MATH 535: Mathematical Methods in Data Science

Lecture Notes: Least Squares

Feb 3, 5, 7, 10, 2025

Motivation

SLIDESHOW

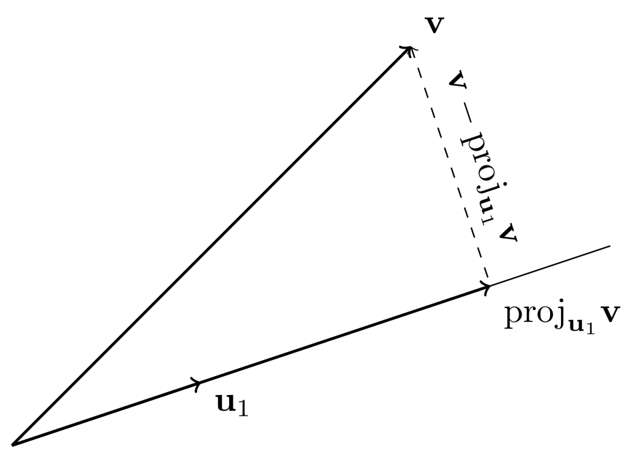
Orthogonal Projection

Geometric Intuition

Consider a vector $\mathbf{v} \notin U$ where U is a linear subspace. We want to find the vector $\mathbf{p}$ in U that is closest to $\mathbf{v}$ in Euclidean norm:

$$\min_{\mathbf{p} \in U} \|\mathbf{p} - \mathbf{v}\|$$

Key geometric insight: The optimal solution $\mathbf{p}^*$ has the property that $\mathbf{v} - \mathbf{p}^*$ is orthogonal to U.



Definition and Main Theorem

Definition 1 (Orthogonal Projection on Orthonormal List)**.** Let $\mathbf{q}_1, \ldots, \mathbf{q}_m$ be an orthonormal list. The orthogonal projection of $\mathbf{v} \in \mathbb{R}^n$ on $\{\mathbf{q}_i\}_{i=1}^m$ is:

$$\mathrm{proj}_{\{\mathbf{q}_i\}_{i=1}^m} \mathbf{v} = \sum_{j=1}^m \langle \mathbf{v}, \mathbf{q}_j \rangle \, \mathbf{q}_j$$

Theorem 1 (Orthogonal Projection)**.** Let $U \subseteq \mathbb{R}^n$ be a linear subspace and $\mathbf{v} \in \mathbb{R}^n$. Then:

a) There exists a unique solution $\mathrm{proj}_U \mathbf{v} := \mathbf{p}^* \in U$ to

$$\min_{\mathbf{p} \in U} \|\mathbf{p} - \mathbf{v}\|$$

b) The solution $\mathbf{p}^* \in U$ is characterized by

$$(*) \qquad \langle \mathbf{v} - \mathbf{p}^*, \mathbf{u} \rangle = 0, \quad \forall \mathbf{u} \in U$$

c) For any orthonormal basis $\mathbf{q}_1, \ldots, \mathbf{q}_m$ of U, $\mathrm{proj}_{\{\mathbf{q}_i\}_{i=1}^m} \mathbf{v} = \mathrm{proj}_U \mathbf{v}$

Proof: 1. For existence, we first show that any vector $\mathbf{p}^*$ satisfying $(*)$ is a minimizer:

For any $\mathbf{p} \in U$, let $\mathbf{u} = \mathbf{p} - \mathbf{p}^*$. Note $\mathbf{u} \in U$. By Pythagoras:

$$\begin{aligned} \|\mathbf{p} - \mathbf{v}\|^2 &= \|\mathbf{p} - \mathbf{p}^* + \mathbf{p}^* - \mathbf{v}\|^2 \\ &= \|\mathbf{p} - \mathbf{p}^*\|^2 + \|\mathbf{p}^* - \mathbf{v}\|^2 \\ &\geq \|\mathbf{p}^* - \mathbf{v}\|^2 \end{aligned}$$

where we used $(*)$ to get the second line. (The inequality is strict unless $\mathbf{p} = \mathbf{p}^*$. We'll need this later.)

Now we construct a vector satisfying $(*)$. Let $\mathbf{q}_1, \ldots, \mathbf{q}_m$ be an orthonormal basis of U and define:

$$\mathbf{p}^* = \sum_{j=1}^m \langle \mathbf{v}, \mathbf{q}_j \rangle \, \mathbf{q}_j$$

For any $\mathbf{u} \in U$, write:

$$\mathbf{u} = \sum_{i=1}^m \alpha_i \mathbf{q}_i \quad \text{where } \alpha_i = \langle \mathbf{u}, \mathbf{q}_i \rangle$$

Then:

$$\begin{aligned} \langle \mathbf{v} - \mathbf{p}^*, \mathbf{u} \rangle &= \left\langle \mathbf{v} - \sum_{j=1}^m \langle \mathbf{v}, \mathbf{q}_j \rangle \, \mathbf{q}_j, \sum_{i=1}^m \alpha_i \mathbf{q}_i \right\rangle \\ &= \sum_{i=1}^m \langle \mathbf{v}, \mathbf{q}_i \rangle \, \alpha_i - \sum_{i=1}^m \sum_{j=1}^m \alpha_i \langle \mathbf{v}, \mathbf{q}_j \rangle \langle \mathbf{q}_j, \mathbf{q}_i \rangle \\ &= \sum_{i=1}^m \langle \mathbf{v}, \mathbf{q}_i \rangle \, \alpha_i - \sum_{j=1}^m \alpha_j \langle \mathbf{v}, \mathbf{q}_j \rangle = 0 \end{aligned}$$

So $\mathbf{p}^*$ satisfies $(*)$. This proves existence.

2. For uniqueness, suppose $\mathbf{p}_1$ and $\mathbf{p}_2$ both satisfy $(*)$. Then by step 1, they are both minimizers. But we showed the inequality in step 1 is strict unless $\mathbf{p} = \mathbf{p}^*$, so we must have $\mathbf{p}_1 = \mathbf{p}_2$. Remains to show that any minimizer must satisfy $(*)$. Let $\mathbf{p}^*$ be a minimizer and suppose, for contradiction, that $(*)$ doesn't hold. Then there exists $\mathbf{u} \in U$ with $\langle \mathbf{v} - \mathbf{p}^*, \mathbf{u} \rangle = c \neq 0$. Consider $\mathbf{p}_t = \mathbf{p}^* + t\mathbf{u}$ for small t. Then:

$$\begin{aligned} \|\mathbf{p}_t - \mathbf{v}\|^2 &= \|(\mathbf{p}^* - \mathbf{v}) + t\mathbf{u}\|^2 \\ &= \|\mathbf{p}^* - \mathbf{v}\|^2 + 2t\langle \mathbf{v} - \mathbf{p}^*, \mathbf{u} \rangle + t^2 \|\mathbf{u}\|^2 \\ &= \|\mathbf{p}^* - \mathbf{v}\|^2 + 2tc + t^2 \|\mathbf{u}\|^2 \end{aligned}$$

For small t with appropriate sign, this is smaller than $\|\mathbf{p}^* - \mathbf{v}\|^2$, contradicting minimality. □

Example

Consider $U = \text{span}(\mathbf{w}_1, \mathbf{w}_2)$ where:

$$\mathbf{w}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{w}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

To find orthogonal projection of $\mathbf{v} = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}$:

1. First construct orthonormal basis of U using Gram-Schmidt:

a) Start with $\mathbf{w}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$. Normalize: $\mathbf{q}_1 = \frac{\mathbf{w}_1}{\|\mathbf{w}_1\|} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$

b) Take $\mathbf{w}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$ and subtract projection onto $\mathbf{q}_1$: $\mathbf{v}_2 = \mathbf{w}_2 - \langle \mathbf{w}_2, \mathbf{q}_1 \rangle \mathbf{q}_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} - \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -1/\sqrt{2} \\ 1 \\ 1/\sqrt{2} \end{pmatrix}$

c) Normalize: $\mathbf{q}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \frac{1}{\sqrt{6}} \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}$

2. Compute inner products:

$$\langle \mathbf{v}, \mathbf{q}_1 \rangle = \frac{2}{\sqrt{2}}, \quad \langle \mathbf{v}, \mathbf{q}_2 \rangle = \frac{2}{\sqrt{6}}$$

3. The projection is:

$$\text{proj}_U \mathbf{v} = \frac{2}{\sqrt{2}} \mathbf{q}_1 + \frac{2}{\sqrt{6}} \mathbf{q}_2 = \begin{pmatrix} 2/3 \\ 2/3 \\ 4/3 \end{pmatrix}$$

Knowledge Check 1: Is $\begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}$ in the span of $\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$ and $\frac{1}{\sqrt{6}} \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}$?

A) Yes

B) No

Matrix Representation of Orthogonal Projection

The orthogonal projection map proj_U is linear:

Theorem 2 (Linearity of Projection)**.** Let $U \subseteq \mathbb{R}^n$ be a subspace. For all $\alpha \in \mathbb{R}$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$:

$$\text{proj}_U(\alpha \mathbf{x} + \mathbf{y}) = \alpha \text{proj}_U \mathbf{x} + \text{proj}_U \mathbf{y}$$

Proof: Let $\mathbf{q}_1, \ldots, \mathbf{q}_m$ be an orthonormal basis of U. Then:

$$\begin{aligned}
\text{proj}_U(\alpha\mathbf{x} + \mathbf{y}) &= \sum_{j=1}^{m} \langle \alpha\mathbf{x} + \mathbf{y}, \mathbf{q}_j \rangle \mathbf{q}_j \\
&= \sum_{j=1}^{m} \{\alpha\langle \mathbf{x}, \mathbf{q}_j \rangle + \langle \mathbf{y}, \mathbf{q}_j \rangle\} \mathbf{q}_j \\
&= \alpha \text{proj}_U \mathbf{x} + \text{proj}_U \mathbf{y}
\end{aligned}$$

$\square$

Therefore, proj_U can be represented by a matrix $P \in \mathbb{R}^{n \times n}$. Let

$$Q = \begin{pmatrix} | & & | \\ \mathbf{q}_1 & \cdots & \mathbf{q}_m \\ | & & | \end{pmatrix}$$

Note that computing

$$Q^T \mathbf{v} = \begin{pmatrix} \langle \mathbf{v}, \mathbf{q}_1 \rangle \\ \vdots \\ \langle \mathbf{v}, \mathbf{q}_m \rangle \end{pmatrix}$$

gives coefficients in basis expansion of $\text{proj}_U \mathbf{v}$. Hence

$$P = QQ^T$$

Example: For $U = \text{span}(\mathbf{q}_1, \mathbf{q}_2)$ where

$$\mathbf{q}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{q}_2 = \frac{1}{\sqrt{6}} \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}$$

The projection matrix is:

$$P = QQ^T = \begin{pmatrix} 2/3 & -1/3 & 1/3 \\ -1/3 & 2/3 & 1/3 \\ 1/3 & 1/3 & 2/3 \end{pmatrix}$$

Verify: For $\mathbf{v} = (0, 0, 2)^T$,

$$P\mathbf{v} = \begin{pmatrix} 2/3 \\ 2/3 \\ 4/3 \end{pmatrix} = \text{proj}_U \mathbf{v}$$

The matrix $P = QQ^T$ is not to be confused with

$$Q^T Q = \begin{pmatrix} \langle \mathbf{q}_1, \mathbf{q}_1 \rangle & \cdots & \langle \mathbf{q}_1, \mathbf{q}_m \rangle \\ \langle \mathbf{q}_2, \mathbf{q}_1 \rangle & \cdots & \langle \mathbf{q}_2, \mathbf{q}_m \rangle \\ \vdots & \ddots & \vdots \\ \langle \mathbf{q}_m, \mathbf{q}_1 \rangle & \cdots & \langle \mathbf{q}_m, \mathbf{q}_m \rangle \end{pmatrix} = I_{m \times m}$$

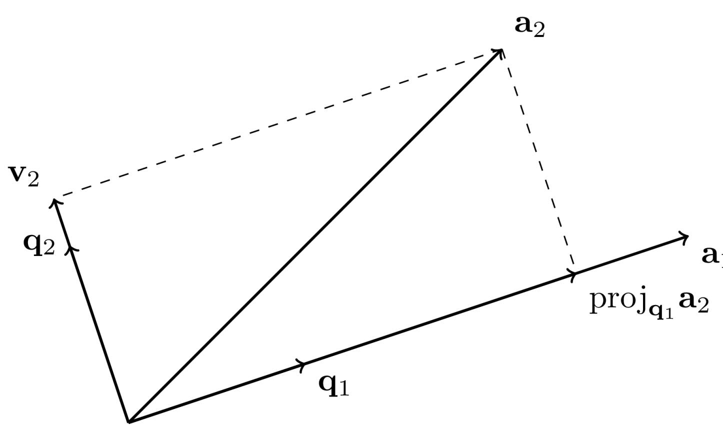
where $I_{m \times m}$ denotes the $m \times m$ identity matrix.

The Gram-Schmidt Procedure

Review of Gram-Schmidt

Given linearly independent vectors $\mathbf{a}_1, \ldots, \mathbf{a}_m$, we want to construct an orthonormal basis of their span. The Gram-Schmidt process works by:

1. Taking each vector $\mathbf{a}_j$ in turn
2. Subtracting its projection onto previously constructed orthonormal vectors
3. Normalizing the result



The key equations are:
Step 1: For first vector, just normalize:

$$\mathbf{q}_1 = \frac{\mathbf{a}_1}{\|\mathbf{a}_1\|}$$

Step 2: For second vector, subtract projection onto $\mathbf{q}_1$:

$$\mathbf{v}_2 = \mathbf{a}_2 - \langle \mathbf{a}_2, \mathbf{q}_1 \rangle \mathbf{q}_1$$
$$\mathbf{q}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|}$$

Step j: For j-th vector, subtract projections onto all previous vectors:

$$\mathbf{v}_j = \mathbf{a}_j - \sum_{i=1}^{j-1} \langle \mathbf{a}_j, \mathbf{q}_i \rangle \mathbf{q}_i$$
$$\mathbf{q}_j = \frac{\mathbf{v}_j}{\|\mathbf{v}_j\|}$$

Example: Consider:

$$\mathbf{a}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{a}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

Then:

$$\mathbf{q}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\1\end{pmatrix}$$

$$\mathbf{v}_2 = \begin{pmatrix}0\\1\\1\end{pmatrix} - \frac{1}{2}\begin{pmatrix}1\\0\\1\end{pmatrix}$$

$$\mathbf{q}_2 = \frac{1}{\sqrt{6}}\begin{pmatrix}-1\\2\\1\end{pmatrix}$$

This gives orthonormal basis $\{\mathbf{q}_1, \mathbf{q}_2\}$ of $\text{span}(\mathbf{a}_1, \mathbf{a}_2)$.

Implementation

SLIDESHOW

Matrix Form of Gram-Schmidt

Let $\mathbf{a}_1, \ldots, \mathbf{a}_m$ be linearly independent vectors. We can express the Gram-Schmidt process in matrix form. Stack the input and output vectors into matrices:

$$A = \begin{pmatrix}| & & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_m \\ | & & |\end{pmatrix} \quad \text{and} \quad Q = \begin{pmatrix}| & & | \\ \mathbf{q}_1 & \cdots & \mathbf{q}_m \\ | & & |\end{pmatrix}$$

From the Gram-Schmidt process, for all j:

$$\mathbf{a}_j = \sum_{i=1}^{j-1} \langle \mathbf{a}_j, \mathbf{q}_i \rangle \mathbf{q}_i + \|\mathbf{v}_j\| \mathbf{q}_j$$

where we define $r_{jj} = \|\mathbf{v}_j\|$ and $r_{ij} = \langle \mathbf{a}_j, \mathbf{q}_i \rangle$ for $i < j$.
Thus, there exists an upper triangular matrix R such that:

$$A = QR \quad \text{where} \quad R = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ 0 & r_{22} & \cdots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_{mm} \end{pmatrix}$$

This is called a QR decomposition of A. Key properties:

- Q has orthonormal columns: $Q^T Q = I_{m \times m}$
- R is upper triangular with positive diagonal entries
- Column j of R contains coefficients expressing $\mathbf{a}_j$ in basis $\{\mathbf{q}_1, \ldots, \mathbf{q}_j\}$

Example: For previous vectors:

$$A = \begin{pmatrix}1 & 0\\0 & 1\\1 & 1\end{pmatrix}, \quad Q = \begin{pmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ 0 & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}}\end{pmatrix}, \quad R = \begin{pmatrix}\sqrt{2} & \frac{1}{\sqrt{2}} \\ 0 & \sqrt{\frac{3}{2}}\end{pmatrix}$$

Verify: $A = QR$ and $Q^T Q = I_{2\times 2}$.

Overdetermined Systems

Normal equations

Consider an $n \times m$ matrix A with linearly independent columns and vector $\mathbf{b} \in \mathbb{R}^n$. When $n > m$ (more equations than variables), the system $A\mathbf{x} = \mathbf{b}$ typically has no solution.

Instead, we seek the "best approximate solution" by minimizing the distance to $\mathbf{b}$:

$$\min_{\mathbf{x} \in \mathbb{R}^m} \|A\mathbf{x} - \mathbf{b}\|$$

Writing A in terms of its columns:

$$A = \begin{pmatrix} | & & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_m \\ | & & | \end{pmatrix}$$

The problem becomes: find coefficients $x_1, \ldots, x_m$ minimizing

$$\left\| \sum_{j=1}^{m} x_j \mathbf{a}_j - \mathbf{b} \right\|^2 = \sum_{i=1}^{n} \left(\sum_{j=1}^{m} a_{ij} x_j - b_i \right)^2$$

Theorem 3 (Normal Equations)**.** A solution $\mathbf{x}^*$ to the least squares problem satisfies:

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

If the columns of A are linearly independent, this solution is unique.

Proof idea: Let $U = \text{col}(A)$. By the Orthogonal Projection Theorem:

- $A\mathbf{x}^*$ must be the orthogonal projection of $\mathbf{b}$ onto U
- This gives $\langle \mathbf{b} - A\mathbf{x}^*, \mathbf{a}_i \rangle = 0$ for all i
- Stacking these equations gives $A^T(\mathbf{b} - A\mathbf{x}^*) = \mathbf{0}$
- When columns are independent, $A^T A$ is invertible, giving uniqueness (see below)

Claim: When A has linearly independent columns, $A^T A$ is invertible.
Proof: Let $\mathbf{x}$ be such that $A^T A\mathbf{x} = \mathbf{0}$. Then:

$$\begin{aligned} A^T A\mathbf{x} = \mathbf{0} &\implies \mathbf{x}^T(A^T A\mathbf{x}) = 0 \\ &\implies (A\mathbf{x})^T(A\mathbf{x}) = 0 \\ &\implies \|A\mathbf{x}\|^2 = 0 \\ &\implies A\mathbf{x} = \mathbf{0} \\ &\implies \mathbf{x} = \mathbf{0} \end{aligned}$$

where the last step uses linear independence of columns. Therefore $A^T A$ is invertible.
Example: Consider fitting a line $y \approx \beta_0 + \beta_1 x$ through three points:

$$(x_1, y_1) = (0, 0), \quad (x_2, y_2) = (1, 0), \quad (x_3, y_3) = (1, 2)$$

The least squares problem minimizes:

$$\sum_{i=1}^{3}(y_i - \{\beta_0 + \beta_1 x_i\})^2$$

Writing in matrix form with a column of ones for the intercept:

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{y} = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}$$

Then:

$$A^T A = \begin{pmatrix} 3 & 2 \\ 2 & 2 \end{pmatrix} \quad \text{and} \quad A^T \mathbf{y} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

Solving $A^T A\boldsymbol{\beta} = A^T\mathbf{y}$ gives the system:

$$\begin{pmatrix} 3 & 2 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

This is equivalent to:

$$3\beta_0 + 2\beta_1 = 2$$
$$2\beta_0 + 2\beta_1 = 2$$

Subtracting the second equation from the first:

$$\beta_0 = 0$$

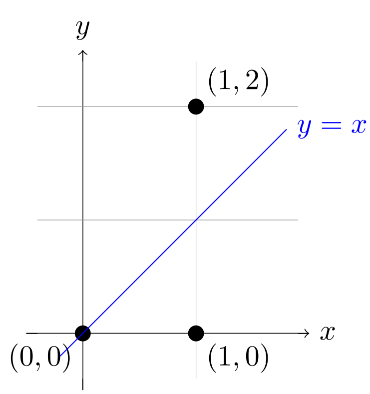
Substituting back:

$$2\beta_1 = 2 \implies \beta_1 = 1$$

Therefore:

$$\boldsymbol{\beta}^* = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

This corresponds to the line $y = x$, which fits the points $(0, 0)$, $(1, 0)$, and $(1, 2)$ in a least squares sense.



Least Squares via QR

Let $A \in \mathbb{R}^{n\times m}$ be a matrix with linearly independent columns and $\mathbf{b} \in \mathbb{R}^n$. Consider solving

$$\min_{\mathbf{x}\in\mathbb{R}^m} \|A\mathbf{x} - \mathbf{b}\|$$

Strategy using QR decomposition:

1. Compute QR decomposition: $A = QR$
2. Then $A\mathbf{x} = QR\mathbf{x}$ and $QQ^T\mathbf{b}$ is projection of $\mathbf{b}$ onto col(A)
3. By normal equations: $A\mathbf{x}^* = QQ^T\mathbf{b}$
4. Substitute $A = QR$:
$$QR\mathbf{x}^* = QQ^T\mathbf{b}$$
5. Multiply both sides by Q^T and use $Q^TQ = I_{m\times m}$:
$$R\mathbf{x}^* = Q^T\mathbf{b}$$
6. Solve this triangular system using back substitution.

Example: Consider again fitting a line $y \approx \beta_0 + \beta_1 x$ through three points:

$$(x_1, y_1) = (0, 0), \quad (x_2, y_2) = (1, 0), \quad (x_3, y_3) = (1, 2)$$

Using previous matrix form:

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{y} = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}$$

Let's compute the QR decomposition using Gram-Schmidt. Start with the columns of A:

$$\mathbf{a}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{a}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

Step 1: Normalize $\mathbf{a}_1$ to get first column of Q:

$$\|\mathbf{a}_1\| = \sqrt{3} \implies \mathbf{q}_1 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Step 2: Compute $\mathbf{v}_2$ by removing projection of $\mathbf{a}_2$ onto $\mathbf{q}_1$:

$$\langle \mathbf{a}_2, \mathbf{q}_1 \rangle = \frac{2}{\sqrt{3}}$$

$$\mathbf{v}_2 = \mathbf{a}_2 - \langle \mathbf{a}_2, \mathbf{q}_1 \rangle \mathbf{q}_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} - \frac{2}{\sqrt{3}} \cdot \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -2/3 \\ 1/3 \\ 1/3 \end{pmatrix}$$

Normalize $\mathbf{v}_2$ to get second column of Q:

$$\|\mathbf{v}_2\| = \sqrt{\frac{2}{3}} \implies \mathbf{q}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \begin{pmatrix} -\sqrt{2/3} \\ 1/\sqrt{6} \\ 1/\sqrt{6} \end{pmatrix}$$

Therefore:

$$Q = \begin{pmatrix} 1/\sqrt{3} & -\sqrt{2/3} \\ 1/\sqrt{3} & 1/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{6} \end{pmatrix}, \quad R = \begin{pmatrix} \sqrt{3} & 2/\sqrt{3} \\ 0 & \sqrt{2/3} \end{pmatrix},$$

where the matrix R comes from writing each column of A in terms of the orthonormal basis $\{\mathbf{q}_1, \mathbf{q}_2\}$.
For the first column:

$$\mathbf{a}_1 = \|\mathbf{a}_1\|\mathbf{q}_1 = \sqrt{3}\mathbf{q}_1$$

So $r_{11} = \sqrt{3}$ and $r_{21} = 0$. For the second column:

$$\mathbf{a}_2 = \langle \mathbf{a}_2, \mathbf{q}_1 \rangle \mathbf{q}_1 + \|\mathbf{v}_2\|\mathbf{q}_2 = \frac{2}{\sqrt{3}}\mathbf{q}_1 + \sqrt{\frac{2}{3}}\mathbf{q}_2$$

So $r_{12} = \frac{2}{\sqrt{3}}$ and $r_{22} = \sqrt{\frac{2}{3}}$.
Then solve:

$$Q^T\mathbf{y} = \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ -\sqrt{2/3} & 1/\sqrt{6} & 1/\sqrt{6} \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 2/\sqrt{3} \\ 2/\sqrt{6} \end{pmatrix}$$

$$R\boldsymbol{\beta}^* = Q^T\mathbf{y}$$

This gives the system:

$$\sqrt{3}\beta_0 + \frac{2}{\sqrt{3}}\beta_1 = \frac{2}{\sqrt{3}}$$
$$\sqrt{\frac{2}{3}}\beta_1 = \frac{2}{\sqrt{6}}$$

From the second equation:

$$\beta_1 = \frac{2/\sqrt{6}}{\sqrt{2/3}} = 1$$

Substituting into the first equation:

$$\sqrt{3}\beta_0 + \frac{2}{\sqrt{3}} = \frac{2}{\sqrt{3}} \implies \beta_0 = 0$$

Therefore:

$$\boldsymbol{\beta}^* = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

This gives the same line $y = x$ that we found using normal equations, but without explicitly forming A^TA.
Key advantages:

- No explicit computation of A^TA needed
- More numerically stable than normal equations

Implementation

SLIDESHOW

Back to the dataset

SLIDESHOW