Solutions to Maths/LA/Tutorial 7 Least Squares

(Rev 24 July 2021)

Q1 Lay5e/Ch6.5/pg364/Ex1

EXAMPLE 1 Find a least-squares solution of the inconsistent system $A\mathbf{x} = \mathbf{b}$ for

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$$

Sol:

SOLUTION To use normal equations (3), compute:

$$A^TA = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}$$

$$A^T\mathbf{b} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

Then the equation $A^TA\mathbf{x} = A^T\mathbf{b}$ becomes

$$\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

Row operations can be used to solve this system, but since A^TA is invertible and 2×2, it is probably faster to compute

$$(A^TA)^{-1} = \frac{1}{84} \begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix}$$

and then to solve $A^TA\mathbf{x} = A^T\mathbf{b}$ as

$$\begin{aligned} \hat{\mathbf{x}} &= (A^TA)^{-1}A^T\mathbf{b} \\ &= \frac{1}{84} \begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix} \begin{bmatrix} 19 \\ 11 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 84 \\ 168 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \end{aligned}$$

In many calculations, A^TA is invertible, but this is not always the case. The next

Q2) Lay5e/Ch6.5/pg366/Ex4

EXAMPLE 4 Find a least-squares solution of $A\mathbf{x} = \mathbf{b}$ for

$$A = \begin{bmatrix} 1 & -6 \\ 1 & -2 \\ 1 & 1 \\ 1 & 7 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -1 \\ 2 \\ 1 \\ 6 \end{bmatrix}$$

SOLUTION Because the columns $\mathbf{a}_1$ and $\mathbf{a}_2$ of A are orthogonal, the orthogonal projection of $\mathbf{b}$ onto Col A is given by

$$\hat{\mathbf{b}} = \frac{\mathbf{b}\cdot\mathbf{a}_1}{\mathbf{a}_1\cdot\mathbf{a}_1}\mathbf{a}_1 + \frac{\mathbf{b}\cdot\mathbf{a}_2}{\mathbf{a}_2\cdot\mathbf{a}_2}\mathbf{a}_2 = \frac{8}{4}\mathbf{a}_1 + \frac{45}{90}\mathbf{a}_2 \tag{5}$$

$$= \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \end{bmatrix} + \begin{bmatrix} -3 \\ -1 \\ 1/2 \\ 7/2 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 5/2 \\ 11/2 \end{bmatrix}$$

Now that $\hat{\mathbf{b}}$ is known, we can solve $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. But this is trivial, since we already know what weights to place on the columns of A to produce $\hat{\mathbf{b}}$. It is clear from (5) that

$$\hat{\mathbf{x}} = \begin{bmatrix} 8/4 \\ 45/90 \end{bmatrix} = \begin{bmatrix} 2 \\ 1/2 \end{bmatrix}$$

■

Q3) Lay5e/ch6.5/pg 364/Ex2

EXAMPLE 2 Find a least-squares solution of $A\mathbf{x} = \mathbf{b}$ for

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

SOLUTION Compute

$$A^T A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix}$$

$$A^T \mathbf{b} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ -4 \\ 2 \\ 6 \end{bmatrix}$$

The augmented matrix for $A^T A\mathbf{x} = A^T\mathbf{b}$ is

$$\begin{bmatrix} 6 & 2 & 2 & 2 & 4 \\ 2 & 2 & 0 & 0 & -4 \\ 2 & 0 & 2 & 0 & 2 \\ 2 & 0 & 0 & 2 & 6 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 1 & 3 \\ 0 & 1 & 0 & -1 & -5 \\ 0 & 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The general solution is $x_1 = 3 - x_4$, $x_2 = -5 + x_4$, $x_3 = -2 + x_4$, and x_4 is free. So the general least-squares solution of $A\mathbf{x} = \mathbf{b}$ has the form

$$\hat{\mathbf{x}} = \begin{bmatrix} 3 \\ -5 \\ -2 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad \blacksquare$$

Q4) Lay4e/ch6.5/pg 368/Q14/

14. Let $A = \begin{bmatrix} 2 & 1 \\ -3 & -4 \\ 3 & 2 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 5 \\ 4 \\ 4 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 4 \\ -5 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} 6 \\ -5 \end{bmatrix}$. Compute $A\mathbf{u}$ and $A\mathbf{v}$, and compare them with $\mathbf{b}$. Is it possible that at least one of $\mathbf{u}$ or $\mathbf{v}$ could be a least-squares solution of $A\mathbf{x} = \mathbf{b}$? (Answer this without computing a least-squares solution.)

Ans Q4) Lay5e/ch6.5/pg 368/Ex14/:

14. One computes that $A\mathbf{u} = \begin{bmatrix} 3 \\ 8 \\ 2 \end{bmatrix}, \mathbf{b} - A\mathbf{u} = \begin{bmatrix} 2 \\ -4 \\ 2 \end{bmatrix}, \|\mathbf{b} - A\mathbf{u}\| = \sqrt{24}$;

$A\mathbf{v} = \begin{bmatrix} 7 \\ 2 \\ 8 \end{bmatrix}, \mathbf{b} - A\mathbf{v} = \begin{bmatrix} -2 \\ 2 \\ -4 \end{bmatrix}, \|\mathbf{b} - A\mathbf{v}\| = \sqrt{24}$.

Since $A\mathbf{u}$ and $A\mathbf{v}$ are equally close to $\mathbf{b}$, and the orthogonal projection is the *unique* closest point in Col A to $\mathbf{b}$, neither $A\mathbf{u}$ nor $A\mathbf{v}$ can be the closest point in Col A to $\mathbf{b}$. Thus neither $\mathbf{u}$ nor $\mathbf{v}$ can be a least-squares solution of $A\mathbf{x} = \mathbf{b}$.

Q5) Lay5e/ch6.5/pg368/Ex17+18

In Exercises 17 and 18, A is an $m \times n$ matrix and $\mathbf{b}$ is in $\mathbb{R}^m$. Mark each statement True or False. Justify each answer.

17. a. The general least-squares problem is to find an $\mathbf{x}$ that makes $A\mathbf{x}$ as close as possible to $\mathbf{b}$.

b. A least-squares solution of $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}}$ that satisfies $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$, where $\hat{\mathbf{b}}$ is the orthogonal projection of $\mathbf{b}$ onto Col A.

c. A least-squares solution of $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}}$ such that $\|\mathbf{b} - A\mathbf{x}\| \leq \|\mathbf{b} - A\hat{\mathbf{x}}\|$ for all $\mathbf{x}$ in $\mathbb{R}^n$.

d. Any solution of $A^TA\mathbf{x} = A^T\mathbf{b}$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$.

e. If the columns of A are linearly independent, then the equation $A\mathbf{x} = \mathbf{b}$ has exactly one least-squares solution.

18. a. If $\mathbf{b}$ is in the column space of A, then every solution of $A\mathbf{x} = \mathbf{b}$ is a least-squares solution.

b. The least-squares solution of $A\mathbf{x} = \mathbf{b}$ is the point in the column space of A closest to $\mathbf{b}$.

c. A least-squares solution of $A\mathbf{x} = \mathbf{b}$ is a list of weights that, when applied to the columns of A, produces the orthogonal projection of $\mathbf{b}$ onto Col A.

d. If $\hat{\mathbf{x}}$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$, then $\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b}$.

e. The normal equations always provide a reliable method for computing least-squares solutions.

f. If A has a QR factorization, say $A = QR$, then the best way to find the least-squares solution of $A\mathbf{x} = \mathbf{b}$ is to compute $\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$.

Ans Q5 (Lay5e/ch6.5/pg368/Ex17+18)

17. **a**. True. See the beginning of the section. The distance from $A\mathbf{x}$ to $\mathbf{b}$ is $\| A\mathbf{x} - \mathbf{b} \|$.

b. True. See the comments about equation (1).

c. False. The inequality points in the wrong direction. See the definition of a least-squares solution.

d. True. See Theorem 13.

e. True. See Theorem 14.

18. **a**. True. See the paragraph following the definition of a least-squares solution.

b. False. If $\hat{\mathbf{x}}$ is the least-squares solution, then $A\hat{\mathbf{x}}$ is the point in the column space of A closest to $\mathbf{b}$. See Figure 1 and the paragraph preceding it.

c. True. See the discussion following equation (1).

d. False. The formula applies only when the columns of A are linearly independent. See Theorem 14.

e. False. See the comments after Example 4.

f. False. See the Numerical Note.

Q5) Lay5e/ch6.5/pg368/Ex17a

17. a. **The general least-squares problem is to find an x that makes $A\mathbf{x}$ as close as possible to b.**

Detailed Ans Q5) Lay5e/ch6.5/pg368/Ex17a, ans in Pg Lay5e/ch6.5/pg362

a. True. See the beginning of the section. The distance from $A\mathbf{x}$ to $\mathbf{b}$ is $\| A\mathbf{x} - \mathbf{b} \|$.

Think of $A\mathbf{x}$ as an *approximation* to $\mathbf{b}$. The smaller the distance between $\mathbf{b}$ and $A\mathbf{x}$, given by $\|\mathbf{b} - A\mathbf{x}\|$, the better the approximation. The **general least-squares problem** is to find an $\mathbf{x}$ that makes $\|\mathbf{b} - A\mathbf{x}\|$ as small as possible. The adjective "least-squares" arises from the fact that $\|\mathbf{b} - A\mathbf{x}\|$ is the square root of a sum of squares.

DEFINITION If A is $m \times n$ and $\mathbf{b}$ is in $\mathbb{R}^m$, a **least-squares solution** of $A\mathbf{x} = \mathbf{b}$ is an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ such that

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

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

The most important aspect of the least-squares problem is that no matter what $\mathbf{x}$ we select, the vector $A\mathbf{x}$ will necessarily be in the column space, Col A. So we seek an $\mathbf{x}$ that makes $A\mathbf{x}$ the closest point in Col A to $\mathbf{b}$. See Figure 1. (Of course, if $\mathbf{b}$ happens to be in Col A, then $\mathbf{b}$ *is* $A\mathbf{x}$ for some $\mathbf{x}$, and such an $\mathbf{x}$ is a "least-squares solution.")

Q5) Lay5e/ch6.5/pg368/Ex17b

b. A least-squares solution of $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}}$ that satisfies $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$, where $\hat{\mathbf{b}}$ is the orthogonal projection of $\mathbf{b}$ onto Col A.

Detailed Ans Q5) Lay5e/ch6.5/pg368/Ex17b, ans in Pg Lay5e/ch6.5/pg363

b. True. See the comments about equation (1).

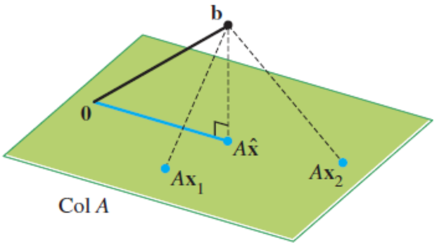


FIGURE 1 The vector $\mathbf{b}$ is closer to $A\hat{\mathbf{x}}$ than to $A\mathbf{x}$ for other $\mathbf{x}$.

Because $\hat{\mathbf{b}}$ is in the column space of A, the equation $A\mathbf{x} = \hat{\mathbf{b}}$ *is* consistent, and there is an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ such that

$$A\hat{\mathbf{x}} = \hat{\mathbf{b}} \tag{1}$$

Since $\hat{\mathbf{b}}$ is the closest point in Col A to $\mathbf{b}$, a vector $\hat{\mathbf{x}}$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$ if and only if $\hat{\mathbf{x}}$ satisfies (1). Such an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ is a list of weights that will build $\hat{\mathbf{b}}$ out of the columns of A. See Figure 2. [There are many solutions of (1) if the equation has free variables.]

Q5) Lay5e/ch6.5/pg368/Ex17d

d. Any solution of $A^TA\mathbf{x} = A^T\mathbf{b}$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$.

Detailed Ans Q5 (Ex17d)

d. True. See Theorem 13.

THEOREM 13 The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the nonempty set of solutions of the normal equations $A^TA\mathbf{x} = A^T\mathbf{b}$.

PROOF As shown above, the set of least-squares solutions is nonempty and each least-squares solution $\hat{\mathbf{x}}$ satisfies the normal equations. Conversely, suppose $\hat{\mathbf{x}}$ satisfies $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$. Then $\hat{\mathbf{x}}$ satisfies (2) above, which shows that $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to the rows of A^T and hence is orthogonal to the columns of A. Since the columns of A span Col A, the vector $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to all of Col A. Hence the equation

$$\mathbf{b} = A\hat{\mathbf{x}} + (\mathbf{b} - A\hat{\mathbf{x}})$$

is a decomposition of $\mathbf{b}$ into the sum of a vector in Col A and a vector orthogonal to Col A. By the uniqueness of the orthogonal decomposition, $A\hat{\mathbf{x}}$ must be the orthogonal projection of $\mathbf{b}$ onto Col A. That is, $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$, and $\hat{\mathbf{x}}$ is a least-squares solution. ■

Q5) Lay5e/ch6.5/pg368/Ex17e

e. If the columns of A are linearly independent, then the equation $A\mathbf{x} = \mathbf{b}$ has exactly one least-squares solution.

Detailed Ans Q5 (Ex17e)

e. True. See Theorem 14.

THEOREM 14

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^TA is invertible.

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

$$\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b} \tag{4}$$

Note: theorem 14 proof in Q6

Q5) Lay5e/ch6.5/pg368/Ex18a

18. a. If $\mathbf{b}$ is in the column space of A, then every solution of $A\mathbf{x} = \mathbf{b}$ is a least-squares solution.

Detailed Ans Q5) Lay5e/ch6.5/pg368/Ex18a

18. **a**. True. See the paragraph following the definition of a least-squares solution.

Think of $A\mathbf{x}$ as an *approximation* to $\mathbf{b}$. The smaller the distance between $\mathbf{b}$ and $A\mathbf{x}$, given by $\|\mathbf{b} - A\mathbf{x}\|$, the better the approximation. The **general least-squares problem** is to find an $\mathbf{x}$ that makes $\|\mathbf{b} - A\mathbf{x}\|$ as small as possible. The adjective "least-squares" arises from the fact that $\|\mathbf{b} - A\mathbf{x}\|$ is the square root of a sum of squares.

DEFINITION

If A is $m \times n$ and $\mathbf{b}$ is in $\mathbb{R}^m$, a **least-squares solution** of $A\mathbf{x} = \mathbf{b}$ is an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ such that

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

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

The most important aspect of the least-squares problem is that no matter what $\mathbf{x}$ we select, the vector $A\mathbf{x}$ will necessarily be in the column space, Col A. So we seek an $\mathbf{x}$ that makes $A\mathbf{x}$ the closest point in Col A to $\mathbf{b}$. See Figure 1. (Of course, if $\mathbf{b}$ happens to be in Col A, then $\mathbf{b}$ *is* $A\mathbf{x}$ for some $\mathbf{x}$, and such an $\mathbf{x}$ is a "least-squares solution.")

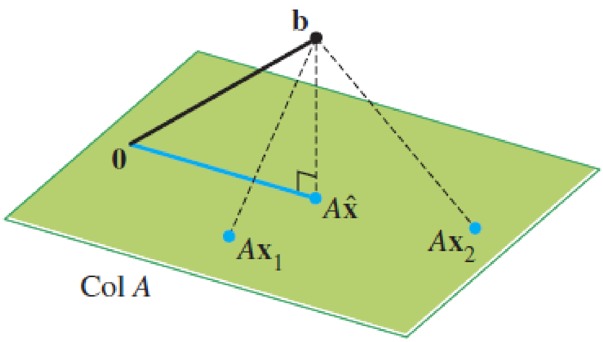


FIGURE 1 The vector $\mathbf{b}$ is closer to $A\hat{\mathbf{x}}$ than to $A\mathbf{x}$ for other $\mathbf{x}$.

Q5) Lay5e/ch6.5/pg368/Ex18c

c. A least-squares solution of $A\mathbf{x} = \mathbf{b}$ is a list of weights that, when applied to the columns of A, produces the orthogonal projection of $\mathbf{b}$ onto Col A.

Detailed Ans Q5) Lay5e/ch6.5/pg368/Ex18c in Lay5e/ch6.5/pg 363

c. True. See the discussion following equation (1).

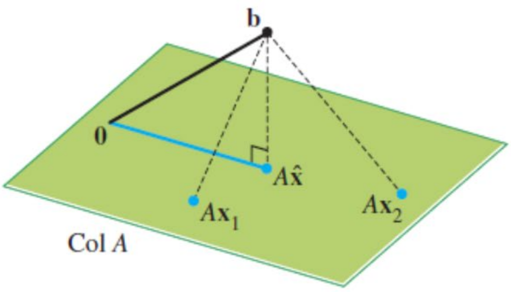


FIGURE 1 The vector $\mathbf{b}$ is closer to $A\hat{\mathbf{x}}$ than to $A\mathbf{x}$ for other $\mathbf{x}$.

Because $\hat{\mathbf{b}}$ is in the column space of A, the equation $A\mathbf{x} = \hat{\mathbf{b}}$ *is* consistent, and there is an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ such that

$$A\hat{\mathbf{x}} = \hat{\mathbf{b}} \tag{1}$$

Since $\hat{\mathbf{b}}$ is the closest point in Col A to $\mathbf{b}$, a vector $\hat{\mathbf{x}}$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$ if and only if $\hat{\mathbf{x}}$ satisfies (1). Such an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ is a list of weights that will build $\hat{\mathbf{b}}$ out of the columns of A. See Figure 2. [There are many solutions of (1) if the equation has free variables.]

Q5) Lay5e/ch6.5/pg368/Ex18d

d. If $\hat{\mathbf{x}}$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$, then $\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b}$.

Ans Q5, (Ex18d)

d. False. The formula applies only when the columns of A are linearly independent. See Theorem 14.

Note: theorem 14 proof in Q6

THEOREM 14

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^TA is invertible.

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

$$\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b} \tag{4}$$

Q5) Lay5e/ch6.5/pg368/Ex18e

e. The normal equations always provide a reliable method for computing least-squares solutions.

Detailed Ans Q5) Lay5e/ch6.5/pg368/Ex18e in (Lay5e/pg366 : Example 4)

e. False. See the comments after Example 4.

In some cases, the normal equations for a least-squares problem can be *ill-conditioned*; that is, small errors in the calculations of the entries of A^TA can sometimes cause relatively large errors in the solution $\hat{\mathbf{x}}$. If the columns of A are linearly independent, the least-squares solution can often be computed more reliably through a QR factorization of A (described in Section 6.4).[1]

Q5) Lay5e/ch6.5/pg368/Ex18f

f. If A has a QR factorization, say $A = QR$, then the best way to find the least-squares solution of $A\mathbf{x} = \mathbf{b}$ is to compute $\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$.

Detailed Ans Q5) Lay5e/ch6.5/pg368/Ex18f in Lay5e/ch6.5/pg367

f. False. See the Numerical Note.

THEOREM 15

Given an $m \times n$ matrix A with linearly independent columns, let $A = QR$ be a QR factorization of A as in Theorem 12. Then, for each $\mathbf{b}$ in $\mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution, given by

$$\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b} \tag{6}$$

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

$$A\hat{\mathbf{x}} = QR\hat{\mathbf{x}} = QRR^{-1}Q^T\mathbf{b} = QQ^T\mathbf{b}$$

By Theorem 12, the columns of Q form an orthonormal basis for Col A. Hence, by Theorem 10, $QQ^T\mathbf{b}$ is the orthogonal projection $\hat{\mathbf{b}}$ of $\mathbf{b}$ onto Col A. Then $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$, which shows that $\hat{\mathbf{x}}$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$. The uniqueness of $\hat{\mathbf{x}}$ follows from Theorem 14. ■

NUMERICAL NOTE

Since R in Theorem 15 is upper triangular, $\hat{\mathbf{x}}$ should be calculated as the exact solution of the equation

$$R\mathbf{x} = Q^T\mathbf{b} \tag{7}$$

It is much faster to solve (7) by back-substitution or row operations than to compute R^{-1} and use (6).

Additional Example Video Ref: "Solving least squares via QR factorization": https://www.youtube.com/watch?v=OCLFZwi40nM

Q6) Lay5e/ch6.5/pg369/Ex19+20+21

Proof Theorem 14:

THEOREM 14

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 indpendent.

c. The matrix A^TA is invertible.

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

$$\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b} \tag{4}$$

The main elements of a proof of Theorem 14 are outlined in Exercises 19–21, which also review concepts from Chapter 4. Formula (4) for $\hat{\mathbf{x}}$ is useful mainly for theoretical purposes and for hand calculations when A^TA is a 2×2 invertible matrix.

19. Let A be an $m \times n$ matrix. Use the steps below to show that a vector $\mathbf{x}$ in $\mathbb{R}^n$ satisfies $A\mathbf{x} = \mathbf{0}$ if and only if $A^TA\mathbf{x} = \mathbf{0}$. This will show that Nul A = Nul A^TA.

a. Show that if $A\mathbf{x} = \mathbf{0}$, then $A^TA\mathbf{x} = \mathbf{0}$.

b. Suppose $A^TA\mathbf{x} = \mathbf{0}$. Explain why $\mathbf{x}^TA^TA\mathbf{x} = \mathbf{0}$, and use this to show that $A\mathbf{x} = \mathbf{0}$.

20. Let A be an $m \times n$ matrix such that A^TA is invertible. Show that the columns of A are linearly independent. [*Careful:* You may not assume that A is invertible; it may not even be square.]

21. Let A be an $m \times n$ matrix whose columns are linearly independent. [*Careful:* A need not be square.]

a. Use Exercise 19 to show that A^TA is an invertible matrix.

b. Explain why A must have at least as many rows as columns.

c. Determine the rank of A.

Detailed Ans Q6) Lay5e/ch6.5/pg369/Ex19,20,21

19. **a**. If $A\mathbf{x} = \mathbf{0}$, then $A^TA\mathbf{x} = A^T\mathbf{0} = \mathbf{0}$. This shows that Nul A is contained in Nul A^TA.

b. If $A^TA\mathbf{x} = \mathbf{0}$, then $\mathbf{x}^TA^TA\mathbf{x} = \mathbf{x}^T\mathbf{0} = 0$. So $(A\mathbf{x})^T(A\mathbf{x}) = 0$, which means that $\|A\mathbf{x}\|^2 = 0$, and hence $A\mathbf{x} = \mathbf{0}$. This shows that Nul A^TA is contained in Nul A.

20. Suppose that $A\mathbf{x} = \mathbf{0}$. Then $A^TA\mathbf{x} = A^T\mathbf{0} = \mathbf{0}$. Since A^TA is invertible, $\mathbf{x}$ must be $\mathbf{0}$. Hence the columns of A are linearly independent.

21. **a**. If A has linearly independent columns, then the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. By Exercise 19, the equation $A^TA\mathbf{x} = \mathbf{0}$ also has only the trivial solution. Since A^TA is a square matrix, it must be invertible by the Invertible Matrix Theorem.

b. Since the n linearly independent columns of A belong to $\mathbb{R}^m$, m could not be less than n.

c. The n linearly independent columns of A form a basis for Col A, so the rank of A is n.

Q7) Given the following table of measured height and weight of 7 individuals,

Dataset

Height (in)	Weight (lb)
48	93
50	105
53	102
55	118
61	121
64	135
73	180

a) Assume that the weight (w) and height (h) are related by the following equation:

$$w = \beta_0 + \beta_1 h$$

find the unknown variable β_0, β_1

b) Calculate and plot the residuals using the β found

Ans Q7A) Find the parameters β

```
% T7_Q81.m
close all; clear all;

h = [ 1 48; 1 50; 1 53; 1 55; 1 61; 1 64; 1 73];
w = [93 105 102 118 121 135 180]';
beta = inv(h'*h)*(h'*w)

input_height = h(:,2);
est_weight_input = beta(1)+beta(2)*input_height;
% only for test data height

all_height     = [0:75];
est_weight_all = beta(1)+beta(2)*all_height;
figure(1);
plot(input_height, w,'r+'); hold on;
title('least squares regession of height vs weight');
xlabel('height (in)');
ylabel('weight (pound)'); grid on;
plot(all_height, est_weight_all, 'g-'); hold on;
```

```
beta =

  -61.1088
    3.1727
```

slope

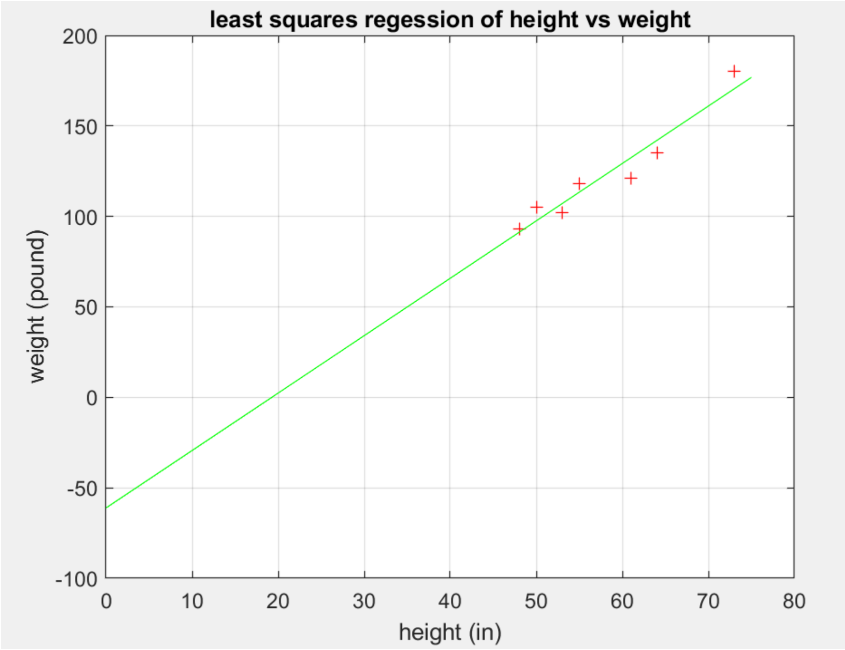
$$\hat{y} = a + bx \quad or \quad \hat{y} = \beta_0 + \beta_1 x$$

y-intercept

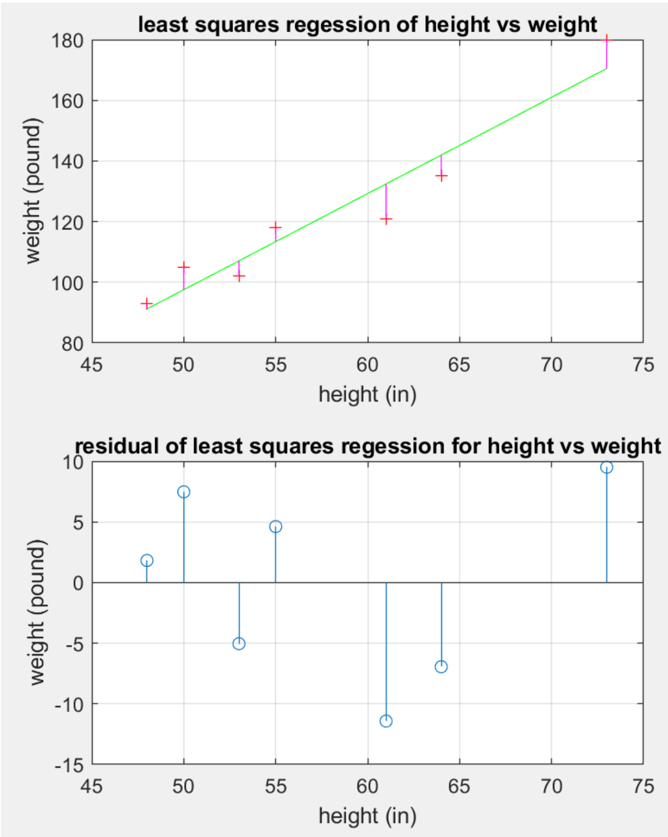
Least-Square Regression Analysis

$$\widehat{weight} = -61.11 + 3.17\,(\,height\,)$$

$$\hat{y} = -61.11 + 3.17x$$

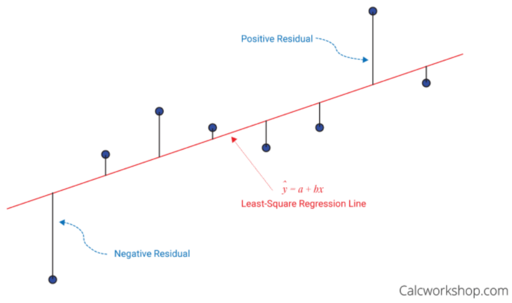


Ans Q7B) plot the residuals

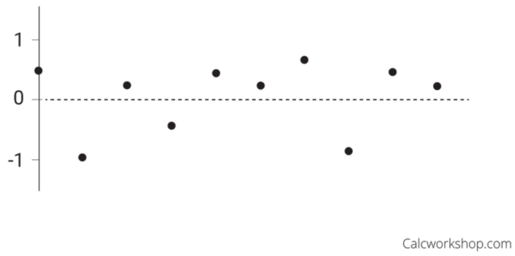


Residuals

Now the **residuals** are the *differences between the observed and predicted values*. It measures the distance from the regression line (predicted value) and the actual observed value. In other words, it helps us to measure error, or how well our regression line “fits” our data. Moreover, we can then visually display our findings and look for variations on a residual plot.



Residual Plot Interpretation



Residual Scatter Plot

Q8) Least squares fit of other curves

Given that y is generate by the following equation:

$$y_{clean} = \beta_0 + \beta_1 x + \beta_2 x^2$$

$$y_{noisy} = y_{clean} + n$$

for $x = -3..3$ and n is some random noise with mean 0,
we recorded the following value pairs:

$$x \quad = [-3, \quad -2, \quad -1, \quad 0, \quad 1, \quad 2, \quad 3]$$

$$y_{clean} = [4.3000 \quad 2.6000 \quad 1.5000 \quad 1.0000 \quad 1.1000 \quad 1.8000 \quad 3.1000];$$

$$n = [-0.0830 \quad 0.2203 \quad -0.4999 \quad -0.1977 \quad -0.3532 \quad -0.4077 \quad -0.3137];$$

$$y_{noisy} = [4.2170 \quad 2.8203 \quad 1.0001 \quad 0.8023 \quad 0.7468 \quad 1.3923 \quad 2.7863];$$

Answer the followings:

A) find the parameters β_{clean} using (x, y_{clean}) and $\beta_{noisy}\ using\ (x, y_{noisy})$;

B) why is the residual error using β_{clean} with the equation $y = \beta_0 + \beta_1 x + \beta_2\, x^2$ while it is not zero when using β_{noisy}?

Ans Q8

See code in Matlab (This is for example only, its not necessary for student to code!)

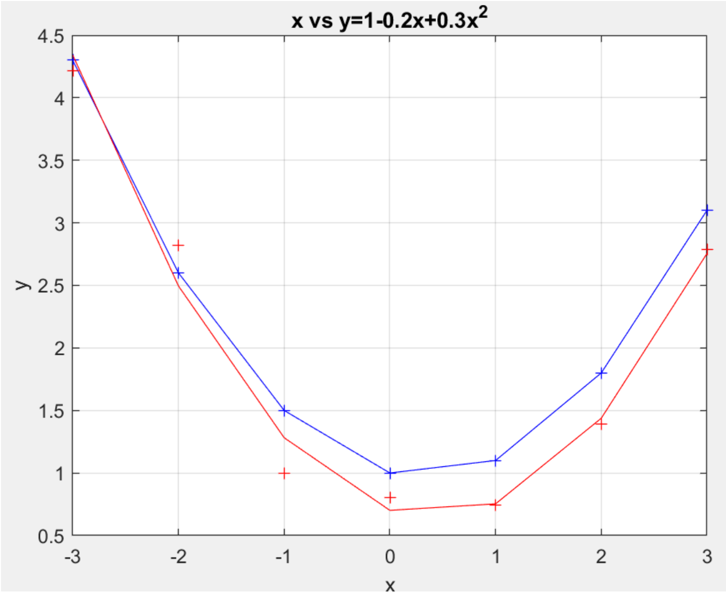
```
clear all
rng(1);
x          = [-3:1:3];
%y_clean   = 1-0.2*x+0.3*x.*x;
% created by this eqn
y_clean    = [4.3000    2.6000    1.5000    1.0000    1.1000    1.8000
3.1000];
noise = rand(length(x),1)'-0.5;
y_noisy = y_clean+noise;
plot(x,y_clean);
hold on;
plot(x,y_noisy,'r+');

X = zeros(length(x),3);
X(:,1) = ones(length(x),1)';
for (i=1:length(x))
   X(i,2) = x(i);
   X(i,3) = x(i)*x(i);
end
y_clean=y_clean(:);
beta = inv(X'*X)*(X'*y_clean);
y_noisy=y_noisy(:);
beta_noisy = inv(X'*X)*(X'*y_noisy);
y_noisy_eq = X*beta_noisy;
plot(x,y_noisy_eq,'r-');
```

Ans Q8A) Output of Matlab code

$$\beta_{clean} = [1.0000 \quad -0.2000 \quad 0.3000]$$

$$\beta_{noisy} = [0.7014 \quad -0.2643 \quad 0.3163]$$



$y_{clean} = \beta_0 + \beta_1 x + \beta_2 x^2$ and the least squares solution using the formulation $y = X\beta_{clean}$ gives the exact solution.

Using β_{noisy} to generate y_{noisy} has residual error due to the presence of n which is not accounted for.

Lay5e/pg373: explanations of theory for this question

Least-Squares Fitting of Other Curves

When data points $(x_1, y_1), \ldots, (x_n, y_n)$ on a scatter plot do not lie close to any line, it may be appropriate to postulate some other functional relationship between x and y.

The next two examples show how to fit data by curves that have the general form

$$y = \beta_0 f_0(x) + \beta_1 f_1(x) + \cdots + \beta_k f_k(x) \tag{2}$$

where $f_0, \ldots, f_k$ are known functions and $\beta_0, \ldots, \beta_k$ are parameters that must be determined. As we will see, equation (2) describes a linear model because it is linear in the unknown parameters.

For a particular value of x, (2) gives a predicted, or "fitted," value of y. The difference between the observed value and the predicted value is the residual. The parameters $\beta_0, \ldots, \beta_k$ must be determined so as to minimize the sum of the squares of the residuals.

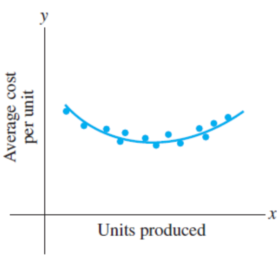


FIGURE 3
Average cost curve.

EXAMPLE 2 Suppose data points $(x_1, y_1), \ldots, (x_n, y_n)$ appear to lie along some sort of parabola instead of a straight line. For instance, if the x-coordinate denotes the production level for a company, and y denotes the average cost per unit of operating at a level of x units per day, then a typical average cost curve looks like a parabola that opens upward (Figure 3). In ecology, a parabolic curve that opens downward is used to model the net primary production of nutrients in a plant, as a function of the surface area of the foliage (Figure 4). Suppose we wish to approximate the data by an equation of the form

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 \tag{3}$$

Describe the linear model that produces a "least-squares fit" of the data by equation (3).

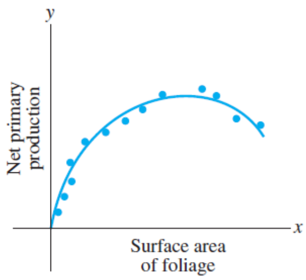


FIGURE 4
Production of nutrients.

Describe the linear model that produces a "least-squares fit" of the data by equation (3).

SOLUTION Equation (3) describes the ideal relationship. Suppose the actual values of the parameters are $\beta_0, \beta_1, \beta_2$. Then the coordinates of the first data point (x_1, y_1) satisfy an equation of the form

$$y_1 = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \epsilon_1$$

where ϵ_1 is the residual error between the observed value y_1 and the predicted y-value $\beta_0 + \beta_1 x_1 + \beta_2 x_1^2$. Each data point determines a similar equation:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \epsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_2 + \beta_2 x_2^2 + \epsilon_2 \\ &\vdots \qquad\qquad \vdots \\ y_n &= \beta_0 + \beta_1 x_n + \beta_2 x_n^2 + \epsilon_n \end{aligned}$$

It is a simple matter to write this system of equations in the form $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$. To find X, inspect the first few rows of the system and look for the pattern.

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}}_{\boldsymbol{\epsilon}}$$

■

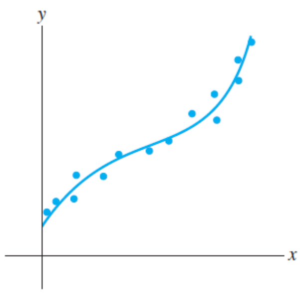


FIGURE 5
Data points along a cubic curve.

EXAMPLE 3 If data points tend to follow a pattern such as in Figure 5, then an appropriate model might be an equation of the form

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$$

Such data, for instance, could come from a company's total costs, as a function of the level of production. Describe the linear model that gives a least-squares fit of this type to data $(x_1, y_1), \dots, (x_n, y_n)$.

SOLUTION By an analysis similar to that in Example 2, we obtain

$$\underset{\text{Observation vector}}{\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}, \quad \underset{\text{Design matrix}}{X = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 \end{bmatrix}}, \quad \underset{\text{Parameter vector}}{\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}}, \quad \underset{\text{Residual vector}}{\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}}$$

■

============================== End ============================