Questions

1. Orthonormality

   (a) Suppose you have a real orthonormal matrix $U$, i.e., the columns of $U$ are orthonormal, and you have real vectors $\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}_1, \mathbf{y}_2$ such that
   
   $\mathbf{y}_1 = U\mathbf{x}_1$
   $\mathbf{y}_2 = U\mathbf{x}_2$
   
   **Calculate** $\langle \mathbf{y}_1, \mathbf{y}_2 \rangle = \mathbf{y}_1^\top \mathbf{y}_2$ in terms of $\langle \mathbf{x}_1, \mathbf{x}_2 \rangle = \mathbf{x}_1^\top \mathbf{x}_2$.
   
   **Answer:** Note that
   
   $\langle \mathbf{y}_1, \mathbf{y}_2 \rangle = \mathbf{y}_1^\top \mathbf{y}_2 = \mathbf{x}_1^\top U^\top U \mathbf{x}_2 = \mathbf{x}_1^\top \mathbf{x}_2 = \langle \mathbf{x}_1, \mathbf{x}_2 \rangle$.

   (b) Following the previous question, express $\|\mathbf{y}_1\|_2^2$ and $\|\mathbf{y}_2\|_2^2$ in terms of $\|\mathbf{x}_1\|_2^2$ and $\|\mathbf{x}_2\|_2^2$.
   
   **Answer:** Note that
   
   $\langle \mathbf{y}_1, \mathbf{y}_1 \rangle = \mathbf{y}_1^\top \mathbf{y}_1 = \mathbf{x}_1^\top U^\top U \mathbf{x}_1 = \mathbf{x}_1^\top \mathbf{x}_1 = \|\mathbf{x}_1\|_2^2$, and
   
   $\langle \mathbf{y}_2, \mathbf{y}_2 \rangle = \mathbf{y}_2^\top \mathbf{y}_2 = \mathbf{x}_2^\top U^\top U \mathbf{x}_2 = \mathbf{x}_2^\top \mathbf{x}_2 = \|\mathbf{x}_2\|_2^2$.

   (c) Suppose you observe data coming from the model $y = \mathbf{a}^\top \mathbf{x}$. Suppose you have $m$ data points $(\mathbf{x}_i, y_i)$. Set up a least squares formulation for estimating $\mathbf{a}$ and find the solution to the least squares.
   
   **Answer:** If we line up the data points to have
   
   $X := \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_m^\top \end{bmatrix}$ and $\mathbf{y} := \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$
   
   we aim to solve
   
   $\hat{\mathbf{a}} = \min_{\mathbf{w}} \|\mathbf{y} - X\mathbf{w}\|_2^2$.
   
   The solution is then
   
   $\hat{\mathbf{a}} = (X^\top X)^{-1} X^\top \mathbf{y}$.

   (d) Now suppose $V$ is an orthonormal square matrix and we observe data points transformed by $V^\top$. In other words, we have observations
   
   $\tilde{\mathbf{x}} = V^\top \mathbf{x}$
   
   with the $y$’s unchanged. Recall that originally we had $y \approx \tilde{\mathbf{a}}^\top \tilde{\mathbf{x}}$. **Suppose we want to write the least squares in terms of $\tilde{\mathbf{x}}$ as $y \approx \tilde{\mathbf{a}}^\top \tilde{\mathbf{x}}$.** Set up a least squares formulation for estimating $\tilde{\mathbf{a}}$ from the $(\tilde{\mathbf{x}}, y)$ pairs and find the solution to the least squares.
**Answer:** We can write
\[ \mathbf{X} := \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_m^\top \end{bmatrix} \cdot V \]

The least squares formulation is
\[ \hat{a} = \min_w \| \mathbf{y} - \mathbf{X}w \|_2^2. \]

The solution is then
\[ \hat{a} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} = (V \Sigma U^\top)^{-1} V \Sigma U^\top \mathbf{y} = V (X^\top X)^{-1} V^\top X^\top \mathbf{y} = V \mathbf{y}. \]

(e) Now suppose that we have the matrix
\[ \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_m^\top \end{bmatrix} := \mathbf{X} = U \Sigma V^\top. \]

and suppose that the orthonormal matrix in the previous part is the \( V \) corresponding to this SVD representation. **Set up a least squares formulation for estimating \( \hat{a} \) and find the solution to the least squares. Is there anything interesting going on?**

**Answer:** From the previous part, we see that the solution is
\[
\hat{a} = V^\top (X^\top X)^{-1} X^\top \mathbf{y} = V^\top (V \Sigma U^\top U \Sigma V^\top)^{-1} V \Sigma U^\top \mathbf{y} = (\Sigma^\top \Sigma)^{-1} \Sigma^\top U^\top \mathbf{y} = \begin{bmatrix}
\frac{1}{\sigma_1} & 0 & \cdots & 0 & 0 \\
0 & \frac{1}{\sigma_2} & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & \cdots & \cdots & \frac{1}{\sigma_n} & 0 \\
0 & \cdots & \cdots & \cdots & 0
\end{bmatrix} \cdot U^\top \mathbf{y}.
\]

The matrix inverse term \((\Sigma^\top \Sigma)^{-1}\) then becomes a diagonal square matrix that can be easily invertible by inverting each of the diagonal coordinates.

2. **Geometric interpretation of the SVD**

In this exercise, we explore the geometric interpretation of symmetric matrices and how this connects to the SVD. We consider how a real \( 2 \times 2 \) matrix acts on the unit circle, transforming it into an ellipse. It turns out that the principal semiaxes of the resulting ellipse are related to the singular values of the matrix, as well as the vectors in the SVD.

(a) Consider the real \( 2 \times 2 \) matrix
\[ A = \begin{pmatrix} 0 & -1 \\ 3 & 0 \end{pmatrix}. \]
Now consider the unit circle in $\mathbb{R}^2$,

$$ S = \left\{ \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} \mid 0 \leq \theta < 2\pi \right\}. $$

**Plot $AS$ on the $\mathbb{R}^2$ plane.**

**Answer:**

$$ AS = \left\{ \begin{pmatrix} -\sin \theta \\ 3 \cos \theta \end{pmatrix} \mid 0 \leq \theta < 2\pi \right\}. $$

The plot should be the ellipse centered at the origin that passes through the points $(0, 3), (0, -3), (-1, 0), (1, 0)$.

(b) We can think of the unit circle as solutions to $\vec{x}^T \vec{x} = 1$ for a two-dimensional vector $\vec{x}$. **Verify that the unit-circle satisfies this equation.** Then, find a diagonal matrix $W$ such that $(AS)^T W (AS) = 1$.

Hint: The SVD of $A$ is

$$ A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. $$

**Answer:**

$$ \vec{x}^T \vec{x} = \begin{pmatrix} \cos \theta & \sin \theta \end{pmatrix} \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} = \cos^2 \theta + \sin^2 \theta = 1 $$

In order to find the diagonal matrix $W$, we let set it to be

$$ W = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}. $$

Then,

$$ (AS)^T W (AS) = \sin^2 \theta + 9b \cos^2 \theta. $$
We would like this to be \(\sin^2 \theta + \cos^2 \theta = 1\). Hence, we choose \(a = 1\) and \(b = \frac{1}{2}\). The corresponding matrix \(W\) is then
\[
W = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}.
\]

(c) Consider the columns of the matrices \(U, V\) obtained in the previous part, and treat them as vectors in \(\mathbb{R}^2\). Let \(U = (\vec{u}_1 \ \vec{u}_2)\), \(V = (\vec{v}_1 \ \vec{v}_2)\). Let \(\sigma_1, \sigma_2\) be the singular values of \(A\), where \(\sigma_1 \geq \sigma_2\).

**Draw in your plot of \(AS\) the vectors \(\sigma_1 \vec{u}_1\) and \(\sigma_2 \vec{u}_2\), drawn from the origin. What do these vectors correspond to geometrically?**

**Answer:** \(\sigma_1 \vec{u}_1 = (0, 3)\) corresponds to the semi-major axis of the ellipse, while \(\sigma_2 \vec{u}_2 = (-1, 0)\) corresponds to the semi-minor axis.

(d) **Repeat what you did above for the matrix** \(A = \begin{pmatrix} 2 & 1 \\ -2 & 1 \end{pmatrix}\).

Do you notice something?

**Hint:** The SVD of \(A\) is
\[
A = \begin{bmatrix} \sqrt{2} & \sqrt{2} \\ -\sqrt{2} & \sqrt{2} \end{bmatrix} \begin{bmatrix} 2 \sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.
\]

**Answer:** Here,
\[
AS = \left\{ \begin{pmatrix} 2 \cos \theta + \sin \theta \\ -2 \cos \theta + \sin \theta \end{pmatrix} | 0 \leq \theta < 2\pi \right\}.
\]

Then, as you should see through your calculations, we aren’t able to find a diagonal \(W\) matrix for \(AS\). Instead, we can find set a full \(W\) matrix and try to solve for \(W\). Letting
\[
W = \begin{bmatrix} a & b \\ c & d \end{bmatrix},
\]

Let’s calculate \((AS)^\top W(AS)\). Remember that we want this to be equal to 1 for all \(\theta\).

\[
(AS)^\top W(AS) = \begin{bmatrix} 2 \cos \theta + \sin \theta & -2 \cos \theta + \sin \theta \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 2 \cos \theta + \sin \theta \\ -2 \cos \theta + \sin \theta \end{bmatrix}
\]
\[
= \begin{bmatrix} (2a - 2c) \cos \theta + (a + c) \sin \theta & (2b - 2d) \cos \theta + (b + d) \sin \theta \end{bmatrix} \begin{bmatrix} 2 \cos \theta + \sin \theta \\ -2 \cos \theta + \sin \theta \end{bmatrix}
\]
\[
= (4a - 4b - 4c + 4d) \cos^2 \theta + (a + b + c + d) \sin^2 \theta + (4a - 4d) \sin \theta \cos \theta
\]

Next, we choose our parameters \(a, b, c, d\) to force this to become \(\cos^2 \theta + \sin^2 \theta = 1\).

\[
4a - 4b - 4c + 4d = 1
\]
\[
a + b + c + d = 1
\]
\[
4a - 4d = 0
\]
There are many pairs of \((a, b, c, d)\) that satisfy these equations. For example, if we want to find a symmetric matrix \(W\) that satisfies this, then we can choose
\[
\begin{pmatrix}
\frac{5}{16} & \frac{3}{16} \\
\frac{3}{16} & \frac{5}{16}
\end{pmatrix}.
\]
Note that
\[
A^\top A = \begin{bmatrix} 8 & 0 \\ 0 & 2 \end{bmatrix} \quad AA^\top = \begin{bmatrix} 5 & -3 \\ -3 & 5 \end{bmatrix}.
\]
each with eigenvalues \(\lambda_1 = 8, \lambda_2 = 2\).
Here, the resemblance of \(AA^\top\) to the \(W\) matrix above is uncanny. In fact, it is clear that \(W = (AA^\top)^{-1}\).
This fact demands an explanation. If we imagine \(S\) as being a matrix or vector in spirit, we can explore as follows \((AS)^T WAS = S^T A^T WAS = 1\). We know that \(S^T S = 1\) and so what we want is \(A^T WA = I\) the identity. Multiplying both sides on the left by the inverse of \(A^T\) gives \(WA = (A^T)^{-1}\) and multiplying both sides on the right by \(A^{-1}\) gives \(W = (A^T)^{-1}A^{-1} = (AA^\top)^{-1}\). This explains this seeming coincidence.
Returning to the given problem, the corresponding set of eigenvectors of \(A^\top A\) is
\[
\begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 \\ 1 \end{bmatrix}.
\]
Running these through \(A\) and normalizing, we get:
\[
\begin{bmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -\frac{\sqrt{2}}{\sqrt{2}^2} \\ \frac{\sqrt{2}}{\sqrt{2}^2} \end{bmatrix}.
\]
Notice that these are also eigenvectors of \(AA^\top\).
Anyway, the corresponding singular values are \(\sqrt{8} = 2\sqrt{2}\) and \(\sqrt{2}\).
The SVD decomposition for \(A\) is therefore:
\[
A = \begin{bmatrix}
\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\
-\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2}
\end{bmatrix} \begin{bmatrix}
2\sqrt{2} & 0 \\
0 & \sqrt{2}
\end{bmatrix} \begin{bmatrix}
1 \\ 0
\end{bmatrix}.
\]

The ellipse and corresponding points are plotted in the above graph.
Notice that the first vector points along the -45 degree line which corresponds to the major axis for the ellipse while the second is along the 45 degree line which is the minor axis for the ellipse. Once again, it is the $U$ matrix whose constituent vectors give these directions.

Why? Because we are viewing the matrix as acting on column vectors, and these are 2-dimensional vectors.

If you want to explore some more on your own, make a wide matrix filled with the output of $A\vec{x}$ for $\vec{x}$ that are drawn randomly from the set $S$ representing the unit circle. (You can create these by using numpy.random.uniform to draw a bunch of samples uniformly from 0 to $2\pi$, and then generate vectors using cosine and sine.) Then ask numpy to take the SVD of the resulting wide matrix. Look at the $U$ matrix that the SVD returns. You will see that it will be pretty close to what we got above. This reflects the power of the SVD to discover the underlying elliptical structure given a bunch of points. Given the evaluation of a matrix $A$ on a bunch of points drawn uniformly from around the unit circle (and in higher dimensions, from the unit hypersphere) it will reveal the major and minor axes for the relevant ellipsoid along with the shape of the ellipsoid through the singular values.

(e) Consider the case where $A$ is a real $n \times n$ symmetric matrix. What do you observe geometrically in this case?

Answer: In this case, $AA^T$ and $A^TA$ are equal and therefore have the same eigenvalues. Hence $U = V$, and geometrically the action of $A$ corresponds to scaling the unit sphere in $\mathbb{R}^n$ along the vectors $\vec{v}_i$ by a factor of $\sigma_i$ for each $i$ to get a hyperellipse.

$A\vec{x} = U\Sigma V^T\vec{x}$. First, $\vec{x}$ is projected onto the eigenvectors of $A^TA$, then the different dimensions are scaled by the singular values, and the vector is reconstituted through a linear combination of the same eigenvalues by applying $U$.

Contributors:

- Kuan-Yun Lee.
- Anant Sahai.
- Lynn Chua.
- Shane Barratt.