6.5 & 6.6: Least-Squares Problems

Math 220: Linear Algebra

We will now look at the case where $A\mathbf{x}=\mathbf{b}$ has no solution. What would be "closest" possible solution $\mathbf{x}$? This is called the Least-Squares problem, and it mirrors our Best-Approximation Theorem from 6.3.

Definition

If A is $m \times n$ and **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

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

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

Here is the derivation for how to find the least squares solution $\hat{x}$.

$A\vec{x} = \vec{b}$ has no solution.

let $\hat{b} = \text{proj}_{\text{col}A} \vec{b}$.

$\Rightarrow A\vec{x} = \hat{b}$ is consistent since $\hat{b} \in \text{col } A$.

we call the solution $\hat{x}$. That is $A\hat{x} = \hat{b}$

$\Rightarrow \vec{b} - \hat{b} \perp \text{col } A$

$\Rightarrow \vec{b} - A\hat{x} \perp \text{col } A$

$\Rightarrow$ For any column a_j of A we have $\vec{a}_j \cdot (\vec{b} - A\hat{x}) = 0$

$\Rightarrow \vec{a}_j^T (\vec{b} - A\hat{x}) = 0$

$\Rightarrow A^T(\vec{b} - A\hat{x}) = \vec{0}$

$\Rightarrow A^T\vec{b} - A^T A\hat{x} = \vec{0}$ and $A^T\vec{b} = A^T A\hat{x}$

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

That is $\hat{x} = (A^T A)^{-1} A^T \vec{b}$ when $A^T A$ is invertible.

6.5 & 6.6: Least-Squares Problems

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

$$A=\begin{bmatrix}-1 & 2\\ 2 & -3\\ -1 & 3\end{bmatrix},\ \mathbf{b}=\begin{bmatrix}4\\1\\2\end{bmatrix}$$

$$A^TA=\begin{bmatrix}-1 & 2 & -1\\ 2 & -3 & 3\end{bmatrix}\begin{bmatrix}-1 & 2\\ 2 & -3\\ -1 & 3\end{bmatrix}=\begin{bmatrix}6 & -11\\ -11 & 22\end{bmatrix}$$

$$A^T\vec{b}=\begin{bmatrix}-1 & 2 & -1\\ 2 & -3 & 3\end{bmatrix}\begin{bmatrix}4\\1\\2\end{bmatrix}=\begin{bmatrix}-4\\11\end{bmatrix}$$

$$(A^TA)^{-1}=\frac{1}{11}\begin{bmatrix}22 & 11\\ 11 & 6\end{bmatrix}=\begin{bmatrix}2 & 1\\ 1 & 6/11\end{bmatrix}$$

AND $$\hat{x}=(A^TA)^{-1}(A^T\vec{b})=\begin{bmatrix}2 & 1\\ 1 & 6/11\end{bmatrix}\begin{bmatrix}-4\\11\end{bmatrix}=\begin{bmatrix}3\\2\end{bmatrix}$$

check

$$A\hat{x}=\begin{bmatrix}-1 & 2\\ 2 & -3\\ -1 & 3\end{bmatrix}\begin{bmatrix}3\\2\end{bmatrix}=\begin{bmatrix}1\\0\\3\end{bmatrix}$$

$$\hat{b}=\begin{bmatrix}1\\0\\3\end{bmatrix}\approx\begin{bmatrix}4\\1\\2\end{bmatrix}=\vec{b}$$

Ex 2: Find a least-squares solution of the inconsistent system $A\mathbf{x}=\mathbf{b}$ for

$$A=\begin{bmatrix}1 & 1 & 0\\ 1 & 1 & 0\\ 1 & 0 & 1\\ 1 & 0 & 1\end{bmatrix},\ \mathbf{b}=\begin{bmatrix}1\\3\\8\\2\end{bmatrix}$$

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

$$A^T\vec{b}=\begin{bmatrix}14\\4\\10\end{bmatrix}$$

To find a least squares soln, we find rref($[A^TA \mid A^T\vec{b}]$)

$$\left[\begin{array}{ccc|c}4 & 2 & 2 & 14\\ 2 & 2 & 0 & 4\\ 2 & 0 & 2 & 10\end{array}\right]\sim\left[\begin{array}{ccc|c}1 & 0 & 1 & 5\\ 0 & 1 & -1 & -3\\ 0 & 0 & 0 & 0\end{array}\right]$$

$$\Rightarrow \hat{x}=\begin{bmatrix}5\\-3\\0\end{bmatrix}+x_3\begin{bmatrix}-1\\1\\1\end{bmatrix}$$

There are infinite number of least-squares solutions

6.5 & 6.6: Least-Squares Problems

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 **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

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

The distance from **b** to *A***x** is called the least - squares error

Ex 3: Find the least-squares error of Ex 1.

recall $\vec{b} = \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix}$ and $\hat{b} = A\hat{x} = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}$

least squares error $= \|\vec{b} - A\hat{x}\| = \left\| \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} \right\| = \sqrt{3^2 + 1^2 + (-1)^2} = \sqrt{11}$

If the columns of A are orthogonal, the least-squares solution is even easier to find.

Ex 4: Verify the columns of A are orthogonal and find a least-squares solution of $A\mathbf{x} = \mathbf{b}$.

$$A = \begin{array}{c} \begin{array}{ccc} \vec{v}_1 & \vec{v}_2 & \vec{v}_3 \end{array} \\ \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \\ -1 & 1 & -1 \end{bmatrix} \end{array}, \quad \mathbf{b} = \begin{bmatrix} 2 \\ 5 \\ 6 \\ 6 \end{bmatrix}$$

$\vec{v}_1 \cdot \vec{v}_2 = 0$

$\vec{v}_1 \cdot \vec{v}_3 = 0$

$\vec{v}_2 \cdot \vec{v}_3 = 0$

$$\hat{b} = \frac{\vec{b} \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 + \frac{\vec{b} \cdot \vec{v}_2}{\vec{v}_2 \cdot \vec{v}_2}\vec{v}_2 + \frac{\vec{b} \cdot \vec{v}_3}{\vec{v}_3 \cdot \vec{v}_3}\vec{v}_3$$

$$= \frac{1}{3}\vec{v}_1 + \frac{14}{3}\vec{v}_2 + \frac{-5}{3}\vec{v}_3$$

$$= \frac{1}{3}\begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \end{bmatrix} + \frac{14}{3}\begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} - \frac{5}{3}\begin{bmatrix} 0 \\ -1 \\ 1 \\ -1 \end{bmatrix}$$

$= \begin{bmatrix} 5 \\ 2 \\ 3 \\ 6 \end{bmatrix}$ and $A\hat{x} = \hat{b}$ w/ $\hat{x} = \begin{bmatrix} 1/3 \\ 14/3 \\ -5/3 \end{bmatrix}$

Practice Problems

1. Let $A = \begin{bmatrix} 1 & -3 & -3 \\ 1 & 5 & 1 \\ 1 & 7 & 2 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 5 \\ -3 \\ -5 \end{bmatrix}$. Find a least-squares solution of $A\mathbf{x} = \mathbf{b}$, and compute the associated least-squares error.

$$A^T A = \begin{bmatrix} 3 & 9 & 0 \\ 9 & 83 & 28 \\ 0 & 28 & 14 \end{bmatrix}$$

$$A^T \vec{b} = \begin{bmatrix} -3 \\ -65 \\ -28 \end{bmatrix}$$

solve $A^T A \hat{x} = A^T \vec{b}$

$$\left[\begin{array}{ccc|c} 3 & 9 & 0 & -3 \\ 9 & 83 & 28 & -65 \\ 0 & 28 & 14 & -28 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & -3/2 & 2 \\ 0 & 1 & 1/2 & -1 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

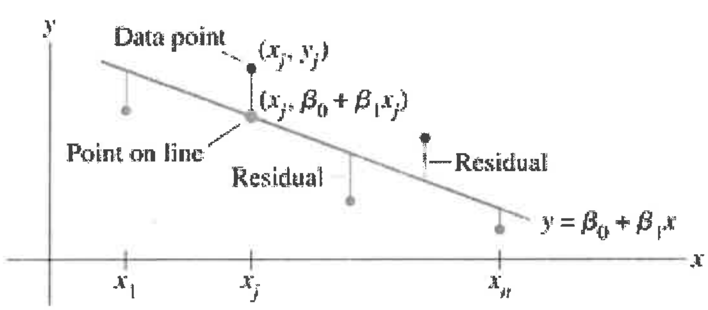
$$\hat{x} = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 3/2 \\ -1/2 \\ 1 \end{bmatrix}$$

since we need "a solution", set $x_3 = 0$.

$$A\hat{x} = A \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ -3 \\ -5 \end{bmatrix}$$

$\Rightarrow \|A\hat{x} - \vec{b}\| = 0$. This is because $\vec{b} \in \operatorname{col} A$.

Now we're going to look at finding a best-fit line for a set of data points, also known as linear-regression.



Predicted y-value		Observed y-value
$\beta_0 + \beta_1 x_1$	$=$	y_1
$\beta_0 + \beta_1 x_2$	$=$	y_2
$\vdots$		$\vdots$
$\beta_0 + \beta_1 x_n$	$=$	y_n

$$X\beta = \mathbf{y}, \quad \text{where } X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

6.5 & 6.6: Least-Squares Problems

Ex 5: Find the equation $y = \beta_0 + \beta_1 x$ of the least-squares line that best fits the data points. $(1,1),(4,2),(8,4),(11,5)$

$$y = \beta_0 + \beta_1 x \Rightarrow \begin{matrix} 1 = \beta_0 + \beta_1 \cdot 1 \\ 2 = \beta_0 + \beta_1 \cdot 4 \\ 4 = \beta_0 + \beta_1 \cdot 8 \\ 5 = \beta_0 + \beta_1 \cdot 11 \end{matrix} \Rightarrow \begin{bmatrix} 1 \\ 2 \\ 4 \\ 5 \end{bmatrix} = \beta_0 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \beta_1 \begin{bmatrix} 1 \\ 4 \\ 8 \\ 11 \end{bmatrix}$$

$$\Rightarrow \underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 4 \\ 1 & 8 \\ 1 & 11 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}}_{\vec{x}} = \underbrace{\begin{bmatrix} 1 \\ 2 \\ 4 \\ 5 \end{bmatrix}}_{\vec{b}}$$

need rref $[A^TA \mid A^T\vec{b}]$ to find $\hat{x}$.

$$A^TA = \begin{bmatrix} 4 & 24 \\ 24 & 202 \end{bmatrix} \text{ and } A^T\vec{b} = \begin{bmatrix} 12 \\ 46 \end{bmatrix}$$

$$\hat{x} = \begin{bmatrix} 15/29 \\ 12/29 \end{bmatrix}$$ which means $y = \frac{15}{29} + \frac{12}{29}x$

check w/ your Texan Friend.

Ex 6: Find the quadratic regression equation $y = \beta_0 + \beta_1 x + \beta_2 x^2$ of the least-squares line that best fits the data points. $(-2,12),(-1,5),(0,3),(1,2),(2,4)$.

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 \Rightarrow \begin{matrix} 12 = \beta_0 + \beta_1(-2) + \beta_2(-2)^2 \\ 5 = \beta_0 + \beta_1(-1) + \beta_2(-1)^2 \\ 3 = \beta_0 + \beta_1(0) + \beta_2(0)^2 \\ 2 = \beta_0 + \beta_1(1) + \beta_2(1)^2 \\ 4 = \beta_0 + \beta_1(2) + \beta_2(2)^2 \end{matrix} \Rightarrow \begin{bmatrix} 12 \\ 5 \\ 3 \\ 2 \\ 4 \end{bmatrix} = \beta_0 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \beta_1 \begin{bmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{bmatrix} + \beta_2 \begin{bmatrix} 4 \\ 1 \\ 0 \\ 1 \\ 4 \end{bmatrix}$$

$$\Rightarrow \underbrace{\begin{bmatrix} 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}}_{\vec{x}} = \underbrace{\begin{bmatrix} 12 \\ 5 \\ 3 \\ 2 \\ 4 \end{bmatrix}}_{\vec{b}}$$

$$A^TA = \begin{bmatrix} 5 & 0 & 10 \\ 0 & 10 & 0 \\ 10 & 0 & 34 \end{bmatrix} \text{ and } A^T\vec{b} = \begin{bmatrix} 26 \\ -19 \\ 71 \end{bmatrix}$$

And we can now reduce to find

$$\hat{x} = \begin{bmatrix} 87/35 \\ -19/10 \\ 19/14 \end{bmatrix}$$ which means

$$y = \frac{87}{35} - \frac{19}{10}x + \frac{19}{14}x^2$$

6.5 & 6.6: Least-Squares Problems

The General Linear Model

In some applications, it is necessary to fit data points with something other than a straight line. In the examples that follow, the matrix equation is still $X\beta = \mathbf{y}$, but the specific form of X changes from one problem to the next. Statisticians usually introduce a **residual vector** $\in$, defined by $\in = \mathbf{y} - X\beta$, and write

$$\mathbf{y} = X\beta + \in$$

Any equation of this form is referred to as a **linear model**. Once X and $\mathbf{y}$ are determined, the goal is to minimize the length of $\in$, which amounts to finding a least-squares solution of $X\beta = \mathbf{y}$. In each case, the least-squares solution $\widehat{\beta}$ is a solution of the normal equations

$$X^T X\beta = X^T\mathbf{y}$$

9. A certain experiment produces the data $(1, 7.9)$, $(2, 5.4)$, and $(3, -.9)$. Describe the model that produces a least-squares fit of these points by a function of the form

$$y = A\cos x + B\sin x$$

$$y = A\cos x + B\sin x \Rightarrow \begin{aligned} 7.9 &= A\cos 1 + B\sin 1 \\ 5.4 &= A\cos 2 + B\sin 2 \\ -0.9 &= A\cos 3 + B\sin 3 \end{aligned}$$

$$\Rightarrow \begin{bmatrix} 7.9 \\ 5.4 \\ -0.9 \end{bmatrix} = A\begin{bmatrix} \cos 1 \\ \cos 2 \\ \cos 3 \end{bmatrix} + B\begin{bmatrix} \sin 1 \\ \sin 2 \\ \sin 3 \end{bmatrix}$$

$$\Rightarrow \underbrace{\begin{bmatrix} \cos 1 & \sin 1 \\ \cos 2 & \sin 2 \\ \cos 3 & \sin 3 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} A \\ B \end{bmatrix}}_{\vec{x}} = \underbrace{\begin{bmatrix} 7.9 \\ 5.4 \\ -0.9 \end{bmatrix}}_{\vec{b}}$$

The approximate solution $\hat{x} = (A^T A)^{-1} A^T \vec{b} = \begin{bmatrix} 2.3421 \\ 7.4475 \end{bmatrix}$

and so the model is $y = 2.3421\cos(x) + 7.4475\sin(x)$.