Goals:

- answer questions and discuss homework questions
- Go over the following Chapter 7 topics:
  - *Using projections (and thus orthonormal bases) to solve problems*
  - *Fitting data to least squares lines*
- Do related problems

If V is a subspace spanned by non-redundant vectors $\mathbf{v_1}, \mathbf{v_2}$, then an orthogonal basis for V is given by $\mathbf{w_1} = \mathbf{v_1}$ and

$$\mathbf{w_2} = \mathbf{v_2} - \frac{\mathbf{v_2} \cdot \mathbf{v_1}}{\mathbf{v_1} \cdot \mathbf{v_1}} \mathbf{v_1}.$$

Find the distance from a vector to a subspace of 2 dimensions or closest vector

Given a vector $\mathbf{x}$ and a subspace $V = \mathrm{Span}\,\{\mathbf{v_1}, \mathbf{v_2}\}$, find (a) the vector in V that is closest to $\mathbf{x}$ and (b) find the distance from V to $\mathbf{x}$.

(a) To find the vector in V that is closest to $\mathbf{x}$

- Use $\mathbf{v_1}, \mathbf{v_2}$ to find an orthogonal basis $\mathbf{w_1}, \mathbf{w_2}$ for V.
- Compute $\mathrm{Proj}_{\mathrm{V}}\mathbf{x} = c_1\mathbf{w_1} + c_2\mathbf{w_2}$ where c_1 and c_2 are the projection (Fourier) coefficients. This is the vector in V closest to $\mathbf{x}$.

(b) The magnitude $\|\mathbf{x} - \mathrm{Proj}_{\mathrm{V}}\mathbf{x}\|$ is the distance from V to $\mathbf{x}$.

Activity 1

Find the point in the span V of $(1, 1, 1, 0)$ and $(0, 3, 3, 6)$ that is closest to $(1, 1, 1, 1)$. What is the distance between V and $(1, 1, 1, 1)$?

Solution to Activity 1

$\mathbf{w_1} = (1, 1, 1, 0)$ and

$$\mathbf{w_2} = (0, 3, 3, 6) - \frac{(0, 3, 3, 6) \cdot (1, 1, 1, 0}{(1, 1, 1, 0) \cdot (1, 1, 1, 0)}(1, 1, 1, 0)$$

$$= (0, 3, 3, 6) - \frac{6}{3}(1, 1, 1, 0) = (0, 3, 3, 6) - (2, 2, 2, 0) = (-2, 1, 1, 6).$$

$$\mathrm{Proj}_V(1, 1, 1, 1) = \frac{3}{3}(1, 1, 1, 0) + \frac{6}{42}(-2, 1, 1, 6)$$
$$= \left(\frac{30}{42}, \frac{48}{42}, \frac{48}{42}, \frac{36}{42}\right) = \left(\frac{5}{7}, \frac{8}{7}, \frac{8}{7}, \frac{6}{7}\right).$$

Distance between $(1, 1, 1, 1)$ and V is

$$\|(1, 1, 1, 1) - \left(\frac{5}{7}, \frac{8}{7}, \frac{8}{7}, \frac{6}{7}\right)\| = \| - \left(\frac{2}{7}, \frac{-1}{7}, \frac{-1}{7}, \frac{1}{7}\right)\| = \sqrt{\frac{1}{7}}.$$

If two planes have normal vectors that are not parallel, they will intersect and the distance between them is zero. If two planes are parallel, how to find the distance between them?

The book outlines a strategy for finding the distance between two lines.

Develop a strategy to find the distance between two parallel planes. Find the distance between $x + y + z = 2$ and $x + y + z = 3$.

- Shift both planes down (or over) the same amount so that one goes through the origin. That plane is now a subspace.
- Take any point on the non-subspace plane and project it onto the normal vector to both planes. The length of this vector gives the distance between the planes.

Shifting both planes down by 2 results in $x + y + z = 0$ and $x + y + z = 1$. The latter plane includes the point $(1, 0, 0)$. Project this point onto the normal vector $(1, 1, 1)$ to obtain $(1/3, 1/3, 1/3)$. This length of this vector is $\sqrt{\frac{1}{3}}$.

Strategy 2

- Shift both planes down (or over) the same amount so that one goes through the origin. That plane is now a subspace.
- Find an orthogonal basis $\mathbf{u}, \mathbf{v}$ for the subspace plane V.
- Find any point $\mathbf{x}$ on the shifted down plane that is not a subspace (call it P) and project it onto the subspace plane V to get $\mathbf{y} = \mathrm{Proj}_V \mathbf{x}$,
- The distance between P and V (which equals the distance between the original planes) is $\|\mathbf{x} - \mathbf{y}\|$.

Shift $x+y+z=2$ and $x+y+z=3$ down 2 in the z direction. The distance between them will not change. So find the distance between $x+y+z=0$ and $x+y+z=1$. The point $(1,0,0)$ is on the latter plane. The vectors $(1,0,-1)$ and $(0,-1,1)$ form a basis for $x+y+z=0$. An orthogonal basis is $(1,0,-1)$ and $(0,-1,1)-\frac{-1}{2}(1,0,-1)=(\frac{1}{2},-1,-\frac{1}{2})$. We can use $(1,-2,1)$ as our second vector. Project $(1,0,0)$ onto the span of these two orthogonal vectors to get $\frac{1}{2}(1,0,-1)+\frac{1}{6}(1,-2,1)=(\frac{2}{3},-\frac{1}{3},-\frac{1}{3})$. The distance between this last projection and $(1,0,0)$ is $\|(\frac{1}{3},\frac{1}{3},\frac{1}{3})\|=\sqrt{\frac{1}{3}}$.

Problem: Given paired data $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, find the line that best fits the data.

Note-you would only do this if the correlation coefficient r that we found earlier in the class is reasonably close to ± 1.

Let

$$\mathbf{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix}, \ \mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \\ y_n \end{pmatrix}, \ \mathbf{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{pmatrix},$$

Least Square Lines Continued

If the data were on a line then there would exist a slope m and a y-intercept b such that $y_j = mx_j + b$ for $j = 1, \ldots n$. That is, we'd have

$$\mathbf{Y} = m\mathbf{X} + b\mathbf{1}.$$

So our goal is to find m and b to minimize $\|\mathbf{Y} - (m\mathbf{X} + b\mathbf{1})\|$.

That is, find the vector in $\mathrm{span}\{\mathbf{X}, \mathbf{1}\}$ (i.e. $m\mathbf{X} + b\mathbf{1}$) that is closest to $\mathbf{Y}$.
Use $\mathbf{w_1} = \mathbf{1}$ and

$$\mathbf{w_2} = \hat{\mathbf{X}} = \mathbf{X} - \left(\frac{\mathbf{X}\cdot\mathbf{1}}{\mathbf{1}\cdot\mathbf{1}}\right)\mathbf{1} = \mathbf{X} - \left(\frac{\sum x_i}{n}\right)\mathbf{1} = \mathbf{X} - \overline{x}\mathbf{1}.$$

Least Squares Lines Continued

Next find the vector in $V = \mathrm{span}\,\{\mathbf{X}, \mathbf{1}\}$ that is closest to $\mathbf{Y}$:

$$\begin{aligned}
\mathrm{Proj}_V \mathbf{Y} &= \left(\frac{\mathbf{Y}\cdot\hat{\mathbf{X}}}{\hat{\mathbf{X}}\cdot\hat{\mathbf{X}}}\right)\hat{\mathbf{X}} + \left(\frac{\mathbf{Y}\cdot\mathbf{1}}{\mathbf{1}\cdot\mathbf{1}}\right)\mathbf{1} \\
&= \left(\frac{\mathbf{Y}\cdot\hat{\mathbf{X}}}{\hat{\mathbf{X}}\cdot\hat{\mathbf{X}}}\right)(\mathbf{X} - \overline{x}\mathbf{1}) + \overline{y}\mathbf{1} \\
&= \left(\frac{\mathbf{Y}\cdot\hat{\mathbf{X}}}{\hat{\mathbf{X}}\cdot\hat{\mathbf{X}}}\right)\mathbf{X} + \left(\overline{y} - \overline{x}\left(\frac{\mathbf{Y}\cdot\hat{\mathbf{X}}}{\hat{\mathbf{X}}\cdot\hat{\mathbf{X}}}\right)\right)\mathbf{1} \\
&= m\mathbf{X} + b\mathbf{1},
\end{aligned}$$

where $m = \frac{\mathbf{Y}\cdot\hat{\mathbf{X}}}{\hat{\mathbf{X}}\cdot\hat{\mathbf{X}}}$ and $b = \overline{y} - \overline{x}\left(\frac{\mathbf{Y}\cdot\hat{\mathbf{X}}}{\hat{\mathbf{X}}\cdot\hat{\mathbf{X}}}\right)$.

To see if this is a good fit, we can compute r or r^2 as in Chapter 2. If $|r|$ is close to 1, then the data can be reasonably approximated with a line. Recall

$\hat{\mathbf{X}} = \mathbf{X} - \overline{x}\mathbf{1}$ and $\hat{\mathbf{Y}} = \mathbf{Y} - \overline{y}\mathbf{1}$. Then

$$r = \frac{\hat{\mathbf{X}} \cdot \hat{\mathbf{Y}}}{\|\hat{\mathbf{X}}\| \|\hat{\mathbf{Y}}\|}.$$

The book shows that

$$\|\mathbf{Y} - (m\mathbf{X} + b\mathbf{1})\|^2 = (1 - r^2)\|\hat{\mathbf{Y}}\|^2.$$

Problem: Find the line of best fit through the points $(2,3),(3,2),(5,1),(6,0)$.

Problem: Find the line of best fit through the points $(2,3),(3,2),(5,1),(6,0)$.

$$\mathbf{X}=(2,3,5,6),\ \ \mathbf{Y}=(3,2,1,0),\ \ \mathbf{1}=(1,1,1,1).$$

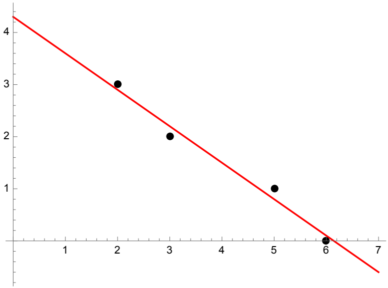
$$\overline{x}=\frac{16}{4}=4,\ \ \hat{\mathbf{X}}=(2,3,5,6)-(4,4,4,4)=(-2,-1,1,2).$$

Also $\overline{y}=\frac{3}{2}$ and $\hat{\mathbf{Y}}=\left(\frac{3}{2},\frac{1}{2},-\frac{1}{2},-\frac{3}{2}\right)$.
So an orthogonal basis for the span of $\mathbf{X}$ and $\mathbf{1}$ is $\hat{\mathbf{X}}$ and $\mathbf{1}$

$$\begin{aligned}
\mathrm{Proj}_{\mathbf{X},\mathbf{1}}\mathbf{Y} = \mathrm{Proj}_{\hat{\mathbf{X}},\mathbf{1}}\mathbf{Y} &= \frac{\mathbf{Y}\cdot\hat{\mathbf{X}}}{\hat{\mathbf{X}}\cdot\hat{\mathbf{X}}}\hat{\mathbf{X}} + \frac{\mathbf{Y}\cdot\mathbf{1}}{\mathbf{1}\cdot\mathbf{1}}\mathbf{1} \\
&= \frac{-7}{10}\hat{\mathbf{X}} + \frac{3}{2}\mathbf{1} = \frac{-7}{10}\left(\mathbf{X} - 4\mathbf{1}\right) + \frac{3}{2}\mathbf{1} \\
&= \frac{-7}{10}\hat{\mathbf{X}} + \frac{43}{10}\mathbf{1}.
\end{aligned}$$

So the line of best fit is $y = -.7x + 4.3$.

The r value is $r = \frac{\hat{\mathbf{X}}\cdot\hat{\mathbf{Y}}}{||\hat{\mathbf{X}}||\,||\hat{\mathbf{Y}}||} = \frac{-7}{\sqrt{10}\sqrt{5}} = -.9899$.


4
3
2
1
1
2
3
4
5
6
7

Activity 4

Given date $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ that are suspected to be close to fitting a parabola

$$y = ax^2 + bx + c,$$

set up a process to find the coefficients a, b, c that give the best choice for the parabola.

Let

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \\ y_n \end{pmatrix}, \ \mathbf{X_1} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix}, \ \mathbf{X_2} = \begin{pmatrix} x_1^2 \\ x_2^2 \\ \vdots \\ x_{n-1}^2 \\ x_n^2 \end{pmatrix}, \mathbf{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{pmatrix},$$

We want to minimize $\|\mathbf{Y} - (a\mathbf{X_2} + b\mathbf{X_1} + c\mathbf{1})\|$.

We'll do this by forming an orthogonal basis for $V = \mathrm{span}\left\{\mathbf{1}, \mathbf{X_1}, \mathbf{X_2}\right\}$ and using that basis to compute $\mathrm{Proj}_V\mathbf{Y}$.

Let $\mathbf{w_1} = \mathbf{1}$,

$$\mathbf{w_2} = \hat{\mathbf{X}}_\mathbf{1} = \mathbf{X_1} - \left(\frac{\mathbf{X_1}\cdot\mathbf{1}}{\mathbf{1}\cdot\mathbf{1}}\right)\mathbf{1}.$$

Let $\mathbf{w_1} = \mathbf{1}$,

$$\mathbf{w_2} = \hat{\mathbf{X}}_1 = \mathbf{X_1} - \left(\frac{\mathbf{X_1} \cdot \mathbf{1}}{\mathbf{1} \cdot \mathbf{1}}\right) \mathbf{1}.$$

We'll do this seriously in a later chapter but we can also find a third orthogonal vector:

$$\mathbf{w_3} = \hat{\mathbf{X}}_2 = \mathbf{X_2} - \left(\frac{\mathbf{X_2} \cdot \mathbf{1}}{\mathbf{1} \cdot \mathbf{1}}\right) \mathbf{1} - \left(\frac{\mathbf{X_2} \cdot \hat{\mathbf{X}}_1}{\hat{\mathbf{X}}_1 \cdot \hat{\mathbf{X}}_1}\right) \hat{\mathbf{X}}_1.$$

Now find

$$\mathrm{Proj}_V \mathbf{Y} = d_2 \hat{\mathbf{X}}_2 + d_1 \hat{\mathbf{X}}_1 + d_0 \mathbf{1},$$

and rewrite this as

$$d_2 \left(\mathbf{X}_2 - \left(\frac{\mathbf{X}_2 \cdot \mathbf{1}}{\mathbf{1} \cdot \mathbf{1}} \right) \mathbf{1} - \left(\frac{\mathbf{X}_2 \cdot \hat{\mathbf{X}}_1}{\hat{\mathbf{X}}_1 \cdot \hat{\mathbf{X}}_1} \right) \left(\mathbf{X}_1 - \left(\frac{\mathbf{X}_1 \cdot \mathbf{1}}{\mathbf{1} \cdot \mathbf{1}} \right) \mathbf{1} \right) \right)$$
$$+ d_1 \left(\mathbf{X}_1 - \left(\frac{\mathbf{X}_1 \cdot \mathbf{1}}{\mathbf{1} \cdot \mathbf{1}} \right) \mathbf{1} \right) + d_0 \mathbf{1}.$$

Finally rewrite again in terms of the original (nonorthogonal) basis to find the coefficients of the best-fitting parabola:

$$\mathrm{Proj}_V \mathbf{Y} = d_2\left(\mathbf{X_2}\right) + \left(d_1 - d_2\left(\frac{\mathbf{X_2}\cdot\hat{\mathbf{X}}_1}{\hat{\mathbf{X}}_1\cdot\hat{\mathbf{X}}_1}\right)\right)\mathbf{X_1}$$
$$+\left(d_0 - d_1\left(\frac{\mathbf{X_1}\cdot\mathbf{1}}{\mathbf{1}\cdot\mathbf{1}}\right) - d_2\left(\frac{\mathbf{X_2}\cdot\mathbf{1}}{\mathbf{1}\cdot\mathbf{1}}\right) + d_2\left(\frac{\mathbf{X_2}\cdot\hat{\mathbf{X}}_1}{\hat{\mathbf{X}}_1\cdot\hat{\mathbf{X}}_1}\right)\left(\frac{\mathbf{X_1}\cdot\mathbf{1}}{\mathbf{1}\cdot\mathbf{1}}\right)\right)\mathbf{1},$$

so the coefficients of $\mathbf{X_2}$, $\mathbf{X_1}$, and $\mathbf{1}$ are a, b, and c, respectively.