Math 54 Section Worksheet 15 Solutions

GSI: Jeremy Meza
Office Hours: Mon 3:30-5:30pm, Zoom ID: 7621822286
Friday April 3, 2020

1 Warm-Up

True or False? If W is a subspace with basis $\{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ and $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is another basis of W that is also orthogonal, then $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ was gotten from $\{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ by Gram-Schmidt. False.

2 Virtually Together

This might be the moment everyone has been waiting for... a first "real world" application. Let's see how orthogonal projections can be used to find *best fit lines* for data.

Say we're given data points $(x_1, y_1), \ldots, (x_n, y_n)$. Technically in the "real world" our data points are usually high-dimensional, which means instead of our data points being vectors in $\mathbb{R}^2$, they are data points in $\mathbb{R}^N$ for some large N. We will stick with $\mathbb{R}^2$ for now, as the process easily generalizes.[1] We would like to find a line that goes through all the data points at once. Unfortunately, this is typically impossible, due to factors such as round-off error and noise in the data collection. There is a method, known as *linear regression* of finding the best line possible. We should first define what it means to be "best", which we do by reframing the question more linear-algebraically.

A line is determined by its slope and intercept, so we wish to find m, b that satisfy $\mathbf{y} = m\mathbf{x} + b$, or

$$\mathbf{y} = A\begin{pmatrix} m \\ b \end{pmatrix}, \qquad \text{where} \quad A = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix}, \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Like we said though, this is not possible if $\mathbf{y}$ is not in $\operatorname{Col} A$. But what vector is in $\operatorname{Col} A$ and is as close as possible to $\mathbf{y}$? That's right, the orthogonal projection of $\mathbf{y}$ onto $\operatorname{Col} A$. Thus, we seek vectors $\left(\begin{smallmatrix} m \\ b \end{smallmatrix}\right)$ that instead satisfy

$$\operatorname{proj}_{\operatorname{Col} A} \mathbf{y} = A\begin{pmatrix} m \\ b \end{pmatrix}$$

We call this vector $\left(\begin{smallmatrix} m \\ b \end{smallmatrix}\right)$ a **least-squares solution**. Note that this solution is not necessarily unique! The orthogonal projection is unique, but if A has a nonzero null space, then there could be several least-squares solutions. Now, here's a slick way to find a least-square solution: we know that $\mathbf{y} - \operatorname{proj}_{\operatorname{Col} A} \mathbf{y}$ must be orthogonal to $\operatorname{Col} A$, and we know that the orthogonal complement of $\operatorname{Col} A$ is $\operatorname{Nul} A^T$. So, a least-squares solution $\left(\begin{smallmatrix} m \\ b \end{smallmatrix}\right)$ must satisfy

$$0 = A^T(y - \operatorname{proj}_{\operatorname{Col} A} \mathbf{y}) = A^T\left(\mathbf{y} - A\begin{pmatrix} m \\ b \end{pmatrix}\right) \implies A^T A\begin{pmatrix} m \\ b \end{pmatrix} = A^T\mathbf{y}$$

which we call the **normal equations**. Solving the normal equations solves for the least-squares solution, which in this particular case solves for the slope m and intercept b of the best-fit line. The reason we call this the "least-squares" solution is because this line minimizes the squared distance $\|\mathbf{y} - (m\mathbf{x} + b)\|^2$.

[1] Alternatively, instead of doing linear regression on a high-dimensional data set, do you think you might know a good way to reduce a high dimensional vector to a vector in $\mathbb{R}^n$ for some much smaller n? That's right, you can orthogonally project! But what space should you project onto? Choosing the right space is at the center of PCA or *principal component analysis*, which in this class goes under the guise of SVD, or *singular value decomposition*.

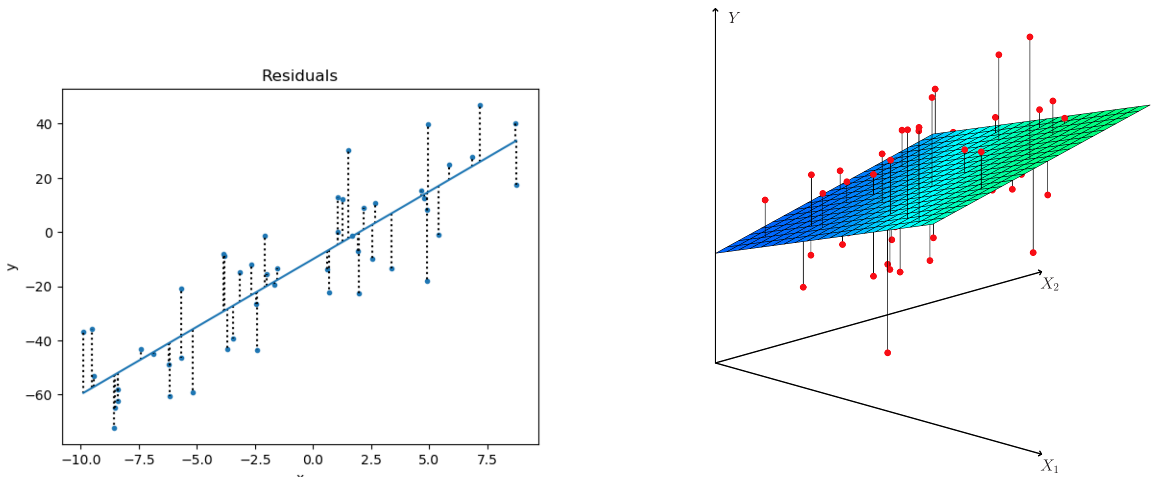

Figure 1: Left: A plot of a least-squares line fit to two-dimensional data points. Following the dashed lines from each point to the line gives the prediction $\widehat{y}_i := mx_i + b$. The line shown minimizes the sum of the squares of the lengths of the dashed lines. Right: In a three-dimensional setting the least squares line becomes a plane.

3 Problems

1. Let $\mathbf{b} = \begin{pmatrix} 4 \\ 0 \\ 2 \end{pmatrix}$ and let $A = \begin{pmatrix} 1 & 0 & -2 \\ 2 & 1 & -1 \\ 0 & -1 & -3 \end{pmatrix}$.

   (a) Find an orthonormal basis for the column space of A.

   (If I had known this calculation was so messy, I wouldn't have assigned it.. oops!) Gram-Schmidt on the columns gives the orthogonal basis $\left\{ \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} -2/5 \\ 1/5 \\ -1 \end{pmatrix} \right\}$. Note that the basis consists of only 2 vectors, since the 3rd column is in the span of the first two. To get an orthonormal basis, we divide the first vector by $\sqrt{5}$ and the second by $\sqrt{6/5}$.

   (b) Find a least squares solution to $A\mathbf{x} = \mathbf{b}$.

   We solve $A^T Ax = A^T b$, which gives $\mathbf{x} = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix} + t \begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix}$ for any $t \in \mathbb{R}$. Note the infinite number of least squares solutions!

   (c) Find the shortest distance from $\mathbf{b}$ to the column space of A.

   Let $\widehat{\mathbf{b}} = A\widehat{\mathbf{x}}$ where $\widehat{\mathbf{x}}$ is any answer from part (b). Then, $\widehat{\mathbf{b}} = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$ and so the shortest distance is

   $$\|\mathbf{b} - \widehat{\mathbf{b}}\| = \|(2, -1, -1)^T\| = \sqrt{6}$$

   Note that no matter our choice of $\widehat{\mathbf{x}}$, the vector $\widehat{\mathbf{b}}$ is always the same. This is because $\widehat{\mathbf{b}}$ is the orthogonal projection of $\mathbf{b}$ onto $\operatorname{Col} A$, and so is unique.

2. Let $A = \begin{pmatrix} 1 & 5 \\ 3 & 1 \\ -2 & 4 \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} 4 \\ -2 \\ -3 \end{pmatrix}$.

(a) Find a least-squares solution of $A\mathbf{x} = \mathbf{b}$.

$\widehat{\mathbf{x}} = \begin{pmatrix} 2/7 \\ 1/7 \end{pmatrix}$.

(b) Find the orthogonal projection of $\mathbf{b}$ onto $\operatorname{Col} A$.

The columns of A are already orthogonal, so we don't need to do Gram-Schmidt. If $\mathbf{a}_1, \mathbf{a}_2$ denote the columns of A, then the projection is

$$\operatorname{proj}_{\operatorname{Col} A} \mathbf{b} = \frac{\mathbf{b} \cdot \mathbf{a}_1}{\mathbf{a}_1 \cdot \mathbf{a}_1}\mathbf{a}_1 + \frac{\mathbf{b} \cdot \mathbf{a}_2}{\mathbf{a}_2 \cdot \mathbf{a}_2}\mathbf{a}_2 = \frac{2}{7}\begin{pmatrix} 1 \\ 3 \\ -2 \end{pmatrix} + \frac{1}{7}\begin{pmatrix} 5 \\ 1 \\ 4 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

(c) Check your answer to part (a) using part (b). We note that

$$A\widehat{\mathbf{x}} = A\begin{pmatrix} 2/7 \\ 1/7 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \operatorname{proj}_{\operatorname{Col} A} \mathbf{b}$$

3. True or False?

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

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

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

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

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

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

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