# Homework 6

**This homework is due on Friday, October 6, 2022, at 11:59PM. Self-grades and HW Resubmissions are due on the following Friday, October 13, 2022, at 11:59PM.**

1. **Study Group Reassignment**

   We hope your study groups from the beginning of the semester have been going well! If you did not fill out the original matching form and would now like to join a group, or if your current study group is not meeting your needs, you can request a new study group via this form. Requests for new study groups are due Friday at 11:59 PM.

**2. Inner Products**

(a) For the following inner product defined on $\mathbb{R}^2$, which inner product properties hold?

$$\langle \vec{x}, \vec{y} \rangle = \begin{bmatrix} 2 \\ 1 \end{bmatrix}^\top (3\vec{x} + 3\vec{y}) \tag{1}$$

i. Symmetry (True/False)

ii. Linearity (True/False)

iii. Positive-Definiteness (True/False)

Explain your answers.

**Solution:**

i. Symmetry TRUE. Note that we have:

$$\langle \vec{x}, \vec{y} \rangle = \begin{bmatrix} 2 \\ 1 \end{bmatrix}^\top (3\vec{x} + 3\vec{y}) = \begin{bmatrix} 2 \\ 1 \end{bmatrix}^\top (3\vec{y} + 3\vec{x}) = \langle \vec{y}, \vec{x} \rangle \tag{2}$$

so $\langle \vec{x}, \vec{y} \rangle = \langle \vec{y}, \vec{x} \rangle$ and symmetry holds.

ii. Linearity FALSE. Let $c \in \mathbb{R}$:

$$\langle c\vec{x}, \vec{y} \rangle = \begin{bmatrix} 2 \\ 1 \end{bmatrix}^\top (3c\vec{x} + 3\vec{y}) \tag{3}$$

However,

$$c \langle \vec{x}, \vec{y} \rangle = \begin{bmatrix} 2 \\ 1 \end{bmatrix}^\top (3c\vec{x} + 3c\vec{y}) \tag{4}$$

Therefore, $\langle c\vec{x}, \vec{y} \rangle \neq c \langle \vec{x}, \vec{y} \rangle$ and linearity does not hold.

iii. Positive-Definiteness FALSE. Let $x = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$. We have:

$$\langle \vec{x}, \vec{x} \rangle = \begin{bmatrix} 2 \\ 1 \end{bmatrix}^\top \left( 3 \begin{bmatrix} -1 \\ -1 \end{bmatrix} + 3 \begin{bmatrix} -1 \\ -1 \end{bmatrix} \right) = \begin{bmatrix} 2 \\ 1 \end{bmatrix}^\top \begin{bmatrix} -6 \\ -6 \end{bmatrix} = -18 \tag{5}$$

Therefore, $\begin{bmatrix} \vec{x} \end{bmatrix} \vec{x} < 0$ and positive-definiteness does not hold.

(b) Consider the following valid inner product over the vector space of 2x2 real matrices $\mathbb{R}^{2\times 2}$, defined as

$$\langle A, B \rangle = \left\langle \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \right\rangle = a_{11}b_{11} + a_{12}b_{12} + a_{21}b_{21} + a_{22}b_{22} \tag{6}$$

for any $A, B \in \mathbb{R}^{2\times 2}$. Given this inner product definiton, what is $\left\| \begin{bmatrix} 2 & 5 \\ 6 & 2 \end{bmatrix} \right\|^2$?

**Solution:** Remember that for any inner product space, $\|A\|^2 = \langle A, A \rangle$. Knowing that, we have:

$$\left\| \begin{bmatrix} 2 & 5 \\ 6 & 2 \end{bmatrix} \right\|^2 = \left\langle \begin{bmatrix} 2 & 5 \\ 6 & 2 \end{bmatrix}, \begin{bmatrix} 2 & 5 \\ 6 & 2 \end{bmatrix} \right\rangle = 2^2 + 5^2 + 6^2 + 2^2 = 69 \tag{7}$$

**3. Least Squares with Shazam**

(a) The application Shazam is able to detect what song is playing by means of an *acoustic footprint*. This is a small set of information that identifies the song. Shazam then checks that footprint in its database, to check for another song that has that footprint. Here is the footprint we obtained via sampling: (we are representing the footprint as a vector)

$$\vec{x}_{\text{sample}} = \begin{bmatrix} 2 \\ 0 \\ -1 \\ 1 \end{bmatrix} \tag{8}$$

Say Shazam has narrowed it down to the following three songs with the corresponding footprints:

"Electric Love - Børns": $\vec{x}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}$

"She's Electric - Oasis": $\vec{x}_2 = \begin{bmatrix} 2 \\ -2 \\ -8 \\ 7 \end{bmatrix}$

"Electric Feel - MGMT": $\vec{x}_3 = \begin{bmatrix} 4 \\ 1 \\ -2 \\ 2 \end{bmatrix}$

Shazam is going to determine which song it is by projecting the footprint of our sample onto each of the song candidates, and ranking the songs based on the normalized inner product of $\vec{x}_{\text{sample}}$ onto the footprints. **Based on this information, which song is playing?**

**Solution:** Correct Answer: Electric Feel. The formula for projection of $\vec{u}$ onto $\vec{v}$ is $\text{proj}_{\vec{v}}(\vec{u}) = \frac{\vec{u}\cdot\vec{v}}{\|\vec{v}\|^2}\vec{v}$, but we only care about the magnitude of this vector, in other words, $\frac{\vec{u}\cdot\vec{v}}{\|\vec{v}\|}$. Magnitudes of projection are 0, $\frac{19}{11}$, $\frac{12}{5}$ respectively, the song which has the projection of largest magnitude is the one that is the most similar.

(b) Shazam wants to partner with Spotify. For its Discover Weekly algorithms, Spotify would like to know what characteristics of a song make it attractive to a first-time listener. Shazam provides the number of Shazams for a set of new songs to Spotify, who combines it with their data on the songs. Here is the table of data that Spotify assembles:

| Shazam Popularity | Tempo | Danceability | Acoustic-ness |
|---|---|---|---|
| 1109 | 100 | 0.8 | 0.6 |
| 5501 | 90 | 0.5 | 0.2 |
| 2031 | 68 | 0.4 | 0.7 |
| 13045 | 120 | 0.9 | 0.2 |

Spotify would now like to use the data it has to predict what the number of shazams for a new song, whose characteristics (Tempo, Danceability, Acoustic-ness) are represented as $\vec{a}_n$. **Which**

**is the correct formula for how Spotify would use Least Squares to calculate this?** Let $M$ be the matrix of Tempo, Danceabilty and Acoustic-ness, and $\vec{b}_n$ be the number of shazams they get:

$$M = \begin{bmatrix} 100 & 90 & 68 & 120 \\ 0.8 & 0.5 & 0.4 & 0.9 \\ 0.6 & 0.2 & 0.7 & 0.2 \end{bmatrix} \tag{9}$$

and

$$\vec{b}_n = \begin{bmatrix} 1109 \\ 5501 \\ 2031 \\ 13045 \end{bmatrix} \tag{10}$$

Options:

i. $(MM^\top)^{-1}M\vec{b}_n$

ii. $\vec{a_n}^\top(MM^\top)^{-1}M\vec{b}_n$

iii. $\vec{a_n}^\top(\vec{b}_n^\top M^\top)^\top$

iv. $\vec{b_n}^\top M\vec{a}_n$

v. $\vec{b_n}^\top M\vec{a}_n$

**Solution:** $\vec{a}_n^\top\left(MM^\top\right)^{-1}M\vec{b}_n$. To find the weights we do $\vec{x}_{\text{approx}} = \left(MM^\top\right)^{-1}M\vec{b}_n$. This boils down to finding the weights that minimize error, giving us the best predictor possible. We then want to run our new data through the model we just made, so we do: $\vec{a}_n^\top\vec{x}_{\text{approx}}$to get our final answer. $\left(A^\top A\right)^{-1}A^\top\vec{b}_n$ is the way we do least squares, but in this case, the $A$ matrix we want to plug into the expression is $M^\top$. This is because $M^\top$ has the data for each characteristic in the columns (each characteristic is one column). Remember that in least squares, we want to find the way in which we should combine the columns in order to best approximate the $\vec{b}$ vector. If we use $M^\top$, we end up finding what linear combination of characteristics best approximates the number of shazams, which is exactly what we want to do - predict Shazams popularity based on the characteristics. $\vec{x}_{\text{approx}} = \left(MM^\top\right)^{-1}M\vec{b}_n$ is not the answer itself, it is just tells us how we should weight the characteristics - we need to run our new data through the model we just made.

(c) Say Spotify gets some new data to incorporate into its data set, the energy of the song. Here is the table with the added data:

| Shazam Popularity | Tempo | Danceability | Acoustic-ness | Energy |
|---|---|---|---|---|
| 1109 | 100 | 0.8 | 0.6 | 0.70 |
| 5501 | 90 | 0.5 | 0.2 | 0.35 |
| 2031 | 68 | 0.4 | 0.7 | 0.55 |
| 13045 | 120 | 0.9 | 0.2 | 0.55 |

**Will it still be possible to run least squares with all of this data?**

**Solution:** The Energy column is linearly dependant on the dancabilty and acoustic-ness - it is their average. This means that the columns of our least squares matrix would be dependant, and this means that least squares cannot work - this is because we have to calculate $\left(A^\top A\right)^{-1}$. If the columns of $A$ are linearly dependant, then $A^\top A$ will not be invertable, breaking our least squares formula. For a more intuitive reason, consider that least squares attempts to answer the question - how do I best linearly combine the columns of $A$ so as to best approximate $\vec{b}$ If the columns of $A$ are dependant, then there would be more than one way to linearly combine the column vectors of $A$ to arrive at the best estimate of $\vec{b}$.

(d) What is the maximum number of features we could have per song, assuming we keep the number of songs the same?

Options:

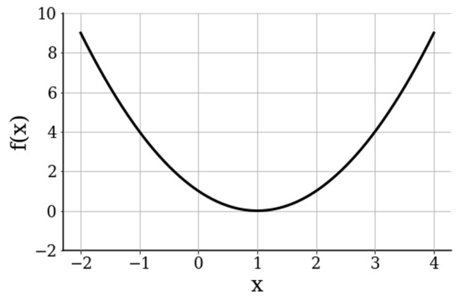
i. 4
ii. 3
iii. 2
iv. 5
v. As many as we want

**Solution:** The maximum number of features is 4. This is because in order for least squares to work, we need to have linearly independent columns, which are of dimension of 4. One can have, at maximum, a set of $n$ vectors in $\mathbb{R}^n$ that are all linearly independent, so the maximum number of features is 4.

**4. Orthogonality**

$f(x)$ and $g(x)$ are both polynomials with degree at most 2. $f(x) = x^2 - 2x + 1$. We define inner product between two polynomials as $\langle f(x), g(x)\rangle = f(0)g(0) + f(1)g(1) + f(2)g(2)$.

If $g(x)$ is orthogonal to $f(x)$, which is possible equation for $g(x)$?




(a) $-x^2 + 2x - 1$

(b) $x^2 + x - 1$

(c) $x - 1$

(d) $x$

**Solution:** Orthogonality means the inner product is zero, namely $\langle f(x), g(x)\rangle = 0$.

(a) $\langle f(x), g(x)\rangle = 1 \times (-1) + 0 \times 0 + 1 \times (-1) = -2$.

(b) $\langle f(x), g(x)\rangle = 1 \times (-1) + 0 \times 1 + 1 \times 5 = 4$.

(c) $\langle f(x), g(x)\rangle = 1 \times (-1) + 0 \times 0 + 1 \times 1 = 0$. $g(x)$ is orthogonal to $f(x)$.

(d) $\langle f(x), g(x)\rangle = 1 \times 0 + 0 \times 1 + 1 \times 2 = 2$.

**5. A Quirky Quantum Question**

(a) In quantum mechanics, states of particles are represented by vectors in a vector space. In this problem, we'll say that all states exist in $\mathbb{R}^2$.

A particular matrix, $\widehat{\mathbf{H}}$ (called the Hamiltonian operator), has the unique property that its eigenvalues represent a particle's allowed energy values. Quantum mechanics tells us that if the values of $\widehat{\mathbf{H}}$ are real, it must be symmetric – that is, it can be written as

$$\widehat{\mathbf{H}} = \begin{bmatrix} a & b \\ b & a \end{bmatrix} \tag{11}$$

**Assume that we know $a > 0$ and $b > 0$. What further condition on $a$ and $b$ forces the allowed energy values (the eigenvalues) to always be nonnegative?**

**Solution:** We must find the eigenvalues of $\hat{\mathbf{H}}$ and determine a condition that forces them to be nonnegative.

$$\widehat{\mathbf{H}}\vec{x} = \lambda\vec{x} \tag{12}$$

$$\widehat{\mathbf{H}}\vec{x} - \lambda\mathbf{I}\vec{x} = \vec{0} \tag{13}$$

$$\left(\widehat{\mathbf{H}} - \lambda\mathbf{I}\right)\vec{x} = \vec{0} \tag{14}$$

$$\begin{vmatrix} (a-\lambda) & b \\ b & (a-\lambda) \end{vmatrix} = 0 \tag{15}$$

$$(a-\lambda)^2 - b^2 = 0 \tag{16}$$

$$(a-\lambda) = \pm b \tag{17}$$

$$\lambda = a \pm b \tag{18}$$

Since $a$ and $b$ are both positive, $a+b$ will never produce a negative eigenvalue. However, in order for $a-b$ to be nonnegative, $a \geq b$.

(b) **Miki experimentally determines that particles associated with the $\widehat{\mathbf{H}}$ matrix from Question 1 have allowed energy values $\lambda_1 = \frac{5}{2}$ and $\lambda_2 = \frac{9}{2}$. Find $a$ and $b$.**

**Solution:** We know from the first question that $\lambda = a \pm b$. Since $\lambda_2 > \lambda_1$ and $a$ and $b$ are both positive, $\lambda_2 = a+b$ and $\lambda_1 = a-b$. Thus, we have two equations and two unknowns:

$$a - \lambda_1 = b \text{ and } \lambda_2 - a = b \tag{19}$$

$$a - \lambda_1 = \lambda_2 - a \tag{20}$$

$$a = \frac{1}{2}(\lambda_2 + \lambda_1) \tag{21}$$

$$a = \frac{1}{2}(\frac{9}{2} + \frac{5}{2}) \tag{22}$$

$$a = \frac{7}{2} = 3.5 \tag{23}$$

Likewise, we can use the same process to solve for b:

$$\lambda_1 + b = a \text{ and } \lambda_2 - b = a \tag{24}$$

$$\lambda_1 + b = \lambda_2 - b \tag{25}$$

$$b = \frac{1}{2}(\lambda_2 - \lambda_1) \tag{26}$$
$$b = \frac{1}{2}(\frac{9}{2} - \frac{5}{2}) \tag{27}$$
$$b = 1 \tag{28}$$

(c) Now, given a new matrix $\widehat{\mathbf{H}} = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$, let the eigenvalues be $\lambda_1 < \lambda_2$ and the normalized eigenvectors be $\vec{v}_{\lambda_1}$ and $\vec{v}_{\lambda_2}$ (corresponding to eigenvalues $\lambda_1$ and $\lambda_2$, respectively, and scaled to a magnitude of 1), span $\mathbb{R}^2$. If a particle is in some state $\vec{v}_s \in \mathbb{R}^2$, then it can be expressed as $\vec{v}_s = \alpha\vec{v}_{\lambda_1} + \beta\vec{v}_{\lambda_2}$, where $\alpha$ and $\beta$ are real constants. If $v_s = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, what are possible magnitudes of $\alpha$? (In quantum mechanics, $\alpha^2$ represents the probability that measuring the particle's energy will yield $\lambda_1$.)

**Solution:** The first step is to find the eigenvectors of $\widehat{\mathbf{H}}$. Based on previous questions, we know that $\lambda_1 = a - b = 3 - 2 = 1$, and $\lambda_2 = a + b = 3 + 2 = 5$. Let's define $\vec{v}_{\lambda_1} = \begin{bmatrix} c_1 \\ d_1 \end{bmatrix}$ and $\vec{v}_{\lambda_1} = \begin{bmatrix} c_2 \\ d_2 \end{bmatrix}$. We can now produce conditions on $c_1$, $d_1$, $c_2$, and $d_2$ to find the eigenvectors:

$$\widehat{\mathbf{H}}\vec{v}_{\lambda_1} = \lambda_1\vec{v}_{\lambda_1} \tag{29}$$
$$\begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} c_1 \\ d_1 \end{bmatrix} = \begin{bmatrix} c_1 \\ d_1 \end{bmatrix} \tag{30}$$
$$\begin{bmatrix} 3c_1 + 2d_1 \\ 2c_1 + 3d_1 \end{bmatrix} = \begin{bmatrix} c_1 \\ d_1 \end{bmatrix} \tag{31}$$

We can take the first row in isolation to produce

$$3c_1 + 2d_1 = c_1 \tag{32}$$
$$c_1 = -d_1 \tag{33}$$

As such, we know that the first eigenvector will have the form $\vec{v}_{\lambda_1} = k_1 \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ for some normalization constant $k_1$. Since we're told that $\left\|\vec{v_{\lambda_1}}\right\| = 1$, $k_1$ must equal $\frac{1}{\sqrt{2}}$. Thus, $\vec{v}_{\lambda_1} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

Now, we can produce another condition using the second eigenvalue:

$$\widehat{\mathbf{H}}\vec{v}_{\lambda_2} = \lambda_2\vec{v}_{\lambda_2} \tag{34}$$
$$\begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} c_2 \\ d_2 \end{bmatrix} = 5\begin{bmatrix} c_2 \\ d_2 \end{bmatrix} \tag{35}$$
$$\begin{bmatrix} 3c_2 + 2d_2 \\ 2c_2 + 3d_2 \end{bmatrix} = 5\begin{bmatrix} c_2 \\ d_2 \end{bmatrix} \tag{36}$$

We can once again take the first row in isolation to produce

$$3c_2 + 2d_2 = 5c_2 \tag{37}$$

$$c_2 = d_2 \tag{38}$$

As such, we know that the second eigenvector will have the form $\vec{v}_{\lambda_2} = k_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ for some normalization constant $k_2$. Since we're told that $\|\vec{v}_{\lambda_2}\| = 1$, $k_2$ must equal $\frac{1}{\sqrt{2}}$. Thus, $\vec{v}_{\lambda_2} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

The last step is to find $\alpha$. Based on the problem statement,

$$\alpha \vec{v}_{\lambda_1} + \beta \vec{v}_{\lambda_2} = \vec{v}_s \tag{39}$$

$$\alpha \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \beta \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{40}$$

$$\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{41}$$

Since we know that $\vec{v}_{\lambda_1}$ and $\vec{v}_{\lambda_2}$ are linearly independent, this equation has a unique solution. $\alpha$ and $\beta$ may be found through row reduction or inversion, but by inspection we can see that $\alpha = -\frac{1}{\sqrt{2}}$ and $\beta = \frac{1}{\sqrt{2}}$. Thus, $|\alpha| = \frac{1}{\sqrt{2}}$.

**6. Discrete System Induction**

Consider the discrete system given by

$$\vec{x}[i+1] = A\vec{x}[i] + B\vec{u}[i] \tag{42}$$

(a) Write $\vec{x}[1]$ in terms of $\vec{x}[0]$ and $\vec{u}[0]$. Then, write $\vec{x}[2]$ in terms of $\vec{x}[0]$, $\vec{u}[1]$, and $\vec{u}[0]$. Lastly, write $\vec{x}[3]$ in terms of $\vec{x}[0]$, $\vec{u}[2]$, $\vec{u}[1]$, and $\vec{u}[0]$.

**Solution:** Firstly, we have

$$\vec{x}[1] = A\vec{x}[0] + B\vec{u}[0] \tag{43}$$

Then, we have

$$\begin{aligned}\vec{x}[2] &= A\vec{x}[1] + B\vec{u}[1] &(44)\\ &= A(A\vec{x}[0] + B\vec{u}[0]) + B\vec{u}[1] &(45)\\ &= A^2\vec{x}[0] + AB\vec{u}[0] + B\vec{u}[1] &(46)\end{aligned}$$

And lastly,

$$\begin{aligned}\vec{x}[3] &= A\vec{x}[2] + B\vec{u}[2] &(47)\\ &= A\left(A^2\vec{x}[0] + AB\vec{u}[0] + B\vec{u}[1]\right) + B\vec{u}[2] &(48)\\ &= A^3\vec{x}[0] + A^2B\vec{u}[0] + AB\vec{u}[1] + B\vec{u}[2] &(49)\end{aligned}$$

(b) We can generalize this to write $\vec{x}[i]$ in terms of $\vec{x}[0]$ and $\vec{u}[i-1], \ldots, \vec{u}[0]$ as follows:

$$\vec{x}[i] = A^i\vec{x}[0] + \sum_{j=0}^{i-1} A^{i-1-j}B\vec{u}[j] \tag{50}$$

**Verify that this equation holds for $i = 1$.** *This is equivalent to testing your base case in induction.*

**Solution:** Plugging in $i = 1$, we have $\vec{x}[1] = A\vec{x}[0] + \sum_{j=0}^{0} A^{i-1-j}B\vec{u}[j] = A\vec{x}[0] + B\vec{u}[0]$

(c) **Show that eq. (50) holds for $\vec{x}[i+1]$.** That is, write $\vec{x}[i+1]$ in terms of $\vec{x}[i]$ and plug in eq. (50) for $\vec{x}[i]$. Show that this simplifies to eq. (50) where we now replace $i$ with $i+1$. *This is equivalent to testing your inductive hypothesis.*

**Solution:** We have that

$$\begin{aligned}\vec{x}[i+1] &= A\vec{x}[i] + B\vec{u}[i] &(51)\\ &= A\left(A^i\vec{x}[0] + \sum_{j=0}^{i-1} A^{i-1-j}B\vec{u}[j]\right) + B\vec{u}[i] &(52)\\ &= A^{i+1}\vec{x}[0] + \sum_{j=0}^{i-1} A^{i-j}B\vec{u}[j] + A^{i-i}B\vec{u}[i] &(53)\\ &= A^{i+1}\vec{x}[0] + \sum_{j=0}^{i} A^{i-j}B\vec{u}[j] &(54)\end{aligned}$$

which is exactly eq. (50) where we plug in $i+1$ for $i$.