Lecture 26

Math 22 Summer 2017
August 16, 2017

Just for today

- §6.4 Finish up
- §6.5 Least-squares problems

§6.4 Gram-Schmidt Review

Consider a basis $\{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ for a subspace $W \subseteq \mathbb{R}^n$.

§6.4 Gram-Schmidt Review

Consider a basis $\{\mathbf{x}_1, \dots, \mathbf{x}_p\}$ for a subspace $W \subseteq \mathbb{R}^n$. We saw that the Gram-Schmidt algorithm produces an orthogonal basis $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$

§6.4 Gram-Schmidt Review

Consider a basis $\{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ for a subspace $W \subseteq \mathbb{R}^n$. We saw that the Gram-Schmidt algorithm produces an orthogonal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ where $\mathbf{v}_1 = \mathbf{x}_1$ and for $k \in \{2, \ldots, p\}$ we have

$$\mathbf{v}_k = \mathbf{x}_k - \left(\sum_{i=1}^{k-1} \frac{\mathbf{x}_k \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i} \mathbf{v}_i \right).$$

Consider a basis $\{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ for a subspace $W \subseteq \mathbb{R}^n$. We saw that the Gram-Schmidt algorithm produces an orthogonal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ where $\mathbf{v}_1 = \mathbf{x}_1$ and for $k \in \{2, \ldots, p\}$ we have

$$\mathbf{v}_k = \mathbf{x}_k - \left(\sum_{i=1}^{k-1} \frac{\mathbf{x}_k \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i} \mathbf{v}_i \right).$$

Moreover, $\mathrm{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} = \mathrm{Span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ for every $k \in \{1, \ldots, p\}$.

§6.4 QR factorization review

§6.4 QR factorization review

Let $A = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_n]$ be an $m \times n$ matrix with linearly independent columns.

§6.4 *QR* factorization review

Let $A = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_n]$ be an $m \times n$ matrix with linearly independent columns. Via the Gram-Schmidt algorithm we can construct an *orthonormal* basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ for $\mathrm{Col}\, A$.

§6.4 QR factorization review

Let $A = [\mathbf{x}_1 \;\cdots\; \mathbf{x}_n]$ be an $m \times n$ matrix with linearly independent columns. Via the Gram-Schmidt algorithm we can construct an *orthonormal* basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ for $\operatorname{Col} A$. Let $Q = [\mathbf{u}_1 \;\cdots\; \mathbf{u}_n]$.

§6.4 QR factorization review

Let $A = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_n]$ be an $m \times n$ matrix with linearly independent columns. Via the Gram-Schmidt algorithm we can construct an *orthonormal* basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ for $\mathrm{Col}\, A$. Let $Q = [\mathbf{u}_1 \ \cdots \ \mathbf{u}_n]$. Since $\mathbf{x}_k \in \mathrm{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$,

§6.4 *QR* factorization review

Let $A = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_n]$ be an $m \times n$ matrix with linearly independent columns. Via the Gram-Schmidt algorithm we can construct an *orthonormal* basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ for $\mathrm{Col}\, A$. Let $Q = [\mathbf{u}_1 \ \cdots \ \mathbf{u}_n]$. Since $\mathbf{x}_k \in \mathrm{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$, we can write

$$\mathbf{x}_k = r_{1k}\mathbf{u}_1 + \cdots + r_{kk}\mathbf{u}_k + 0\mathbf{u}_{k+1} + \cdots + 0\mathbf{u}_n = Q\mathbf{r}_k, \quad \mathbf{r}_k = \begin{bmatrix} r_{1k} \\ \vdots \\ r_{kk} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Let $A = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_n]$ be an $m \times n$ matrix with linearly independent columns. Via the Gram-Schmidt algorithm we can construct an *orthonormal* basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ for $\mathrm{Col}\, A$. Let $Q = [\mathbf{u}_1 \ \cdots \ \mathbf{u}_n]$. Since $\mathbf{x}_k \in \mathrm{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$, we can write

$$\mathbf{x}_k = r_{1k}\mathbf{u}_1 + \cdots + r_{kk}\mathbf{u}_k + 0\mathbf{u}_{k+1} + \cdots + 0\mathbf{u}_n = Q\mathbf{r}_k, \quad \mathbf{r}_k = \begin{bmatrix} r_{1k} \\ \vdots \\ r_{kk} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Then $R := [\mathbf{r}_1 \ \cdots \ \mathbf{r}_n]$ is upper triangular

§6.4 QR factorization review

Let $A = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_n]$ be an $m \times n$ matrix with linearly independent columns. Via the Gram-Schmidt algorithm we can construct an *orthonormal* basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ for $\operatorname{Col} A$. Let $Q = [\mathbf{u}_1 \ \cdots \ \mathbf{u}_n]$. Since $\mathbf{x}_k \in \operatorname{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$, we can write

$$\mathbf{x}_k = r_{1k}\mathbf{u}_1 + \cdots + r_{kk}\mathbf{u}_k + 0\mathbf{u}_{k+1} + \cdots + 0\mathbf{u}_n = Q\mathbf{r}_k, \quad \mathbf{r}_k = \begin{bmatrix} r_{1k} \\ \vdots \\ r_{kk} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Then $R := [\mathbf{r}_1 \ \cdots \ \mathbf{r}_n]$ is upper triangular and $A = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_n] = [Q\mathbf{r}_1 \ \cdots \ Q\mathbf{r}_n] = QR$.

§6.4 *QR* factorization review

Let $A = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_n]$ be an $m \times n$ matrix with linearly independent columns. Via the Gram-Schmidt algorithm we can construct an *orthonormal* basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ for $\mathrm{Col}\, A$. Let $Q = [\mathbf{u}_1 \ \cdots \ \mathbf{u}_n]$. Since $\mathbf{x}_k \in \mathrm{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$, we can write

$$\mathbf{x}_k = r_{1k}\mathbf{u}_1 + \cdots + r_{kk}\mathbf{u}_k + 0\mathbf{u}_{k+1} + \cdots + 0\mathbf{u}_n = Q\mathbf{r}_k, \quad \mathbf{r}_k = \begin{bmatrix} r_{1k} \\ \vdots \\ r_{kk} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Then $R := [\mathbf{r}_1 \ \cdots \ \mathbf{r}_n]$ is upper triangular and $A = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_n] = [Q\mathbf{r}_1 \ \cdots \ Q\mathbf{r}_n] = QR$. How do we guarantee that the diagonal of R is nonnegative?

§6.5 Least-squares solutions

§6.5 Least-squares solutions

Definition

Definition

Let A be an $m \times n$ matrix.

Definition

Let A be an $m \times n$ matrix. Let $\mathbf{b} \in \mathbb{R}^n$.

§6.5 Least-squares solutions

Definition

Let A be an $m \times n$ matrix. Let $\mathbf{b} \in \mathbb{R}^n$. A **least-squares solution** of the matrix equation $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}} \in \mathbb{R}^n$ such that

§6.5 Least-squares solutions

Definition

Let A be an $m \times n$ matrix. Let $\mathbf{b} \in \mathbb{R}^n$. A **least-squares solution** of the matrix equation $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}} \in \mathbb{R}^n$ such that

$$\|A\hat{\mathbf{x}} - \mathbf{b}\| \leq \|A\mathbf{x} - \mathbf{b}\|$$

§6.5 Least-squares solutions

Definition

Let A be an $m \times n$ matrix. Let $\mathbf{b} \in \mathbb{R}^n$. A **least-squares solution** of the matrix equation $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}} \in \mathbb{R}^n$ such that

$$\|A\hat{\mathbf{x}} - \mathbf{b}\| \leq \|A\mathbf{x} - \mathbf{b}\|$$

for all $\mathbf{x} \in \mathbb{R}^n$.

Definition

Let A be an $m \times n$ matrix. Let $\mathbf{b} \in \mathbb{R}^n$. A **least-squares solution** of the matrix equation $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}} \in \mathbb{R}^n$ such that

$$\|A\hat{\mathbf{x}} - \mathbf{b}\| \leq \|A\mathbf{x} - \mathbf{b}\|$$

for all $\mathbf{x} \in \mathbb{R}^n$.

The **least-squares error** is $\|\mathbf{b} - A\hat{\mathbf{x}}\|$

Definition

Let A be an $m \times n$ matrix. Let $\mathbf{b} \in \mathbb{R}^n$. A **least-squares solution** of the matrix equation $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}} \in \mathbb{R}^n$ such that

$$\|A\hat{\mathbf{x}} - \mathbf{b}\| \leq \|A\mathbf{x} - \mathbf{b}\|$$

for all $\mathbf{x} \in \mathbb{R}^n$.

The **least-squares error** is $\|\mathbf{b} - A\hat{\mathbf{x}}\|$ (the distance between $\mathbf{b}$ and $A\hat{\mathbf{x}}$).

§6.5 Least-squares solutions

Definition

Let A be an $m \times n$ matrix. Let $\mathbf{b} \in \mathbb{R}^n$. A **least-squares solution** of the matrix equation $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}} \in \mathbb{R}^n$ such that

$$\|A\hat{\mathbf{x}} - \mathbf{b}\| \leq \|A\mathbf{x} - \mathbf{b}\|$$

for all $\mathbf{x} \in \mathbb{R}^n$.

The **least-squares error** is $\|\mathbf{b} - A\hat{\mathbf{x}}\|$ (the distance between $\mathbf{b}$ and $A\hat{\mathbf{x}}$).

We can obtain a least-squares solution via projection.

§6.5 Least-squares solutions

Definition

Let A be an $m \times n$ matrix. Let $\mathbf{b} \in \mathbb{R}^n$. A **least-squares solution** of the matrix equation $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}} \in \mathbb{R}^n$ such that

$$\|A\hat{\mathbf{x}} - \mathbf{b}\| \leq \|A\mathbf{x} - \mathbf{b}\|$$

for all $\mathbf{x} \in \mathbb{R}^n$.

The **least-squares error** is $\|\mathbf{b} - A\hat{\mathbf{x}}\|$ (the distance between $\mathbf{b}$ and $A\hat{\mathbf{x}}$).

We can obtain a least-squares solution via projection. Let $W = \operatorname{Col} A$.

Definition

Let A be an $m \times n$ matrix. Let $\mathbf{b} \in \mathbb{R}^n$. A **least-squares solution** of the matrix equation $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}} \in \mathbb{R}^n$ such that

$$\|A\hat{\mathbf{x}} - \mathbf{b}\| \leq \|A\mathbf{x} - \mathbf{b}\|$$

for all $\mathbf{x} \in \mathbb{R}^n$.

The **least-squares error** is $\|\mathbf{b} - A\hat{\mathbf{x}}\|$ (the distance between $\mathbf{b}$ and $A\hat{\mathbf{x}}$).

We can obtain a least-squares solution via projection. Let $W = \mathrm{Col}\, A$. Let $\hat{\mathbf{b}} = \mathrm{proj}_W \mathbf{b}$.

Definition

Let A be an $m \times n$ matrix. Let $\mathbf{b} \in \mathbb{R}^n$. A **least-squares solution** of the matrix equation $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}} \in \mathbb{R}^n$ such that

$$\|A\hat{\mathbf{x}} - \mathbf{b}\| \leq \|A\mathbf{x} - \mathbf{b}\|$$

for all $\mathbf{x} \in \mathbb{R}^n$.

The **least-squares error** is $\|\mathbf{b} - A\hat{\mathbf{x}}\|$ (the distance between $\mathbf{b}$ and $A\hat{\mathbf{x}}$).

We can obtain a least-squares solution via projection. Let $W = \operatorname{Col} A$. Let $\hat{\mathbf{b}} = \operatorname{proj}_W \mathbf{b}$. Then $A\mathbf{x} = \hat{\mathbf{b}}$ is consistent,

§6.5 Least-squares solutions

Definition

Let A be an $m \times n$ matrix. Let $\mathbf{b} \in \mathbb{R}^n$. A **least-squares solution** of the matrix equation $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}} \in \mathbb{R}^n$ such that

$$\|A\hat{\mathbf{x}} - \mathbf{b}\| \leq \|A\mathbf{x} - \mathbf{b}\|$$

for all $\mathbf{x} \in \mathbb{R}^n$.

The **least-squares error** is $\|\mathbf{b} - A\hat{\mathbf{x}}\|$ (the distance between $\mathbf{b}$ and $A\hat{\mathbf{x}}$).

We can obtain a least-squares solution via projection. Let $W = \mathrm{Col}\, A$. Let $\hat{\mathbf{b}} = \mathrm{proj}_W\, \mathbf{b}$. Then $A\mathbf{x} = \hat{\mathbf{b}}$ is consistent, and any solution $\hat{\mathbf{x}}$ is a least-squares solution.

§6.5 Least-squares solutions

Definition

Let A be an $m \times n$ matrix. Let $\mathbf{b} \in \mathbb{R}^n$. A **least-squares solution** of the matrix equation $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}} \in \mathbb{R}^n$ such that

$$\|A\hat{\mathbf{x}} - \mathbf{b}\| \leq \|A\mathbf{x} - \mathbf{b}\|$$

for all $\mathbf{x} \in \mathbb{R}^n$.

The **least-squares error** is $\|\mathbf{b} - A\hat{\mathbf{x}}\|$ (the distance between $\mathbf{b}$ and $A\hat{\mathbf{x}}$).

We can obtain a least-squares solution via projection. Let $W = \mathrm{Col}\, A$. Let $\hat{\mathbf{b}} = \mathrm{proj}_W \mathbf{b}$. Then $A\mathbf{x} = \hat{\mathbf{b}}$ is consistent, and any solution $\hat{\mathbf{x}}$ is a least-squares solution. Why?

Definition

Let A be an $m \times n$ matrix. Let $\mathbf{b} \in \mathbb{R}^n$. A **least-squares solution** of the matrix equation $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}} \in \mathbb{R}^n$ such that

$$\|A\hat{\mathbf{x}} - \mathbf{b}\| \leq \|A\mathbf{x} - \mathbf{b}\|$$

for all $\mathbf{x} \in \mathbb{R}^n$.

The **least-squares error** is $\|\mathbf{b} - A\hat{\mathbf{x}}\|$ (the distance between $\mathbf{b}$ and $A\hat{\mathbf{x}}$).

We can obtain a least-squares solution via projection. Let $W = \operatorname{Col} A$. Let $\hat{\mathbf{b}} = \operatorname{proj}_W \mathbf{b}$. Then $A\mathbf{x} = \hat{\mathbf{b}}$ is consistent, and any solution $\hat{\mathbf{x}}$ is a least-squares solution. Why? Best approximation theorem.

§6.5 Least-squares solutions

Definition

Let A be an $m \times n$ matrix. Let $\mathbf{b} \in \mathbb{R}^n$. A **least-squares solution** of the matrix equation $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}} \in \mathbb{R}^n$ such that

$$\|A\hat{\mathbf{x}} - \mathbf{b}\| \leq \|A\mathbf{x} - \mathbf{b}\|$$

for all $\mathbf{x} \in \mathbb{R}^n$.

The **least-squares error** is $\|\mathbf{b} - A\hat{\mathbf{x}}\|$ (the distance between $\mathbf{b}$ and $A\hat{\mathbf{x}}$).

We can obtain a least-squares solution via projection. Let $W = \operatorname{Col} A$. Let $\hat{\mathbf{b}} = \operatorname{proj}_W \mathbf{b}$. Then $A\mathbf{x} = \hat{\mathbf{b}}$ is consistent, and any solution $\hat{\mathbf{x}}$ is a least-squares solution. Why? Best approximation theorem.

Let's see how the details work.

§6.5 Example ☕

§6.5 Example

Let $A = [\mathbf{a}_1\ \mathbf{a}_2] = \begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 8 & 1 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}$.

§6.5 Example ☕

Let $A = [\mathbf{a}_1\ \mathbf{a}_2] = \begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 8 & 1 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}$. Is $A\mathbf{x} = \mathbf{b}$ consistent?

§6.5 Example

Let $A = [\mathbf{a}_1 \ \mathbf{a}_2] = \begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 8 & 1 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}$. Is $A\mathbf{x} = \mathbf{b}$ consistent?

No!

§6.5 Example ☕

Let $A = [\mathbf{a}_1\ \mathbf{a}_2] = \begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 8 & 1 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}$. Is $A\mathbf{x} = \mathbf{b}$ consistent?

No! Let's find a least-squares solution.

§6.5 Example ☕

Let $A = [\mathbf{a}_1\ \mathbf{a}_2] = \begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 8 & 1 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}$. Is $A\mathbf{x} = \mathbf{b}$ consistent?

No! Let's find a least-squares solution.

Let $W = \mathrm{Col}\, A$ and verify that $\{\mathbf{a}_1, \mathbf{a}_2\}$ is a basis.

§6.5 Example ☕

Let $A = [\mathbf{a}_1\ \mathbf{a}_2] = \begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 8 & 1 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}$. Is $A\mathbf{x} = \mathbf{b}$ consistent?

No! Let's find a least-squares solution.

Let $W = \mathrm{Col}\, A$ and verify that $\{\mathbf{a}_1, \mathbf{a}_2\}$ is a basis. Let $\hat{\mathbf{b}} = \mathrm{proj}_W\, \mathbf{b}$.

§6.5 Example

Let $A = [\mathbf{a}_1\ \mathbf{a}_2] = \begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 8 & 1 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}$. Is $A\mathbf{x} = \mathbf{b}$ consistent?

No! Let's find a least-squares solution.

Let $W = \mathrm{Col}\, A$ and verify that $\{\mathbf{a}_1, \mathbf{a}_2\}$ is a basis. Let $\hat{\mathbf{b}} = \mathrm{proj}_W \mathbf{b}$. To compute $\hat{\mathbf{b}}$ there are many options.

Let $A = [\mathbf{a}_1\ \mathbf{a}_2] = \begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 8 & 1 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}$. Is $A\mathbf{x} = \mathbf{b}$ consistent?

No! Let's find a least-squares solution.

Let $W = \mathrm{Col}\, A$ and verify that $\{\mathbf{a}_1, \mathbf{a}_2\}$ is a basis. Let $\hat{\mathbf{b}} = \mathrm{proj}_W\, \mathbf{b}$. To compute $\hat{\mathbf{b}}$ there are many options. One is to use an orthogonal basis for W.

Let $A = [\mathbf{a}_1\ \mathbf{a}_2] = \begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 8 & 1 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}$. Is $A\mathbf{x} = \mathbf{b}$ consistent?

No! Let's find a least-squares solution.

Let $W = \mathrm{Col}\, A$ and verify that $\{\mathbf{a}_1, \mathbf{a}_2\}$ is a basis. Let $\hat{\mathbf{b}} = \mathrm{proj}_W\, \mathbf{b}$. To compute $\hat{\mathbf{b}}$ there are many options. One is to use an orthogonal basis for W. Glad we know how to find these!

§6.5 Example ☕

Let $A = [\mathbf{a}_1\ \mathbf{a}_2] = \begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 8 & 1 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}$. Is $A\mathbf{x} = \mathbf{b}$ consistent?

No! Let's find a least-squares solution.

Let $W = \operatorname{Col} A$ and verify that $\{\mathbf{a}_1, \mathbf{a}_2\}$ is a basis. Let $\hat{\mathbf{b}} = \operatorname{proj}_W \mathbf{b}$. To compute $\hat{\mathbf{b}}$ there are many options. One is to use an orthogonal basis for W. Glad we know how to find these! Let $\mathbf{v}_1 = \mathbf{a}_1$.

Let $A = [\mathbf{a}_1\ \mathbf{a}_2] = \begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 8 & 1 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}$. Is $A\mathbf{x} = \mathbf{b}$ consistent?

No! Let's find a least-squares solution.

Let $W = \mathrm{Col}\, A$ and verify that $\{\mathbf{a}_1, \mathbf{a}_2\}$ is a basis. Let $\hat{\mathbf{b}} = \mathrm{proj}_W \mathbf{b}$. To compute $\hat{\mathbf{b}}$ there are many options. One is to use an orthogonal basis for W. Glad we know how to find these! Let $\mathbf{v}_1 = \mathbf{a}_1$. Let

$$\mathbf{v}_2 = \mathbf{a}_2 - \frac{\mathbf{a}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 = \begin{bmatrix} 49/71 \\ 16/71 \\ -6/71 \\ -17/71 \end{bmatrix}.$$

Let $A = [\mathbf{a}_1\ \mathbf{a}_2] = \begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 8 & 1 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}$. Is $A\mathbf{x} = \mathbf{b}$ consistent?

No! Let's find a least-squares solution.

Let $W = \mathrm{Col}\, A$ and verify that $\{\mathbf{a}_1, \mathbf{a}_2\}$ is a basis. Let $\hat{\mathbf{b}} = \mathrm{proj}_W \mathbf{b}$. To compute $\hat{\mathbf{b}}$ there are many options. One is to use an orthogonal basis for W. Glad we know how to find these! Let $\mathbf{v}_1 = \mathbf{a}_1$. Let

$$\mathbf{v}_2 = \mathbf{a}_2 - \frac{\mathbf{a}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 = \begin{bmatrix} 49/71 \\ 16/71 \\ -6/71 \\ -17/71 \end{bmatrix}.$$

So W has orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2\}$.

§6.5 Example ☕ continued

We can now compute

$$\hat{\mathbf{b}} = \frac{\mathbf{b} \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 + \frac{\mathbf{b} \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 = \begin{bmatrix} 1 \\ 29/14 \\ 39/14 \\ 22/7 \end{bmatrix}.$$

We can now compute

$$\hat{\mathbf{b}} = \frac{\mathbf{b}\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 + \frac{\mathbf{b}\cdot\mathbf{v}_2}{\mathbf{v}_2\cdot\mathbf{v}_2}\mathbf{v}_2 = \begin{bmatrix} 1 \\ 29/14 \\ 39/14 \\ 22/7 \end{bmatrix}.$$

Now to find a least-squares solution we can simply use the augmented matrix

We can now compute

$$\hat{\mathbf{b}} = \frac{\mathbf{b}\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 + \frac{\mathbf{b}\cdot\mathbf{v}_2}{\mathbf{v}_2\cdot\mathbf{v}_2}\mathbf{v}_2 = \begin{bmatrix} 1 \\ 29/14 \\ 39/14 \\ 22/7 \end{bmatrix}.$$

Now to find a least-squares solution we can simply use the augmented matrix

$$[A\ \hat{\mathbf{b}}] \sim \begin{bmatrix} 1 & 0 & 5/14 \\ 0 & 1 & 2/7 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

We can now compute

$$\hat{\mathbf{b}} = \frac{\mathbf{b}\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 + \frac{\mathbf{b}\cdot\mathbf{v}_2}{\mathbf{v}_2\cdot\mathbf{v}_2}\mathbf{v}_2 = \begin{bmatrix} 1 \\ 29/14 \\ 39/14 \\ 22/7 \end{bmatrix}.$$

Now to find a least-squares solution we can simply use the augmented matrix

$$[A\ \hat{\mathbf{b}}] \sim \begin{bmatrix} 1 & 0 & 5/14 \\ 0 & 1 & 2/7 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

to get that $\hat{\mathbf{x}} = \begin{bmatrix} 5/14 \\ 2/7 \end{bmatrix}$ is a least-squares solutions to $A\mathbf{x} = \mathbf{b}$.

§6.5 Example ☕ continued

§6.5 Example ☕ continued

Alternatively, we can use the QR factorization of A to compute $\hat{\mathbf{b}}$.

§6.5 Example ☕ continued

Alternatively, we can use the QR factorization of A to compute $\hat{\mathbf{b}}$. Let $\mathbf{u}_i = \mathbf{v}_i / \|\mathbf{v}_i\|$ for $i = 1, 2$.

Alternatively, we can use the QR factorization of A to compute $\hat{\mathbf{b}}$. Let $\mathbf{u}_i = \mathbf{v}_i / \|\mathbf{v}_i\|$ for $i = 1, 2$. Then

$$Q = [\mathbf{u}_1 \ \mathbf{u}_2] = \begin{bmatrix} 2/\sqrt{142} & (49/71)/\sqrt{42/71} \\ 5/\sqrt{142} & (16/71)/\sqrt{42/71} \\ 7/\sqrt{142} & (-6/71)/\sqrt{42/71} \\ 8/\sqrt{142} & (-17/71)/\sqrt{42/71} \end{bmatrix}$$

Alternatively, we can use the QR factorization of A to compute $\hat{\mathbf{b}}$. Let $\mathbf{u}_i = \mathbf{v}_i / \|\mathbf{v}_i\|$ for $i = 1, 2$. Then

$$Q = [\mathbf{u}_1\ \mathbf{u}_2] = \begin{bmatrix} 2/\sqrt{142} & (49/71)/\sqrt{42/71} \\ 5/\sqrt{142} & (16/71)/\sqrt{42/71} \\ 7/\sqrt{142} & (-6/71)/\sqrt{42/71} \\ 8/\sqrt{142} & (-17/71)/\sqrt{42/71} \end{bmatrix}$$

and

$$R = Q^T A = \begin{bmatrix} \sqrt{142} & (11/71)\sqrt{142} \\ 0 & \sqrt{42/71} \end{bmatrix}$$

Alternatively, we can use the QR factorization of A to compute $\hat{\mathbf{b}}$. Let $\mathbf{u}_i = \mathbf{v}_i / \|\mathbf{v}_i\|$ for $i = 1, 2$. Then

$$Q = [\mathbf{u}_1\ \mathbf{u}_2] = \begin{bmatrix} 2/\sqrt{142} & (49/71)/\sqrt{42/71} \\ 5/\sqrt{142} & (16/71)/\sqrt{42/71} \\ 7/\sqrt{142} & (-6/71)/\sqrt{42/71} \\ 8/\sqrt{142} & (-17/71)/\sqrt{42/71} \end{bmatrix}$$

and

$$R = Q^T A = \begin{bmatrix} \sqrt{142} & (11/71)\sqrt{142} \\ 0 & \sqrt{42/71} \end{bmatrix}$$

and

$$QQ^T = \begin{bmatrix} 5/6 & 1/3 & 0 & -1/6 \\ 1/3 & 11/42 & 3/14 & 4/21 \\ 0 & 3/14 & 5/14 & 3/7 \\ -1/6 & 4/21 & 3/7 & 23/42 \end{bmatrix}.$$

Alternatively, we can use the QR factorization of A to compute $\hat{\mathbf{b}}$. Let $\mathbf{u}_i = \mathbf{v}_i / \|\mathbf{v}_i\|$ for $i = 1, 2$. Then

$$Q = [\mathbf{u}_1 \ \mathbf{u}_2] = \begin{bmatrix} 2/\sqrt{142} & (49/71)/\sqrt{42/71} \\ 5/\sqrt{142} & (16/71)/\sqrt{42/71} \\ 7/\sqrt{142} & (-6/71)/\sqrt{42/71} \\ 8/\sqrt{142} & (-17/71)/\sqrt{42/71} \end{bmatrix}$$

and

$$R = Q^T A = \begin{bmatrix} \sqrt{142} & (11/71)\sqrt{142} \\ 0 & \sqrt{42/71} \end{bmatrix}$$

and

$$QQ^T = \begin{bmatrix} 5/6 & 1/3 & 0 & -1/6 \\ 1/3 & 11/42 & 3/14 & 4/21 \\ 0 & 3/14 & 5/14 & 3/7 \\ -1/6 & 4/21 & 3/7 & 23/42 \end{bmatrix}.$$

What is $\hat{\mathbf{b}}$?

Alternatively, we can use the QR factorization of A to compute $\hat{\mathbf{b}}$. Let $\mathbf{u}_i = \mathbf{v}_i / \|\mathbf{v}_i\|$ for $i = 1, 2$. Then

$$Q = [\mathbf{u}_1\ \mathbf{u}_2] = \begin{bmatrix} 2/\sqrt{142} & (49/71)/\sqrt{42/71} \\ 5/\sqrt{142} & (16/71)/\sqrt{42/71} \\ 7/\sqrt{142} & (-6/71)/\sqrt{42/71} \\ 8/\sqrt{142} & (-17/71)/\sqrt{42/71} \end{bmatrix}$$

and

$$R = Q^T A = \begin{bmatrix} \sqrt{142} & (11/71)\sqrt{142} \\ 0 & \sqrt{42/71} \end{bmatrix}$$

and

$$QQ^T = \begin{bmatrix} 5/6 & 1/3 & 0 & -1/6 \\ 1/3 & 11/42 & 3/14 & 4/21 \\ 0 & 3/14 & 5/14 & 3/7 \\ -1/6 & 4/21 & 3/7 & 23/42 \end{bmatrix}.$$

What is $\hat{\mathbf{b}}$? $\hat{\mathbf{b}} = QQ^T\mathbf{b}$.

Alternatively, we can use the QR factorization of A to compute $\hat{\mathbf{b}}$. Let $\mathbf{u}_i = \mathbf{v}_i / \|\mathbf{v}_i\|$ for $i = 1, 2$. Then

$$Q = [\mathbf{u}_1\ \mathbf{u}_2] = \begin{bmatrix} 2/\sqrt{142} & (49/71)/\sqrt{42/71} \\ 5/\sqrt{142} & (16/71)/\sqrt{42/71} \\ 7/\sqrt{142} & (-6/71)/\sqrt{42/71} \\ 8/\sqrt{142} & (-17/71)/\sqrt{42/71} \end{bmatrix}$$

and

$$R = Q^T A = \begin{bmatrix} \sqrt{142} & (11/71)\sqrt{142} \\ 0 & \sqrt{42/71} \end{bmatrix}$$

and

$$QQ^T = \begin{bmatrix} 5/6 & 1/3 & 0 & -1/6 \\ 1/3 & 11/42 & 3/14 & 4/21 \\ 0 & 3/14 & 5/14 & 3/7 \\ -1/6 & 4/21 & 3/7 & 23/42 \end{bmatrix}.$$

What is $\hat{\mathbf{b}}$? $\hat{\mathbf{b}} = QQ^T\mathbf{b}$. We then find $\hat{\mathbf{x}} = [5/14\ 2/7]^T$ as before.

§6.5 Theorem 13

In the previous example, most of the work was in computing the projection $\hat{\mathbf{b}}$.

§6.5 Theorem 13

In the previous example, most of the work was in computing the projection $\hat{\mathbf{b}}$. It turns out that we don't need to do this to obtain a least-squares solution $\hat{\mathbf{x}}$.

§6.5 Theorem 13

In the previous example, most of the work was in computing the projection $\hat{\mathbf{b}}$. It turns out that we don't need to do this to obtain a least-squares solution $\hat{\mathbf{x}}$.

Theorem

In the previous example, most of the work was in computing the projection $\hat{\mathbf{b}}$. It turns out that we don't need to do this to obtain a least-squares solution $\hat{\mathbf{x}}$.

Theorem

The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ is precisely the solutions of $\boxed{A^T A\mathbf{x} = A^T\mathbf{b}}$.

In the previous example, most of the work was in computing the projection $\hat{\mathbf{b}}$. It turns out that we don't need to do this to obtain a least-squares solution $\hat{\mathbf{x}}$.

Theorem

The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ is precisely the solutions of $\boxed{A^T A\mathbf{x} = A^T\mathbf{b}}$. *The linear system represented by the boxed equation represents a system of linear equations called the* ***normal equations*** *for $A\mathbf{x} = \mathbf{b}$.*

§6.5 Proof of Theorem 13

Proof.

Proof.

Let $A = [\mathbf{a}_1 \ \ldots \ \mathbf{a}_n]$. Let $W = \operatorname{Col} A$.

Proof.

Let $A = [\mathbf{a}_1 \ \ldots \ \mathbf{a}_n]$. Let $W = \mathrm{Col}\, A$.

$(\subseteq)$

Proof.

Let $A = [\mathbf{a}_1 \ \ldots \ \mathbf{a}_n]$. Let $W = \mathrm{Col}\, A$.

$(\subseteq)$
Suppose $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$.

Proof.

Let $A = [\mathbf{a}_1 \ \dots \ \mathbf{a}_n]$. Let $W = \operatorname{Col} A$.

$(\subseteq)$
Suppose $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. Then $\mathbf{b} - \hat{\mathbf{b}} = \mathbf{b} - A\hat{\mathbf{x}}$ is in $W^\perp$.

§6.5 Proof of Theorem 13

Proof.

Let $A = [\mathbf{a}_1 \ \ldots \ \mathbf{a}_n]$. Let $W = \mathrm{Col}\, A$.

($\subseteq$)
Suppose $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. Then $\mathbf{b} - \hat{\mathbf{b}} = \mathbf{b} - A\hat{\mathbf{x}}$ is in $W^{\perp}$. This means that $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to every column of A

Proof.

Let $A = [\mathbf{a}_1 \ \ldots \ \mathbf{a}_n]$. Let $W = \mathrm{Col}\, A$.

($\subseteq$)
Suppose $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. Then $\mathbf{b} - \hat{\mathbf{b}} = \mathbf{b} - A\hat{\mathbf{x}}$ is in $W^{\perp}$. This means that $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to every column of A (every row of A^T).

Proof.

Let $A = [\mathbf{a}_1 \ \ldots \ \mathbf{a}_n]$. Let $W = \operatorname{Col} A$.

($\subseteq$)
Suppose $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. Then $\mathbf{b} - \hat{\mathbf{b}} = \mathbf{b} - A\hat{\mathbf{x}}$ is in $W^{\perp}$. This means that $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to every column of A (every row of A^T). But this means that $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$

Proof.

Let $A = [\mathbf{a}_1 \ \dots \ \mathbf{a}_n]$. Let $W = \mathrm{Col}\, A$.

($\subseteq$)
Suppose $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. Then $\mathbf{b} - \hat{\mathbf{b}} = \mathbf{b} - A\hat{\mathbf{x}}$ is in $W^{\perp}$. This means that $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to every column of A (every row of A^T). But this means that $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$ so that $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$.

Proof.

Let $A = [\mathbf{a}_1 \ \ldots \ \mathbf{a}_n]$. Let $W = \operatorname{Col} A$.

($\subseteq$)
Suppose $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. Then $\mathbf{b} - \hat{\mathbf{b}} = \mathbf{b} - A\hat{\mathbf{x}}$ is in $W^{\perp}$. This means that $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to every column of A (every row of A^T). But this means that $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$ so that $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$.

($\supseteq$)

Proof.

Let $A = [\mathbf{a}_1 \ \ldots \ \mathbf{a}_n]$. Let $W = \mathrm{Col}\, A$.

($\subseteq$)
Suppose $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. Then $\mathbf{b} - \hat{\mathbf{b}} = \mathbf{b} - A\hat{\mathbf{x}}$ is in $W^{\perp}$. This means that $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to every column of A (every row of A^T). But this means that $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$ so that $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$.

($\supseteq$)
Conversely, suppose $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$.

Proof.

Let $A = [\mathbf{a}_1 \ \dots \ \mathbf{a}_n]$. Let $W = \mathrm{Col}\, A$.

($\subseteq$)
Suppose $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. Then $\mathbf{b} - \hat{\mathbf{b}} = \mathbf{b} - A\hat{\mathbf{x}}$ is in $W^{\perp}$. This means that $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to every column of A (every row of A^T). But this means that $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$ so that $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$.

($\supseteq$)
Conversely, suppose $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$. Then $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$ so that $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to the rows of A^T

§6.5 Proof of Theorem 13

Proof.

Let $A = [\mathbf{a}_1 \ \ldots \ \mathbf{a}_n]$. Let $W = \mathrm{Col}\, A$.

($\subseteq$)
Suppose $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. Then $\mathbf{b} - \hat{\mathbf{b}} = \mathbf{b} - A\hat{\mathbf{x}}$ is in $W^\perp$. This means that $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to every column of A (every row of A^T). But this means that $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$ so that $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$.

($\supseteq$)
Conversely, suppose $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$. Then $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$ so that $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to the rows of A^T (columns of A).

Proof.

Let $A = [\mathbf{a}_1 \ \ldots \ \mathbf{a}_n]$. Let $W = \mathrm{Col}\, A$.

($\subseteq$)
Suppose $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. Then $\mathbf{b} - \hat{\mathbf{b}} = \mathbf{b} - A\hat{\mathbf{x}}$ is in $W^{\perp}$. This means that $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to every column of A (every row of A^T). But this means that $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$ so that $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$.

($\supseteq$)
Conversely, suppose $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$. Then $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$ so that $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to the rows of A^T (columns of A). Thus $\mathbf{b} - A\hat{\mathbf{x}} \in W^{\perp}$.

Proof.

Let $A = [\mathbf{a}_1 \ \dots \ \mathbf{a}_n]$. Let $W = \mathrm{Col}\, A$.

($\subseteq$)
Suppose $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. Then $\mathbf{b} - \hat{\mathbf{b}} = \mathbf{b} - A\hat{\mathbf{x}}$ is in $W^\perp$. This means that $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to every column of A (every row of A^T). But this means that $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$ so that $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$.

($\supseteq$)
Conversely, suppose $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$. Then $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$ so that $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to the rows of A^T (columns of A). Thus $\mathbf{b} - A\hat{\mathbf{x}} \in W^\perp$. Moreover, we have

$$\mathbf{b} = \underbrace{A\hat{\mathbf{x}}}_{\in W} + \underbrace{\mathbf{b} - A\hat{\mathbf{x}}}_{\in W^\perp}$$

Proof.

Let $A = [\mathbf{a}_1 \ \ldots \ \mathbf{a}_n]$. Let $W = \mathrm{Col}\, A$.

($\subseteq$)
Suppose $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. Then $\mathbf{b} - \hat{\mathbf{b}} = \mathbf{b} - A\hat{\mathbf{x}}$ is in $W^\perp$. This means that $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to every column of A (every row of A^T). But this means that $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$ so that $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$.

($\supseteq$)
Conversely, suppose $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$. Then $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$ so that $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to the rows of A^T (columns of A). Thus $\mathbf{b} - A\hat{\mathbf{x}} \in W^\perp$. Moreover, we have

$$\mathbf{b} = \underbrace{A\hat{\mathbf{x}}}_{\in W} + \underbrace{\mathbf{b} - A\hat{\mathbf{x}}}_{\in W^\perp}$$

By the uniqueness of orthogonal decompositions, $A\hat{\mathbf{x}}$ must be the projection of $\mathbf{b}$ onto W.

Proof.

Let $A = [\mathbf{a}_1 \ \ldots \ \mathbf{a}_n]$. Let $W = \mathrm{Col}\, A$.

($\subseteq$)
Suppose $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. Then $\mathbf{b} - \hat{\mathbf{b}} = \mathbf{b} - A\hat{\mathbf{x}}$ is in $W^\perp$. This means that $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to every column of A (every row of A^T). But this means that $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$ so that $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$.

($\supseteq$)
Conversely, suppose $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$. Then $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$ so that $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to the rows of A^T (columns of A). Thus $\mathbf{b} - A\hat{\mathbf{x}} \in W^\perp$. Moreover, we have

$$\mathbf{b} = \underbrace{A\hat{\mathbf{x}}}_{\in W} + \underbrace{\mathbf{b} - A\hat{\mathbf{x}}}_{\in W^\perp}$$

By the uniqueness of orthogonal decompositions, $A\hat{\mathbf{x}}$ must be the projection of $\mathbf{b}$ onto W. That is, $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. □

§6.5 Example ☕ continued

§6.5 Example ☕ continued

Let's revisit our example using the previous theorem.

§6.5 Example ☕ continued

Let's revisit our example using the previous theorem.
We are given A and $\mathbf{b}$.

§6.5 Example ☕ continued

Let's revisit our example using the previous theorem.
We are given A and $\mathbf{b}$. We compute

Let's revisit our example using the previous theorem.
We are given A and $\mathbf{b}$. We compute

$$A^T A = \begin{bmatrix} 2 & 5 & 7 & 8 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 8 & 1 \end{bmatrix} = \begin{bmatrix} 142 & 22 \\ 22 & 4 \end{bmatrix}$$

§6.5 Example ☕ continued

Let's revisit our example using the previous theorem.
We are given A and $\mathbf{b}$. We compute

$$A^T A = \begin{bmatrix} 2 & 5 & 7 & 8 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 8 & 1 \end{bmatrix} = \begin{bmatrix} 142 & 22 \\ 22 & 4 \end{bmatrix}$$

$$A^T \mathbf{b} = \begin{bmatrix} 2 & 5 & 7 & 8 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 57 \\ 9 \end{bmatrix}$$

Let's revisit our example using the previous theorem.
We are given A and $\mathbf{b}$. We compute

$$A^T A = \begin{bmatrix} 2 & 5 & 7 & 8 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 8 & 1 \end{bmatrix} = \begin{bmatrix} 142 & 22 \\ 22 & 4 \end{bmatrix}$$

$$A^T \mathbf{b} = \begin{bmatrix} 2 & 5 & 7 & 8 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 57 \\ 9 \end{bmatrix}$$

How do we find $\hat{\mathbf{x}}$?

§6.5 Example ☕ continued

Let's revisit our example using the previous theorem.
We are given A and $\mathbf{b}$. We compute

$$A^T A = \begin{bmatrix} 2 & 5 & 7 & 8 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 8 & 1 \end{bmatrix} = \begin{bmatrix} 142 & 22 \\ 22 & 4 \end{bmatrix}$$

$$A^T \mathbf{b} = \begin{bmatrix} 2 & 5 & 7 & 8 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 57 \\ 9 \end{bmatrix}$$

How do we find $\hat{\mathbf{x}}$?

$$\begin{bmatrix} 142 & 22 & 57 \\ 22 & 4 & 9 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 5/14 \\ 0 & 1 & 2/7 \end{bmatrix}.$$

§6.5 Example

§6.5 Example

Are least-squares solutions always unique?

§6.5 Example

Are least-squares solutions always unique? No!

§6.5 Example

Are least-squares solutions always unique? No!

$$\text{Let } A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix}.$$

Are least-squares solutions always unique? No!

Let $A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix}$. Then

$$A^T A = \begin{bmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix}, \quad A^T \mathbf{b} = \begin{bmatrix} 4 \\ -4 \\ 2 \\ 6 \end{bmatrix},$$

Are least-squares solutions always unique? No!

Let $A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}$, $\quad \mathbf{b} = \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix}$. Then

$$A^T A = \begin{bmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix}, \quad A^T\mathbf{b} = \begin{bmatrix} 4 \\ -4 \\ 2 \\ 6 \end{bmatrix},$$

and

$$\begin{bmatrix} A^T A & A^T\mathbf{b} \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 1 & 3 \\ 0 & 1 & 0 & -1 & -5 \\ 0 & 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

§6.5 Theorem 14

§6.5 Theorem 14

Given the contrast between example ☕ and example △,
one might like to know when a least-squares solution is unique...

Given the contrast between example ☕ and example ⟁,
one might like to know when a least-squares solution is unique...

Theorem

Given the contrast between example ☕ and example △,
one might like to know when a least-squares solution is unique...

Theorem

Let A be an $m \times n$ matrix.

§6.5 Theorem 14

Given the contrast between example ☕ and example ◬,
one might like to know when a least-squares solution is unique...

Theorem

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

§6.5 Theorem 14

Given the contrast between example ☕ and example △,
one might like to know when a least-squares solution is unique...

Theorem

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

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

§6.5 Theorem 14

Given the contrast between example ☕ and example △,
one might like to know when a least-squares solution is unique...

Theorem

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

1. *The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution for every $\mathbf{b} \in \mathbb{R}^m$*
2. *The columns of A are linearly independent*

§6.5 Theorem 14

Given the contrast between example ☕ and example ◬,
one might like to know when a least-squares solution is unique...

Theorem

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

1. *The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution for every $\mathbf{b} \in \mathbb{R}^m$*
2. *The columns of A are linearly independent*
3. *$A^T A$ is invertible*

§6.5 Theorem 14

Given the contrast between example ☕ and example ⧊,
one might like to know when a least-squares solution is unique...

Theorem

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

1. *The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution for every $\mathbf{b} \in \mathbb{R}^m$*
2. *The columns of A are linearly independent*
3. *A^TA is invertible*

When these hold, we have $\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b}$.

§6.5 Theorem 14

Given the contrast between example ☕ and example △, one might like to know when a least-squares solution is unique...

Theorem

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

1. *The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution for every $\mathbf{b} \in \mathbb{R}^m$*
2. *The columns of A are linearly independent*
3. *$A^T A$ is invertible*

When these hold, we have $\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}$.

Due to time constraints we will omit the proof.

§6.5 Theorem 14

Given the contrast between example ☕ and example △,
one might like to know when a least-squares solution is unique...

Theorem

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

1. *The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution for every $\mathbf{b} \in \mathbb{R}^m$*
2. *The columns of A are linearly independent*
3. *$A^T A$ is invertible*

When these hold, we have $\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}$.

Due to time constraints we will omit the proof.

Question:

§6.5 Theorem 14

Given the contrast between example ☕ and example △,
one might like to know when a least-squares solution is unique...

Theorem

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

1. *The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution for every $\mathbf{b} \in \mathbb{R}^m$*
2. *The columns of A are linearly independent*
3. *$A^T A$ is invertible*

When these hold, we have $\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}$.

Due to time constraints we will omit the proof.

Question: What does this theorem tell us about $A^T A$ and the columns of A from example △.

§6.5 Theorem 14

Given the contrast between example ☕ and example △,
one might like to know when a least-squares solution is unique...

Theorem

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

1. *The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution for every $\mathbf{b} \in \mathbb{R}^m$*
2. *The columns of A are linearly independent*
3. *$A^T A$ is invertible*

When these hold, we have $\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}$.

Due to time constraints we will omit the proof.

Question: What does this theorem tell us about $A^T A$ and the columns of A from example △.

$A^T A$ is not invertible, and the columns of A are linearly dependent.

§6.5 Example ☕ continued

We found previously, that this example has a unique least-squares solution.

§6.5 Example ☕ continued

We found previously, that this example has a unique least-squares solution. By the previous theorem we can compute

We found previously, that this example has a unique least-squares solution. By the previous theorem we can compute

$$
\begin{aligned}
\hat{\mathbf{x}} &= (A^T A)^{-1} A^T \mathbf{b} \\
&= \begin{bmatrix} 1/21 & -11/42 \\ -11/42 & 71/42 \end{bmatrix} A^T \mathbf{b} \\
&= \begin{bmatrix} -1/6 & -1/42 & 1/14 & 5/42 \\ 7/6 & 8/21 & -1/7 & -17/42 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} \\
&= \begin{bmatrix} 5/14 \\ 2/7 \end{bmatrix}.
\end{aligned}
$$

§6.5 Theorem 15

Theorem

§6.5 Theorem 15

Theorem

Let A be an $m \times n$ matrix with linearly independent columns.

Theorem

Let A be an $m \times n$ matrix with linearly independent columns. Let $A = QR$ be a QR factorization of A.

§6.5 Theorem 15

Theorem

Let A be an $m \times n$ matrix with linearly independent columns. Let $A = QR$ be a QR factorization of A. Then for every $\mathbf{b} \in \mathbb{R}^m$,

§6.5 Theorem 15

Theorem

Let A be an $m \times n$ matrix with linearly independent columns. Let $A = QR$ be a QR factorization of A. Then for every $\mathbf{b} \in \mathbb{R}^m$, The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution

Theorem

Let A be an $m \times n$ matrix with linearly independent columns. Let $A = QR$ be a QR factorization of A. Then for every $\mathbf{b} \in \mathbb{R}^m$, The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution given by

$$\hat{\mathbf{x}} = R^{-1} Q^T \mathbf{b}.$$

Theorem

Let A be an $m \times n$ matrix with linearly independent columns. Let $A = QR$ be a QR factorization of A. Then for every $\mathbf{b} \in \mathbb{R}^m$, The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution given by

$$\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}.$$

Proof.

Theorem

Let A be an $m \times n$ matrix with linearly independent columns. Let $A = QR$ be a QR factorization of A. Then for every $\mathbf{b} \in \mathbb{R}^m$, The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution given by

$$\hat{\mathbf{x}} = R^{-1} Q^T \mathbf{b}.$$

Proof.

By the previous theorem, $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution,

Theorem

Let A be an $m \times n$ matrix with linearly independent columns. Let $A = QR$ be a QR factorization of A. Then for every $\mathbf{b} \in \mathbb{R}^m$, The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution given by

$$\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}.$$

Proof.

By the previous theorem, $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution, so we just need to check that the given $\hat{\mathbf{x}}$ works.

Theorem

Let A be an $m \times n$ matrix with linearly independent columns. Let $A = QR$ be a QR factorization of A. Then for every $\mathbf{b} \in \mathbb{R}^m$, The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution given by

$$\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}.$$

Proof.

By the previous theorem, $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution, so we just need to check that the given $\hat{\mathbf{x}}$ works. But

$$\begin{aligned} A\hat{\mathbf{x}} &= QR\hat{\mathbf{x}} \\ &= QRR^{-1}Q^T\mathbf{b} \\ &= QQ^T\mathbf{b} = \hat{\mathbf{b}}. \end{aligned}$$

□

§6.5 Example ☕ continued

§6.5 Example ☕ continued

Using the previously computed Q and R,

Using the previously computed Q and R, we verify the theorem in this example as follows.

Using the previously computed Q and R, we verify the theorem in this example as follows.

$$\begin{aligned}\hat{\mathbf{x}} &= R^{-1}Q^T\mathbf{b} \\ &= \begin{bmatrix} -1/6 & -1/42 & 1/14 & 5/42 \\ 7/6 & 8/21 & -1/7 & -17/42 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} \\ &= \begin{bmatrix} 5/14 \\ 2/7 \end{bmatrix}.\end{aligned}$$

Using the previously computed Q and R, we verify the theorem in this example as follows.

$$\begin{aligned}\hat{\mathbf{x}} &= R^{-1}Q^T\mathbf{b} \\ &= \begin{bmatrix} -1/6 & -1/42 & 1/14 & 5/42 \\ 7/6 & 8/21 & -1/7 & -17/42 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} \\ &= \begin{bmatrix} 5/14 \\ 2/7 \end{bmatrix}.\end{aligned}$$

Note that this also shows the unilluminating fact that $(A^TA)^{-1}A^T = R^{-1}Q^T$ when defined.

§6.5 Example ☕ concluded

§6.5 Example ☕ concluded

Consider the quantitative data $(2, 1), (5, 2), (7, 3), (8, 3)$ in the (t, y)-plane.

§6.5 Example ☕ concluded

Consider the quantitative data $(2,1),(5,2),(7,3),(8,3)$ in the (t,y)-plane. How do we find a line $y = ct + d$ that "best fits" this data?

Consider the quantitative data $(2, 1), (5, 2), (7, 3), (8, 3)$ in the (t, y)-plane. How do we find a line $y = ct + d$ that "best fits" this data? Well, to do this we need a notion of error.

§6.5 Example ☕ concluded

Consider the quantitative data $(2, 1), (5, 2), (7, 3), (8, 3)$ in the (t, y)-plane. How do we find a line $y = ct + d$ that "best fits" this data? Well, to do this we need a notion of error. One might measure this error using "squared residuals".

§6.5 Example ☕ concluded

Consider the quantitative data $(2, 1), (5, 2), (7, 3), (8, 3)$ in the (t, y)-plane. How do we find a line $y = ct + d$ that "best fits" this data? Well, to do this we need a notion of error. One might measure this error using "squared residuals". Draw a picture.

§6.5 Example ☕ concluded

Consider the quantitative data $(2,1),(5,2),(7,3),(8,3)$ in the (t,y)-plane. How do we find a line $y = ct + d$ that "best fits" this data? Well, to do this we need a notion of error. One might measure this error using "squared residuals". Draw a picture. In our case, this error is given by

$$\sum_{i=1}^{4}(y_i - ct_i - d)^2 = (1-2c-d)^2 + (2-5c-d)^2 + (3-7c-d)^2 + (3-8c-3)^2.$$

§6.5 Example ☕ concluded

Consider the quantitative data $(2,1),(5,2),(7,3),(8,3)$ in the (t,y)-plane. How do we find a line $y = ct + d$ that "best fits" this data? Well, to do this we need a notion of error. One might measure this error using "squared residuals". Draw a picture. In our case, this error is given by

$$\sum_{i=1}^{4}(y_i - ct_i - d)^2 = (1-2c-d)^2 + (2-5c-d)^2 + (3-7c-d)^2 + (3-8c-3)^2.$$

If we let $A = \begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 8 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} c \\ d \end{bmatrix},$

§6.5 Example ☕ concluded

Consider the quantitative data $(2,1),(5,2),(7,3),(8,3)$ in the (t,y)-plane. How do we find a line $y = ct + d$ that "best fits" this data? Well, to do this we need a notion of error. One might measure this error using "squared residuals". Draw a picture. In our case, this error is given by

$$\sum_{i=1}^{4}(y_i - ct_i - d)^2 = (1-2c-d)^2+(2-5c-d)^2+(3-7c-d)^2+(3-8c-3)^2.$$

If we let $A = \begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 8 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} c \\ d \end{bmatrix}$, then the above error for a given $\mathbf{x}$ is precisely $\|\mathbf{b} - A\mathbf{x}\|^2$.

§6.5 Example ☕ concluded

Consider the quantitative data $(2,1),(5,2),(7,3),(8,3)$ in the (t,y)-plane. How do we find a line $y = ct + d$ that "best fits" this data? Well, to do this we need a notion of error. One might measure this error using "squared residuals". Draw a picture. In our case, this error is given by

$$\sum_{i=1}^{4}(y_i - ct_i - d)^2 = (1-2c-d)^2 + (2-5c-d)^2 + (3-7c-d)^2 + (3-8c-3)^2.$$

If we let $A = \begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 8 & 1 \end{bmatrix}$, $\quad \mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}$, $\quad \mathbf{x} = \begin{bmatrix} c \\ d \end{bmatrix}$, then the above error for a given $\mathbf{x}$ is precisely $\|\mathbf{b} - A\mathbf{x}\|^2$. This quantity is minimized precisely when

§6.5 Example ☕ concluded

Consider the quantitative data $(2, 1), (5, 2), (7, 3), (8, 3)$ in the (t, y)-plane. How do we find a line $y = ct + d$ that "best fits" this data? Well, to do this we need a notion of error. One might measure this error using "squared residuals". Draw a picture. In our case, this error is given by

$$\sum_{i=1}^{4}(y_i - ct_i - d)^2 = (1-2c-d)^2 + (2-5c-d)^2 + (3-7c-d)^2 + (3-8c-3)^2.$$

If we let $A = \begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 8 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} c \\ d \end{bmatrix}$, then the above error for a given $\mathbf{x}$ is precisely $\|\mathbf{b} - A\mathbf{x}\|^2$. This quantity is minimized precisely when you guessed it

§6.5 Example ☕ concluded

Consider the quantitative data $(2,1),(5,2),(7,3),(8,3)$ in the (t,y)-plane. How do we find a line $y = ct + d$ that "best fits" this data? Well, to do this we need a notion of error. One might measure this error using "squared residuals". Draw a picture. In our case, this error is given by

$$\sum_{i=1}^{4}(y_i - ct_i - d)^2 = (1-2c-d)^2 + (2-5c-d)^2 + (3-7c-d)^2 + (3-8c-3)^2.$$

If we let $A = \begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 8 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} c \\ d \end{bmatrix}$, then the above error for a given $\mathbf{x}$ is precisely $\|\mathbf{b} - A\mathbf{x}\|^2$. This quantity is minimized precisely when you guessed it $\mathbf{x}$ is a least-squares solution to $A\mathbf{x} = \mathbf{b}$.

§6.5 Example ☕ concluded

Consider the quantitative data $(2,1),(5,2),(7,3),(8,3)$ in the (t,y)-plane. How do we find a line $y = ct + d$ that "best fits" this data? Well, to do this we need a notion of error. One might measure this error using "squared residuals". Draw a picture. In our case, this error is given by

$$\sum_{i=1}^{4}(y_i - ct_i - d)^2 = (1-2c-d)^2 + (2-5c-d)^2 + (3-7c-d)^2 + (3-8c-3)^2.$$

If we let $A = \begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \\ 8 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} c \\ d \end{bmatrix}$, then the above error for a given $\mathbf{x}$ is precisely $\|\mathbf{b} - A\mathbf{x}\|^2$. This quantity is minimized precisely when you guessed it $\mathbf{x}$ is a least-squares solution to $A\mathbf{x} = \mathbf{b}$. $\hat{\mathbf{x}} = [5/14\ 2/7]^T$.

§6.5 Example ☕ concluded

§6.5 Example ☕ concluded

From the least-squares solution we obtain

§6.5 Example ☕ concluded

From the least-squares solution we obtain the line $y = (5/14)t + (2/7)$

§6.5 Example ☕ concluded

From the least-squares solution we obtain
the line $y = (5/14)t + (2/7)$
that best fits

From the least-squares solution we obtain
the line $y = (5/14)t + (2/7)$
that best fits
the data (2,1),(5,2),(7,3),(8,3) :

From the least-squares solution we obtain
the line $y = (5/14)t + (2/7)$
that best fits
the data (2,1),(5,2),(7,3),(8,3) :