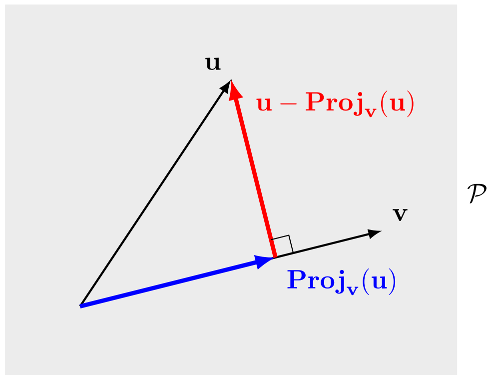
Math 51
Lecture 7 – July 7, 2023

Goal: orthogonal basis for a plane in $\mathbb{R}^n$, linear regression

In Lecture 5, we saw how to find an orthogonal basis from a basis by using linear combinations and dot products. Now that we know about projections from Lecture 6, we can find an orthogonal basis much more systematically. Given two vectors $\mathbf{u}$ and $\mathbf{v}$ in a plane $\mathcal{P}$ in $\mathbb{R}^n$, we know that $\mathbf{Proj}_{\mathbf{v}}(\mathbf{u})$ and $\mathbf{u} - \mathbf{Proj}_{\mathbf{v}}(\mathbf{u})$ lie on the plane and are orthogonal to each other.



Theorem 7.1.1. Suppose $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ are nonzero and not scalar multiples of each other. Then, the vectors

$$\mathbf{y} \qquad \text{and} \qquad \mathbf{x}' = \mathbf{x} - \mathbf{Proj}_{\mathbf{y}}(\mathbf{x})$$

form an orthogonal basis for the plane $\text{span}(\mathbf{x}, \mathbf{y})$. Note that this implies that the plane is 2-dimensional. Another orthogonal basis for this plane is $\{\mathbf{x}, \mathbf{y}' = \mathbf{y} - \mathbf{Proj}_{\mathbf{x}}(\mathbf{y})\}$.

Example 1. Consider the plane $\mathcal{P}$ in $\mathbb{R}^3$ that is spanned by $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ 5 \\ 1 \end{bmatrix}$. Find an orthogonal basis for $\mathcal{P}$ that contains $\mathbf{u}$.

We see that $\text{Proj}_{\vec{u}}(\vec{v}) = \left(\frac{\vec{v}\cdot\vec{u}}{\vec{u}\cdot\vec{u}}\right)\vec{u} = \frac{14}{14}\vec{u} = \vec{u}$. So,

$$\vec{v} - \text{Proj}_{\vec{u}}(\vec{v}) = \vec{v} - \vec{u} = \begin{bmatrix} 0 \\ 3 \\ -2 \end{bmatrix},$$

and thus, $\left\{\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ -2 \end{bmatrix}\right\}$ is an orthogonal basis for P.

Find an orthogonal basis for $\mathcal{P}$ that contains $\mathbf{v}$.

Similarly as above, $\text{Proj}_{\vec{v}}(\vec{u}) = \frac{14}{27}\vec{v}$, and so,

$$\vec{u} - \text{Proj}_{\vec{v}}(\vec{u}) = \vec{u} - \frac{14}{27}\vec{v} = \begin{bmatrix} \frac{13}{27} \\ -\frac{16}{27} \\ \frac{67}{27} \end{bmatrix} = \frac{1}{27}\begin{bmatrix} 13 \\ -16 \\ 67 \end{bmatrix}.$$

Hence, $\left\{\begin{bmatrix} 1 \\ 5 \\ 1 \end{bmatrix}, \begin{bmatrix} 13 \\ -16 \\ 67 \end{bmatrix}\right\}$ is an orthogonal basis for P.

Example 2. Find the projection of $\mathbf{u} = \begin{bmatrix} 2 \\ 1 \\ 0 \\ -3 \end{bmatrix}$ onto the plane $\mathcal{P}$ spanned by $\mathbf{v} = \begin{bmatrix} 1 \\ 3 \\ -1 \\ 2 \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$.

First, we need an orthogonal basis for $P = \text{span}(\vec{v}, \vec{w})$. Since $\text{Proj}_{\vec{w}}(\vec{v}) = \vec{w}$,

$$\vec{v} - \text{Proj}_{\vec{w}}(\vec{v}) = \vec{v} - \vec{w} = \begin{bmatrix} 1 \\ 2 \\ -2 \\ 2 \end{bmatrix}.$$

Thus, $\left\{ \overset{\vec{w}'}{\begin{bmatrix} 1 \\ 2 \\ -2 \\ 2 \end{bmatrix}}, \overset{\vec{w}}{\begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}} \right\}$ is an orthogonal basis for P. Hence,

$$\text{Proj}_P(\vec{u}) = \text{Proj}_{\vec{w}'}(\vec{u}) + \text{Proj}_{\vec{w}}(\vec{u}) = -\frac{2}{13}\vec{w}' + \frac{1}{2}\vec{w} = \begin{bmatrix} -\frac{2}{13} \\ \frac{5}{26} \\ \frac{21}{26} \\ -\frac{4}{13} \end{bmatrix} = \frac{1}{26}\begin{bmatrix} -4 \\ 5 \\ 21 \\ -8 \end{bmatrix}.$$

<u>Note</u> If you pick $\left\{ \begin{bmatrix} -2 \\ 9 \\ 17 \\ -4 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ -1 \\ 2 \end{bmatrix} \right\}$ as an orthogonal basis, you will get the same answer.

Example 3. Find an orthogonal basis for $\mathbb{R}^3$ that contains $\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$.

We need an orthogonal basis for P defined by $x + y - z = 0$. Two (arbitrary) vectors on P are $\underset{\vec{u}}{\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}}$ and $\underset{\vec{v}}{\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}}$. Then, $\text{Proj}_{\vec{v}}(\vec{u}) = \frac{5}{2}\vec{v}$, and so,

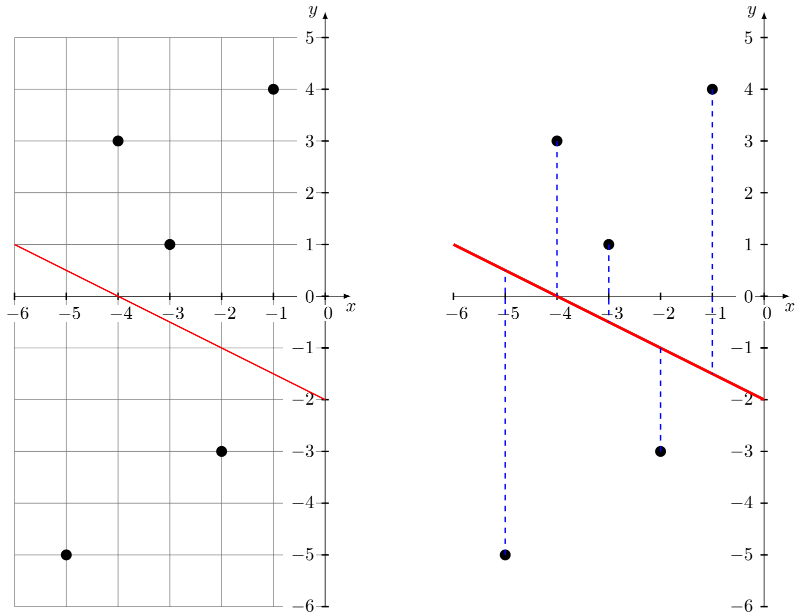
$$\vec{u} - \text{Proj}_{\vec{v}}(\vec{u}) = \begin{bmatrix} 1 \\ -\frac{1}{2} \\ \frac{1}{2} \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}.$$

Hence, $\left\{ \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} \right\}$ is an orthogonal basis for P, and so, $\boxed{\left\{ \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} \right\}}$ is an orthogonal basis for $\mathbb{R}^3$.

Linear regression is the process of finding a line that models a number of data points as closely as possible. We will explain this process by walking through the following data set:

$$(-5,-5) \qquad (-4,3) \qquad (-3,1) \qquad (-2,-3) \qquad (-1,4).$$

Let us look at the plot in the xy-coordinate plane and approximate with the red line.



Clearly, the red line is not a "good" approximate of the data. But how do we really measure how good an approximate is? One way is called the **least squares method**. If our best fit line is given by $y = mx + b$, then the approximation/prediction made for a given x_i is $mx_i + b$. Thus, the error in approximation would be

$$y_i - (mx_i + b).$$

The goal of the least squares method is to come up with m and b that minimizes the quantity

$$\text{SSE(sum of square errors)} \longrightarrow \sum_{i=1}^{n} (y_i - (mx_i + b))^2. \qquad \vec{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

Let us assume that not all data points lie on a vertical line.

1. Define $\mathbf{X} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ and $\mathbf{Y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$. Note that $m\mathbf{X} + b\mathbf{1} = \begin{bmatrix} mx_1 + b \\ \vdots \\ mx_n + b \end{bmatrix}$ list the predicted y-values, by $y = mx + b$, for $x_1, \ldots, x_n$. Our goal is to minimize

$$\sum_{i=1}^{n} (y_i - (mx_i + b))^2 = \|\mathbf{Y} - (m\mathbf{X} + b\mathbf{1})\|^2.$$

2. We can think about the vector $\mathbf{Y} - (m\mathbf{X} + b\mathbf{1})$ as the displacement vector from $\mathbf{Y}$ to a vector $m\mathbf{X} + b\mathbf{1}$, which lies in the plane $\mathcal{P}$ spanned by $\mathbf{X}$ and $\mathbf{1}$. Hence, minimizing $\|\mathbf{Y} - (m\mathbf{X} + b\mathbf{1})\|^2$ (or equivalently minimizing $\|\mathbf{Y} - (m\mathbf{X} + b\mathbf{1})\|$) is finding the point on the plane *closest* to $\mathbf{Y}$. We know this closest point is the projection of $\mathbf{Y}$ onto $\mathcal{P}$. So, we need to compute an orthogonal basis for $\mathcal{P}$. It turns out that $\left\{\mathbf{1}, \widehat{\mathbf{X}}\right\}$ form an orthogonal basis for $\mathcal{P}$, where $\widehat{\mathbf{X}} = \mathbf{X} - \bar{x}\mathbf{1}$. $\longleftarrow$ Verify this!

3. Project $\mathbf{Y}$ onto $\mathcal{P}$:

$$\mathbf{Proj}_{\mathcal{P}}(\mathbf{Y}) = \mathbf{Proj}_{\mathbf{1}}(\mathbf{Y}) + \mathbf{Proj}_{\widehat{\mathbf{X}}}(\mathbf{Y}) = \bar{y}\mathbf{1} + \left(\frac{\mathbf{Y} \cdot \widehat{\mathbf{X}}}{\widehat{\mathbf{X}} \cdot \widehat{\mathbf{X}}}\right)\widehat{\mathbf{X}}.$$

$$= m\vec{X} + b\vec{1}.$$

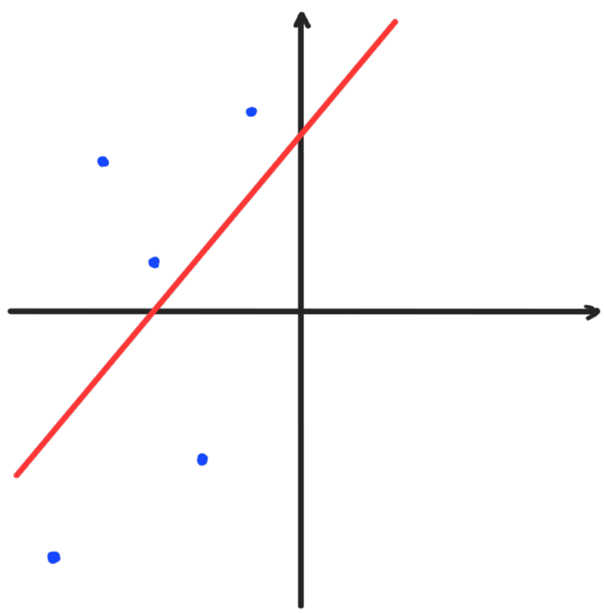
Example 4. Find the best fit line for the above data.

$$(-5,-5) \qquad (-4,3) \qquad (-3,1) \qquad (-2,-3) \qquad (-1,4)$$

We get $\vec{X} = \begin{bmatrix} -5 \\ -4 \\ -3 \\ -2 \\ -1 \end{bmatrix}$ and $\vec{Y} = \begin{bmatrix} -5 \\ 3 \\ 1 \\ -3 \\ 4 \end{bmatrix}$. Since $\bar{x} = -3$, we get $\hat{X} = \vec{X} - (-3)\vec{1} = \begin{bmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{bmatrix}$. So,

$$\begin{aligned} \mathrm{Proj}_P(\vec{Y}) &= \mathrm{Proj}_{\hat{X}}(\vec{Y}) + \mathrm{Proj}_{\vec{1}}(\vec{Y}) = \frac{12}{10}\hat{X} + \frac{0}{5}\vec{1} \\ &= \frac{6}{5}(\vec{X} + 3\vec{1}) = \underbrace{\frac{6}{5}}_{m}\vec{X} + \underbrace{\frac{18}{5}}_{b}\vec{1}. \end{aligned}$$

Hence, the line of best fit is $\boxed{y = \frac{6}{5}x + \frac{18}{5}.}$



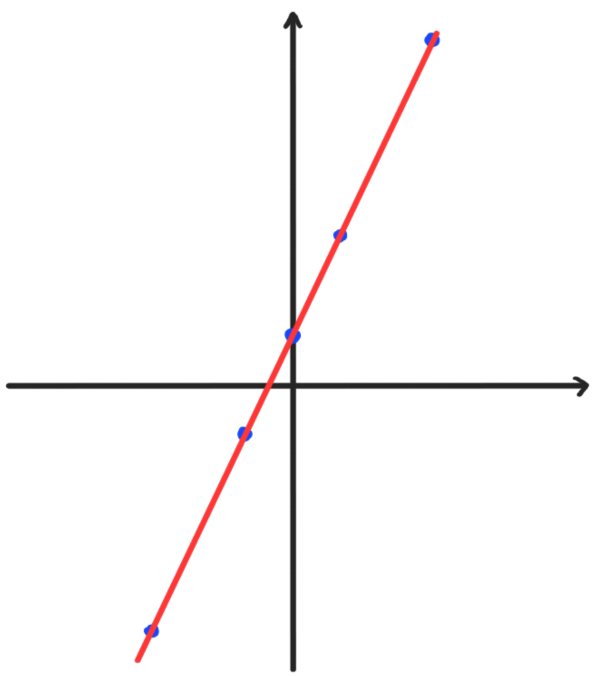
Example 5. Find the best fit line for

$$(-3,-5) \qquad (-1,-1) \qquad (0,1) \qquad (1,3) \qquad (3,7)$$

Since $\vec{X} = \begin{bmatrix} -3 \\ -1 \\ 0 \\ 1 \\ 3 \end{bmatrix}$ and $\vec{Y} = \begin{bmatrix} -5 \\ -1 \\ 1 \\ 3 \\ 7 \end{bmatrix}$, $\bar{x} = 0$, and so, $\hat{\vec{X}} = \vec{X} - 0\vec{1} = \begin{bmatrix} -3 \\ -1 \\ 0 \\ 1 \\ 3 \end{bmatrix}$. Hence,

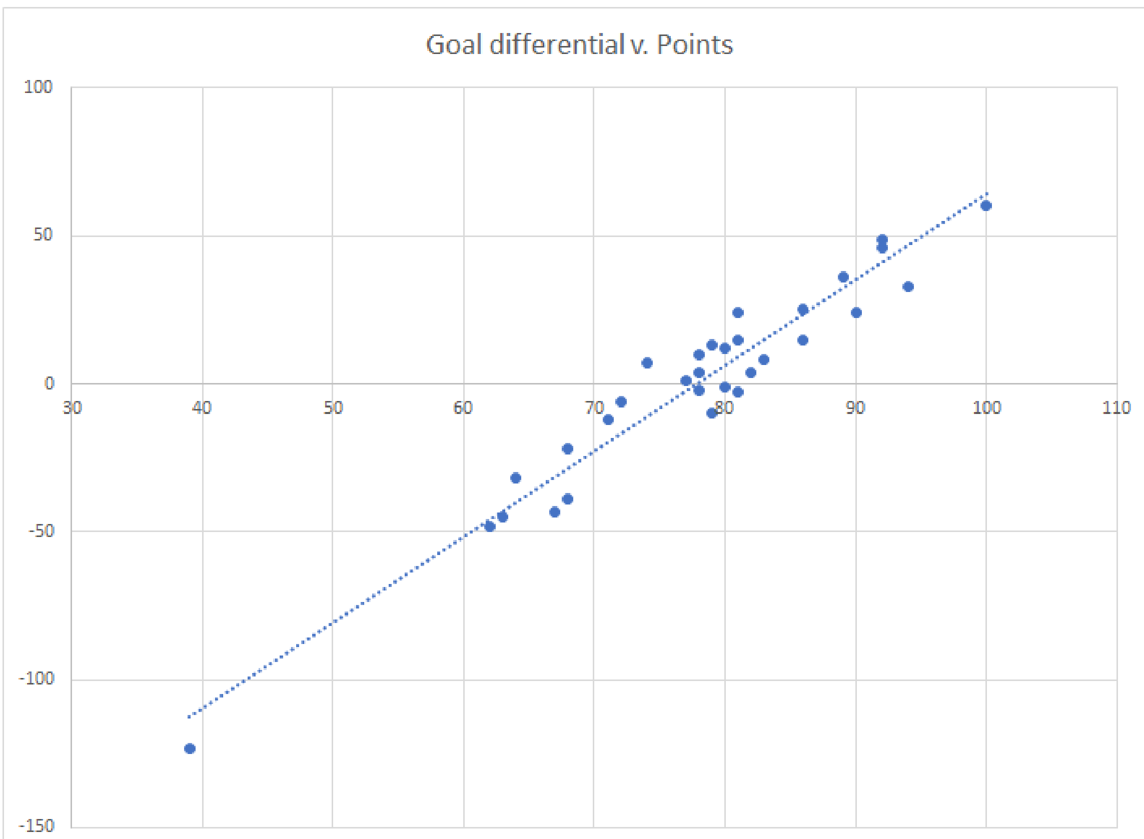
$$\begin{aligned} \mathrm{Proj}_P(\vec{Y}) &= \mathrm{Proj}_{\hat{\vec{X}}}(\vec{Y}) + \mathrm{Proj}_{\vec{1}}(\vec{Y}) = \frac{40}{20}\hat{\vec{X}} + \frac{5}{5}\vec{1} = 2\hat{\vec{X}} + 1\vec{1} \\ &= \underset{m}{\underline{2}}\vec{X} + \underset{b}{\underline{1}}\vec{1}. \end{aligned}$$

Thus, $y = 2x + 1$ is the line of best fit.



Example 6. Consider the following data from the 2019–2020 NHL (National Hockey League) season.

Team	Points	Goal Diff.	Team	Points	Goal Diff.
Boston Bruins	100	60	Calgary Flames	79	–10
St. Louis Blues	94	33	Vancouver Canucks	78	10
Colorado Avalanche	92	46	Florida Panthers	78	4
Tampa Bay Lightning	92	49	Nashville Predators	78	–2
Washington Capitals	90	24	Minnesota Wild	77	1
Philadelphia Flyers	89	36	Arizona Coyotes	74	7
Vegas Golden Knights	86	15	Chicago Blackhawks	72	–6
Pittsburgh Penguins	86	25	Montreal Canadiens	71	–12
Edmonton Oilers	83	8	Buffalo Sabres	68	–22
Dallas Stars	82	4	New Jersey Devils	68	–39
Columbus Blue Jackets	81	–3	**Anaheim Ducks**	67	–43
Toronto Maple Leafs	81	15	Los Angeles Kings	64	–32
Carolina Hurricanes	81	24	San Jose Sharks	63	–45
Winnipeg Jets	80	12	Ottawa Senators	62	–48
New York Islanders	80	–1	Detroit Red Wings	39	–123
New York Rangers	79	13			



Example 7. Suppose that the collection of 51 data points

$$(x_1, y_1),\ (x_2, y_2),\ \ldots,\ (x_{51}, y_{51})$$

does not lie on a single line and has $y = 3x - 4$ as its line of best fit.

(a) Let $\mathcal{P} = \mathrm{span}(\mathbf{1}, \mathbf{X})$. The projection

$$\mathbf{Proj}_{\mathcal{P}}(\mathbf{Y})$$

is a linear combination $s\mathbf{1} + t\mathbf{X}$, where s and t are scalars. Either determine s and t with brief justification, or else explain why there is not enough information given to do so.

Since the line of best fit is $y = 3x - 4$, it follows that

$$\mathrm{Proj}_{\mathcal{P}}(\vec{Y}) = \underbrace{3}_{t}\vec{X} - \underbrace{4}_{s}\vec{1}$$

For the rest of this problem, now consider a new "transformed" data set based on the original data above, given by the following:

$$(2x_1 - 4,\ -y_1),\ (2x_2 - 4,\ -y_2),\ \ldots,\ (2x_{51} - 4,\ -y_{51}).$$

(b) Find expressions, in terms of $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{1}$, for the 51-vectors $\mathbf{X}'$ and $\mathbf{Y}'$ which consist of the x- and y-values (respectively) for the new data set.

The ith data point of the transformed data set is

$$(2x_i - 4, -y_i).$$

Thus, $\boxed{\vec{X}' = 2\vec{X} - 4\vec{1}}$ and $\boxed{\vec{Y}' = -\vec{Y}}$.

(c) As defined above, we have $\mathcal{P} = \mathrm{span}(\mathbf{1}, \mathbf{X})$; it is also true, which you do not have to prove, that $\mathcal{P}$ is equal to $\mathrm{span}(\mathbf{1}, \mathbf{X}')$. Compute

$$\mathbf{Proj}_{\mathcal{P}}(\mathbf{Y}')$$

as a linear combination of $\mathbf{1}$ and $\mathbf{X}'$. Show all the steps of your reasoning.

We see that

$$\mathrm{Proj}_{\mathcal{P}}(\vec{Y}') = -\mathrm{Proj}_{\mathcal{P}}(\vec{Y}) = -(3\vec{X} - 4\vec{1}) = -3\vec{X} + 4\vec{1}$$

$$= -3\left(\tfrac{1}{2}\vec{X}' + 2\vec{1}\right) + 4\vec{1} = \boxed{-\tfrac{3}{2}\vec{X}' - 2\vec{1}}.$$

This implies that the line of best fit for the transformed data is

$$y = -\tfrac{3}{2}x - 2.$$