.

As a consolation prize, let's find $\hat{\mathbf{x}}$ such that $A\hat{\mathbf{x}}$ is the closest vector in Col A to $\mathbf{b}$; that is,

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\| \text{ for all } \mathbf{x} \text{ in } \mathbb{R}^n.$$

Such a sol'n $\hat{\mathbf{x}}$ is called a least-squares solution since $\hat{\mathbf{x}}$ minimizes $\|\mathbf{b} - A\mathbf{x}\| = \|\mathbf{z}\| = \sqrt{\sum \mathbf{z}_i^2}$.

Let $\hat{\mathbf{b}} = \text{proj}_{\text{Col } A}\mathbf{b}$. Then $\hat{\mathbf{b}}$ is the closest vector in Col A to $\mathbf{b}$, and $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$ is consistent.

We know how to do each of these steps:

- Compute $\hat{\mathbf{b}} = \text{proj}_{\text{Col } A}\mathbf{b}$. (How?)
- Solve $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$.

But there is a simpler way!

Least-Squares Solutions

We wish to find least-squares solutions $\hat{\mathbf{x}}$ to $A\mathbf{x} = \mathbf{b}$.

Such least-squares solutions satisfy $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$, where $\hat{\mathbf{b}} = \text{proj}_{\text{Col } A}\mathbf{b}$.

Least-Squares Solutions

We wish to find least-squares solutions $\hat{\mathbf{x}}$ to $A\mathbf{x} = \mathbf{b}$.

Such least-squares solutions satisfy $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$, where $\hat{\mathbf{b}} = \text{proj}_{\text{Col } A}\mathbf{b}$.

By the Orthogonal Decomposition Theorem, $\mathbf{z} = \mathbf{b} - \hat{\mathbf{b}}$ is orthogonal to Col A.
Thus, $\mathbf{z} = \mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to each column $\mathbf{a}_j$ of the matrix A.

$$\mathbf{a}_j \cdot (\mathbf{b} - A\hat{\mathbf{x}}) = 0 \text{ for each } j.$$
$$\mathbf{a}_j^T (\mathbf{b} - A\hat{\mathbf{x}}) = 0 \text{ for each } j.$$

Least-Squares Solutions

We wish to find least-squares solutions $\hat{\mathbf{x}}$ to $A\mathbf{x} = \mathbf{b}$.

Such least-squares solutions satisfy $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$, where $\hat{\mathbf{b}} = \text{proj}_{\text{Col } A}\mathbf{b}$.

By the Orthogonal Decomposition Theorem, $\mathbf{z} = \mathbf{b} - \hat{\mathbf{b}}$ is orthogonal to Col A.
Thus, $\mathbf{z} = \mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to each column $\mathbf{a}_j$ of the matrix A.

$$\begin{aligned} \mathbf{a}_j \cdot (\mathbf{b} - A\hat{\mathbf{x}}) &= 0 \text{ for each } j. \\ \mathbf{a}_j^T(\mathbf{b} - A\hat{\mathbf{x}}) &= 0 \text{ for each } j. \\ A^T(\mathbf{b} - A\hat{\mathbf{x}}) &= \mathbf{0}. \end{aligned}$$

Rearranging, we have that $(A^TA)\hat{\mathbf{x}} = A^T\mathbf{b}$. These are called the normal equations.

Least-Squares Solutions

We wish to find least-squares solutions $\hat{\mathbf{x}}$ to $A\mathbf{x} = \mathbf{b}$.

Such least-squares solutions satisfy $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$, where $\hat{\mathbf{b}} = \text{proj}_{\text{Col } A}\mathbf{b}$.

By the Orthogonal Decomposition Theorem, $\mathbf{z} = \mathbf{b} - \hat{\mathbf{b}}$ is orthogonal to Col A.
Thus, $\mathbf{z} = \mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to each column $\mathbf{a}_j$ of the matrix A.

$$\begin{aligned} \mathbf{a}_j \cdot (\mathbf{b} - A\hat{\mathbf{x}}) &= 0 \text{ for each } j. \\ \mathbf{a}_j^T(\mathbf{b} - A\hat{\mathbf{x}}) &= 0 \text{ for each } j. \\ A^T(\mathbf{b} - A\hat{\mathbf{x}}) &= \mathbf{0}. \end{aligned}$$

Rearranging, we have that $(A^TA)\hat{\mathbf{x}} = A^T\mathbf{b}$. These are called the normal equations.

Theorem

The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ is given by the solutions of the normal equations.

Example

Find the least-squares solutions of $A\mathbf{x} = \mathbf{b}$, where $A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$.

Example

Find the least-squares solutions of $A\mathbf{x} = \mathbf{b}$, where $A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$.

Notice that $\left[\begin{array}{cc|c} 4 & 0 & 2 \\ 0 & 2 & 0 \\ 1 & 1 & 11 \end{array}\right] \xrightarrow{\text{RREF}} \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right]$. Hence the system is inconsistent.

Example

Find the least-squares solutions of $A\mathbf{x} = \mathbf{b}$, where $A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$.

Notice that $\left[\begin{array}{cc|c} 4 & 0 & 2 \\ 0 & 2 & 0 \\ 1 & 1 & 11 \end{array}\right] \xrightarrow{\text{RREF}} \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right]$. Hence the system is inconsistent.

We solve the normal equations:

$$A^T A\mathbf{x} = A^T\mathbf{b}.$$

$$\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}\mathbf{x} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

one sol: $\hat{\mathbf{x}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

Example

Find the least-squares solutions of $A\mathbf{x} = \mathbf{b}$, where $A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$.

Notice that $\left[\begin{array}{cc|c} 4 & 0 & 2 \\ 0 & 2 & 0 \\ 1 & 1 & 11 \end{array}\right] \xrightarrow{\text{RREF}} \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right]$. Hence the system is inconsistent.

We solve the normal equations:

$$A^T A\mathbf{x} = A^T\mathbf{b}.$$

$$\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 19 \\ 11 \end{bmatrix} \qquad \text{one sol: } \hat{\mathbf{x}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Compare:

$$\text{G–S: } \left\{ \begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -4/17 \\ 2 \\ 16/17 \end{bmatrix} \right\}, \quad \hat{\mathbf{b}} = \text{proj}_{\text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}} \mathbf{b} = \left(\frac{\mathbf{b} \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \right) \mathbf{v}_1 + \left(\frac{\mathbf{b} \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2} \right) \mathbf{v}_2 = \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix} = A\hat{\mathbf{x}}.$$

Unique solution to the Least-Squares Problem

Theorem

The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ is given by the solutions of the normal equations

$$A^T A\mathbf{x} = A^T\mathbf{b}.$$

Unique solution to the Least-Squares Problem

Theorem

The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ is given by the solutions of the normal equations

$$A^T A\mathbf{x} = A^T\mathbf{b}.$$

Theorem

Let A be an $m \times n$ matrix. The following statements are logically equivalent:

(a) The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution for each $\mathbf{b}$ in $\mathbb{R}^m$.

(b) The columns of A are linearly independent.

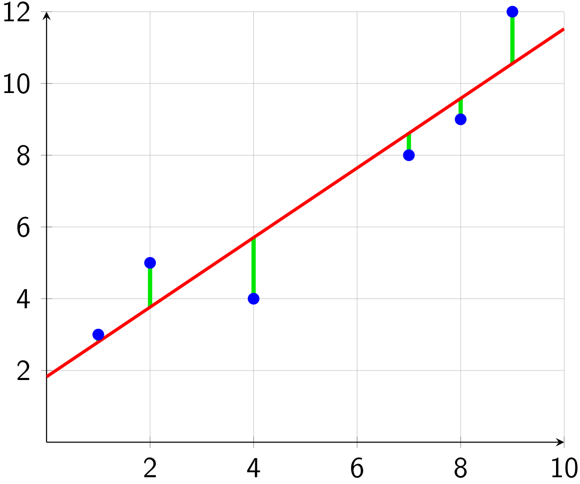
(c) The matrix $A^T A$ is invertible.

When these statements are satisfied, the least-squares solution $\hat{\mathbf{x}}$ is given by

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

Best-Fit Line

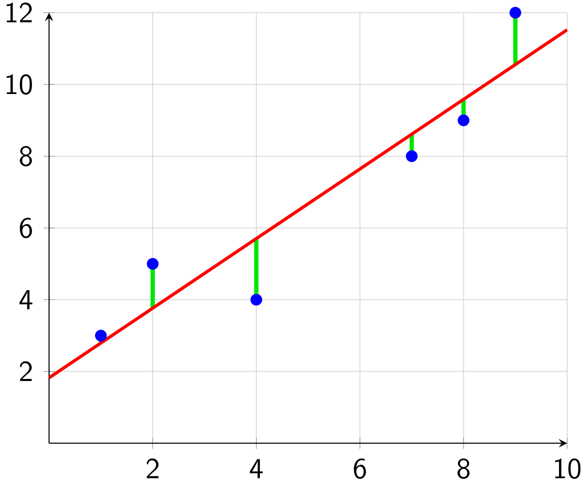
Suppose we have n data points $(x_1, y_1), \ldots, (x_n, y_n)$. What is the line that best fits this data?



We want a line of the form $y = \beta_0 + \beta_1 x$.

Best-Fit Line

Suppose we have n data points $(x_1, y_1), \ldots, (x_n, y_n)$. What is the line that best fits this data?



We want a line of the form $y = \beta_0 + \beta_1 x$.

Which line best fits the data?

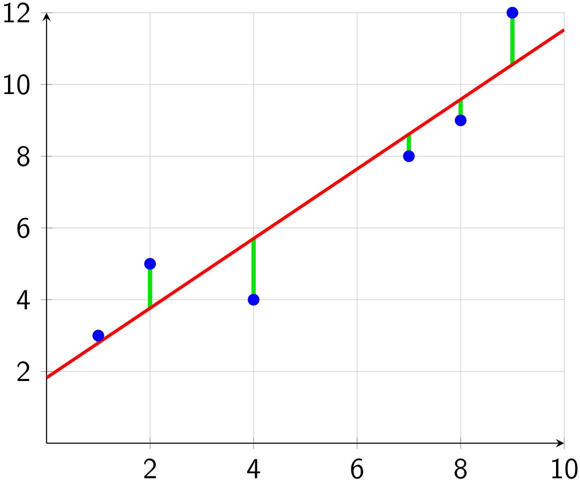
We want to minimize the sum of squares

$$\sqrt{\sum_{i=1}^{n} \Big(y_i - (\beta_0 + \beta_1 x_i) \Big)^2}.$$

$y_i - (\beta_0 + \beta_1 x_i)$ is called the residual.

Best-Fit Line

Suppose we have n data points $(x_1, y_1), \ldots, (x_n, y_n)$. What is the line that best fits this data?



We want a line of the form $y = \beta_0 + \beta_1 x$.

Which line best fits the data?

We want to minimize the sum of squares

$$\sqrt{\sum_{i=1}^{n} \Big(y_i - (\beta_0 + \beta_1 x_i) \Big)^2}.$$

$y_i - (\beta_0 + \beta_1 x_i)$ is called the residual.

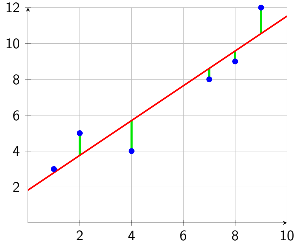
Can we formulate as a least-squares problem?

Yes!

Best-Fit Line

We would like to solve:

$$
\begin{aligned}
y_1 &= \beta_0 + \beta_1 x_1 \\
y_2 &= \beta_0 + \beta_1 x_2 \\
y_3 &= \beta_0 + \beta_1 x_3 \\
&\vdots \\
y_n &= \beta_0 + \beta_1 x_n
\end{aligned}
$$



Best-Fit Line

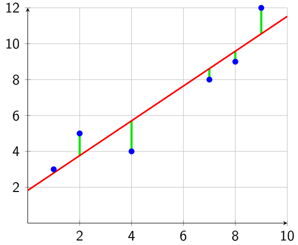
We would like to solve:

$$
\begin{aligned}
y_1 &= \beta_0 + \beta_1 x_1 \\
y_2 &= \beta_0 + \beta_1 x_2 \\
y_3 &= \beta_0 + \beta_1 x_3 \\
&\vdots \\
y_n &= \beta_0 + \beta_1 x_n
\end{aligned}
$$

As a matrix equation:

$$
\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}
$$

$$
\mathbf{y} = A\boldsymbol{\beta}
$$



Best-Fit Line

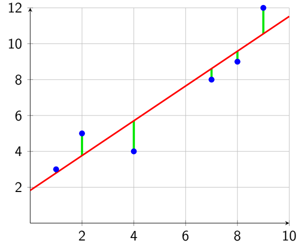
We would like to solve:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_1 \\ y_2 &= \beta_0 + \beta_1 x_2 \\ y_3 &= \beta_0 + \beta_1 x_3 \\ &\vdots \\ y_n &= \beta_0 + \beta_1 x_n \end{aligned}$$

As a matrix equation:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

$$\mathbf{y} = A\boldsymbol{\beta}$$

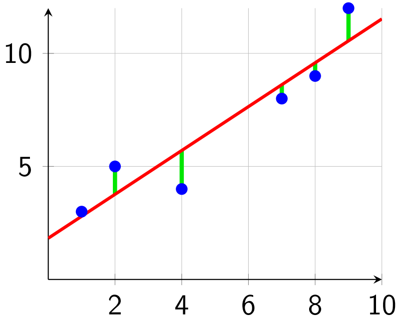


- A is called the model matrix (or design matrix).
- $\mathbf{y}$ is the vector of observed responses.
- $\boldsymbol{\beta}$ is the vector of regression coefficients.

Computing the least-squares solution of $\mathbf{y} = A\boldsymbol{\beta}$ is equivalent to finding values for the regression coefficients β_0 and β_1 that minimize the sum of square residuals.

Example

Suppose we have data $(1, 3), (2, 5), (4, 4), (7, 8), (8, 9), (9, 12)$. What is the best-fit line?

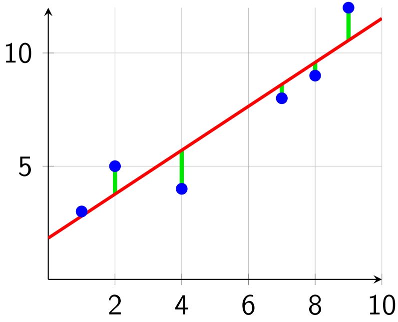


We want a least-squares solution of $\mathbf{y} = A\boldsymbol{\beta}$.

$$\begin{bmatrix} 3 \\ 5 \\ 4 \\ 8 \\ 9 \\ 12 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 4 \\ 1 & 7 \\ 1 & 8 \\ 1 & 9 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}.$$

Example

Suppose we have data $(1,3),(2,5),(4,4),(7,8),(8,9),(9,12)$. What is the best-fit line?



We want a least-squares solution of $\mathbf{y} = A\boldsymbol{\beta}$.

$$\begin{bmatrix} 3 \\ 5 \\ 4 \\ 8 \\ 9 \\ 12 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 4 \\ 1 & 7 \\ 1 & 8 \\ 1 & 9 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}.$$

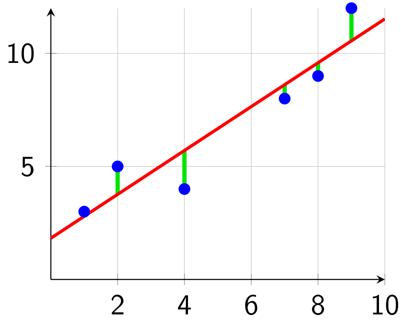
We solve the normal equations by computing

$$\boldsymbol{\beta} = (A^T A)^{-1} A^T \mathbf{y} = \begin{bmatrix} 1.824 \\ 0.970 \end{bmatrix}.$$

So $y = 1.824 + 0.970x$ is the best-fit line.

Example

Suppose we have data $(1,3),(2,5),(4,4),(7,8),(8,9),(9,12)$. What is the best-fit line?



We want a least-squares solution of $\mathbf{y} = A\boldsymbol{\beta}$.

$$\begin{bmatrix} 3 \\ 5 \\ 4 \\ 8 \\ 9 \\ 12 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 4 \\ 1 & 7 \\ 1 & 8 \\ 1 & 9 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}.$$

We solve the normal equations by computing

$$\boldsymbol{\beta} = (A^T A)^{-1} A^T \mathbf{y} = \begin{bmatrix} 1.824 \\ 0.970 \end{bmatrix}.$$

So $y = 1.824 + 0.970x$ is the best-fit line.

We can quantify the error by $\mathbf{y} = A\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \hat{\mathbf{y}} + \mathbf{z}$, $\|\mathbf{z}\| \approx 2.699$.

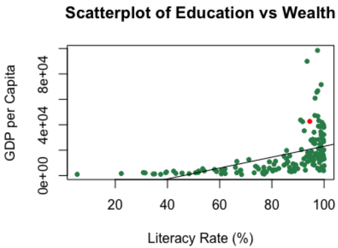
Predicting Wealth Based on Literacy Rate

Data scientists often begin with a simple model, and then determine whether predictions increase when new predictors are added. Let's first consider the following potential relationship:

- Let y denote the Gross Domestic Product (GDP) per capita of a country (in thousands of dollars).
- Let x_1 denote the literacy rate of the country's population (as a percentage).
- We collect a dataset that consists of n observations.
- Based on our data, what is the best model of the form $y = \beta_0 + \beta_1 x_1$?

$$\begin{aligned} 2.079 &= \beta_0 + \beta_1(31.4) \\ 13.440 &= \beta_0 + \beta_1(98.1) \\ 11.324 &= \beta_0 + \beta_1(81.4) \\ &\vdots \\ 3.537 &= \beta_0 + \beta_1(88.7) \end{aligned}$$

$$\begin{bmatrix} 2.07 \\ 13.44 \\ 11.324 \\ \vdots \\ 3.537 \end{bmatrix} = \begin{bmatrix} 1 & 31.4 \\ 1 & 98.1 \\ 1 & 81.4 \\ \vdots & \vdots \\ 1 & 88.7 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$



Interpreting the Results

Let's imagine that we randomly select four countries record the most recent data for each country's GDP per capita and literacy rate. $\widehat{\texttt{Wealth}} = -1.938 + 0.126(\texttt{Education})$

$$A = \begin{bmatrix} 1 & 31 \\ 1 & 98 \\ 1 & 81 \\ 1 & 89 \end{bmatrix} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} 2 \\ 13 \\ 11 \\ 4 \end{bmatrix}$$

Interpreting the Results

Let's imagine that we randomly select four countries record the most recent data for each country's GDP per capita and literacy rate. $\widehat{\texttt{Wealth}} = -1.938 + 0.126(\texttt{Education})$

$$A = \begin{bmatrix} 1 & 31 \\ 1 & 98 \\ 1 & 81 \\ 1 & 89 \end{bmatrix} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} 2 \\ 13 \\ 11 \\ 4 \end{bmatrix} \quad A^T A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 31 & 98 & 81 & 89 \end{bmatrix} \begin{bmatrix} 1 & 31 \\ 1 & 98 \\ 1 & 81 \\ 1 & 89 \end{bmatrix} = \begin{bmatrix} 4 & 299 \\ 299 & 25{,}047 \end{bmatrix}$$

For the least-squares solution to $A\boldsymbol{\beta} = \mathbf{y}$, we solve the normal equations $A^T A\boldsymbol{\beta} = A^T\mathbf{y}$:

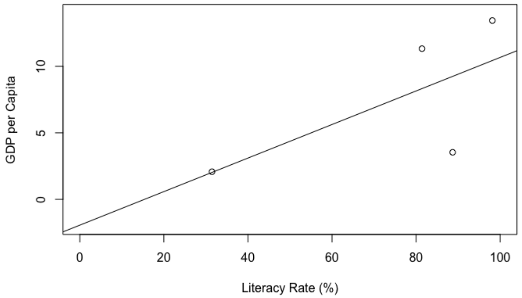
$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \left(\begin{bmatrix} 4 & 299 \\ 299 & 25{,}047 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 31 & 98 & 81 & 89 \end{bmatrix} \begin{bmatrix} 2 \\ 13 \\ 11 \\ 4 \end{bmatrix} \approx \begin{bmatrix} -1.938 \\ 0.126 \end{bmatrix}$$

Fitting a Model for Predicting Wealth of a Nation

Let's imagine that we randomly select four countries record the most recent data for each country's GDP per capita and literacy rate.

$$\widehat{\texttt{Wealth}} = -1.938 + 0.126(\texttt{Education})$$

1. The literacy rate in the United States in 2021 is approximately 80%. Based on our model predict the GDP per capita of the US? (Note the actual value is \$69,734.)
2. Interpret the practical meaning of the slope and vertical intercept of the linear model.



Multiple Regression

We can include other factors and also fit them.

Wealth	Literacy	Life Exp	Area
2	31	65	653
13	98	79	27
11	81	77	2381
4	89	61	387

$$A = \begin{bmatrix} 1 & 31 & 65 & 653 \\ 1 & 98 & 79 & 27 \\ 1 & 81 & 77 & 2381 \\ 1 & 89 & 61 & 387 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 2 \\ 13 \\ 11 \\ 4 \end{bmatrix}.$$

We wish to find the regression coefficients $\beta_0, \beta_1, \beta_2$, and β_3 that will give us the best fitting model of the form

$$\texttt{wealth} = \beta_0 + \beta_1(\texttt{Literacy}) + \beta_2(\texttt{LifeExp}) + \beta_3(\texttt{Area}) + \epsilon.$$

Multiple Regression

We can include other factors and also fit them.

Wealth	Literacy	Life Exp	Area
2	31	65	653
13	98	79	27
11	81	77	2381
4	89	61	387

$$A = \begin{bmatrix} 1 & 31 & 65 & 653 \\ 1 & 98 & 79 & 27 \\ 1 & 81 & 77 & 2381 \\ 1 & 89 & 61 & 387 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 2 \\ 13 \\ 11 \\ 4 \end{bmatrix}.$$

We wish to find the regression coefficients $\beta_0, \beta_1, \beta_2$, and β_3 that will give us the best fitting model of the form

$$\texttt{wealth} = \beta_0 + \beta_1(\texttt{Literacy}) + \beta_2(\texttt{LifeExp}) + \beta_3(\texttt{Area}) + \epsilon.$$

Solving the normal equations, we obtain

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = (X^T X)^{-1} X^T \mathbf{y} \approx \begin{bmatrix} -30.458 \\ 0.067 \\ 0.467 \\ 0.00003 \end{bmatrix}$$

$$\widehat{\texttt{wealth}} = -30.458 + 0.067(\texttt{Literacy}) + 0.467(\texttt{LifeExp}) + 0.00003(\texttt{Area})$$

Fitting Other Models

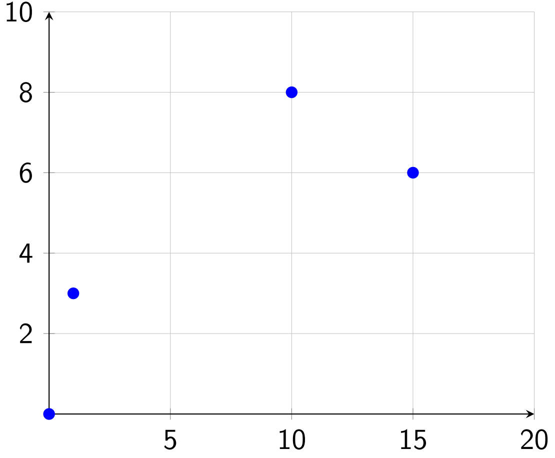
We call any model which is linear in the coefficients β's a linear model.

For example:

- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ includes two factors.
- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$ includes an interaction term.
- $y = \beta_0 + \beta_1 x + \beta_2 x^2$ is a linear model that includes a second-order term.

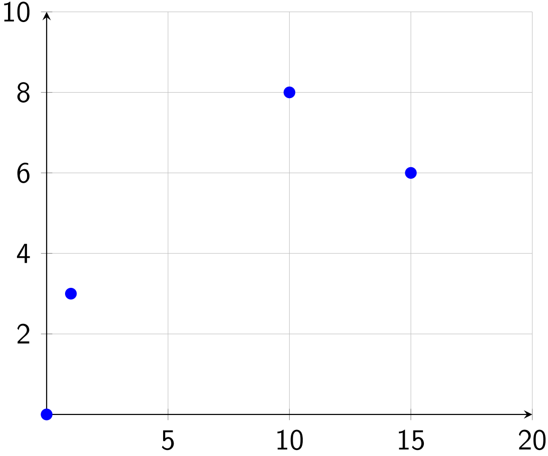
Fitting a Quadratic Polynomial

Consider a ball thrown from $(0, 0)$. The height is measured at the following distances: $(1, 3), (10, 8), (15, 6)$. Where do we predict the ball will hit the ground?



Fitting a Quadratic Polynomial

Consider a ball thrown from $(0,0)$. The height is measured at the following distances: $(1,3),(10,8),(15,6)$. Where do we predict the ball will hit the ground?

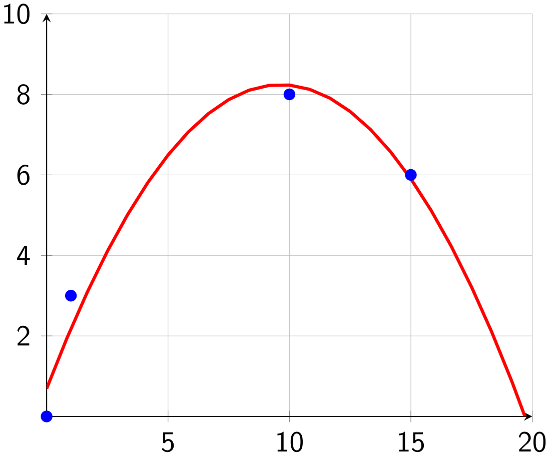


We use a model of $y=\beta_0+\beta_1 x+\beta_2 x^2$.

$$\begin{bmatrix}0\\3\\8\\6\end{bmatrix}=\begin{bmatrix}1 & 0 & 0^2\\1 & 1 & 1^2\\1 & 10 & 10^2\\1 & 15 & 15^2\end{bmatrix}\begin{bmatrix}\beta_0\\\beta_1\\\beta_2\end{bmatrix}.$$

Fitting a Quadratic Polynomial

Consider a ball thrown from $(0,0)$. The height is measured at the following distances: $(1,3),(10,8),(15,6)$. Where do we predict the ball will hit the ground?



We use a model of $y = \beta_0 + \beta_1 x + \beta_2 x^2$.

$$\begin{bmatrix} 0 \\ 3 \\ 8 \\ 6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0^2 \\ 1 & 1 & 1^2 \\ 1 & 10 & 10^2 \\ 1 & 15 & 15^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}.$$

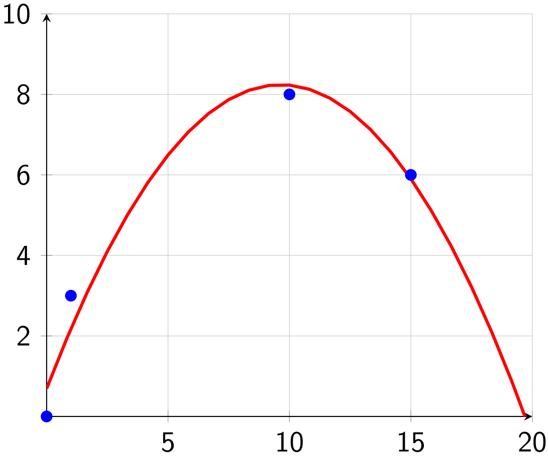
We solve the normal equations by computing

$$\boldsymbol{\beta} = (A^T A)^{-1} A^T \mathbf{y} = \begin{bmatrix} 0.691 \\ 1.567 \\ -0.0813 \end{bmatrix}.$$

So $y = 0.691 + 1.567x - 0.0813x^2$ is the best fit.

Fitting a Quadratic Polynomial

Consider a ball thrown from $(0,0)$. The height is measured at the following distances: $(1,3),(10,8),(15,6)$. Where do we predict the ball will hit the ground?



We use a model of $y = \beta_0 + \beta_1 x + \beta_2 x^2$.

$$\begin{bmatrix} 0 \\ 3 \\ 8 \\ 6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0^2 \\ 1 & 1 & 1^2 \\ 1 & 10 & 10^2 \\ 1 & 15 & 15^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}.$$

We solve the normal equations by computing

$$\boldsymbol{\beta} = (A^T A)^{-1} A^T \mathbf{y} = \begin{bmatrix} 0.691 \\ 1.567 \\ -0.0813 \end{bmatrix}.$$

So $y = 0.691 + 1.567x - 0.0813x^2$ is the best fit.

Using the quadratic formula, height $= 0$ at 19.703.