1. Give an example of a 2×2 matrix that has no real eigenvectors. Justify your solution with intuition (without solving completely for the eigenvectors and eigenvalues).

   Consider a Givens rotation matrix

$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

   with $\theta \neq n\pi$ for any integer n. This matrix will rotate any vector counterclockwise by θ radians. This means that no vector will be mapped to the same direction it was before, so the matrix has no real eigenvectors.

2. Consider an $n \times p$ matrix A. Show that the number of linear independent rows is the same as the number of linearly independent columns.
   Hint: Write $A = CR$ where C is a matrix of the linearly independent columns of A. Why can we write A like this? Then consider the CR product in the "row" interpretation of matrix multiplication.

   Suppose there are m linearly independent columns of A. Place these columns in an $n \times m$ matrix C. Then A can be written as $A = CR$ where R is an $m \times p$ matrix. We can also think of the rows of A being linear combinations of the rows of R, so the number of linearly independent rows is at most m. Thus, the number of linearly independent rows is less than or equal to the number of linearly independent columns!

   But now consider the matrix A^T. By the same argument, the number of linearly independent rows of A^T is less than the number of linearly independent columns of A^T. But the number of transpose just switches rows and columns. Thus, the number of linearly independent rows and columns in A must be equal.

3. Let A be an $m \times n$ matrix (assume $m > n$). The full singular value factorization $A = U\Sigma V^{\mathrm{T}}$ contains more information than necessary to reconstruct A.

   (a) What are the smallest matrices $\tilde{U}$, $\tilde{\Sigma}$ and $\tilde{V}^{\mathrm{T}}$ such that $\tilde{U}\tilde{\Sigma}\tilde{V}^{\mathrm{T}} = A$?

   This is called the *reduced* SVD. The dimensions are: $\tilde{U}$ is $m \times n$, $\tilde{V}$ is $n \times n$, and $\tilde{\Sigma}$ is $n \times n$. Since all the entries below n in Σ are zero, the entries below n in ΣV^T are all zero, and so do not factor into the multiplication by U. Thus, the columns to the *right* of n in U are not needed, and the rows *below* n in Σ are not needed. All of V^T is needed.

   (b) Let $U = \begin{bmatrix} \tilde{U} & \hat{U} \end{bmatrix}$. That is, think about U from the full singular value factorization as a block matrix consisting of the matrix $\tilde{U}$ found in part (a) and the remaining (unneeded) columns $\hat{U}$.

Find expressions for $\tilde{U}^{\mathrm{T}}\tilde{U}$ and $\tilde{U}\tilde{U}^{\mathrm{T}}$.

Recall that $U^TU = I$, so

$$U^TU = \begin{bmatrix} \tilde{U}^T \\ \hat{U}^T \end{bmatrix} \begin{bmatrix} \tilde{U} & \hat{U} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$$

From these we can see that $\tilde{U}^T\tilde{U} = I$ (due to the upper left entry of U^TU). Now consider UU^T:

$$\begin{aligned} UU^T = I = \begin{bmatrix} \tilde{U} & \hat{U} \end{bmatrix} \begin{bmatrix} \tilde{U}^T \\ \hat{U}^T \end{bmatrix} &= \tilde{U}\tilde{U}^T + \hat{U}\hat{U}^T \\ \implies \tilde{U}\tilde{U}^T &= I - \hat{U}\hat{U}^T \end{aligned}$$

Unfortunately, we cannot say much more about $\tilde{U}\tilde{U}^T$. We no longer have the nice result that $U^TU = UU^T = I$. This is a disadvantage of the reduced SVD.

(c) Use the *reduced* singular value factorization obtained in part (a) to find an expression for the matrix $H = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$. How many matrices must be inverted (diagonal and orthogonal matrices don't count)?

$$\begin{aligned} H &= A(A^TA)^{-1}A^T \\ &= (\tilde{U}\tilde{\Sigma}\tilde{V}^T)((\tilde{U}\tilde{\Sigma}\tilde{V}^T)^T(\tilde{U}\tilde{\Sigma}\tilde{V}^T))^{-1}(\tilde{U}\tilde{\Sigma}\tilde{V}^T)^T \\ &= \tilde{U}\tilde{\Sigma}\tilde{V}^T(\tilde{V}\tilde{\Sigma}\tilde{U}^T\tilde{U}\tilde{\Sigma}\tilde{V}^T)^{-1}\tilde{V}\tilde{\Sigma}\tilde{U}^T \\ &= \tilde{U}\tilde{\Sigma}\tilde{V}^T(\tilde{V}\tilde{\Sigma}^2\tilde{V}^T)^{-1}\tilde{V}\tilde{\Sigma}\tilde{U}^T \\ &= \tilde{U}\tilde{\Sigma}(\tilde{V}^T\tilde{V})\tilde{\Sigma}^{-2}(\tilde{V}^T\tilde{V})\tilde{\Sigma}\tilde{U}^T \\ &= \tilde{U}(\tilde{\Sigma}\tilde{\Sigma}^{-2}\tilde{\Sigma})\tilde{U}^T \\ &= \tilde{U}\tilde{U}^T \end{aligned}$$

No matrices needed to be inverted to compute H. Only a single matrix product must be computed.

4. Let x and y be vectors of m elements. The least squares solution for a best-fit line for a plot of y versus x is

$$\hat{\beta} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y$$

where

$$X = \begin{bmatrix} | & | \\ 1 & x \\ | & | \end{bmatrix}$$

(a) Suppose you know the **full** singular value factorization $X = U\Sigma V^{\mathrm{T}}$. Find an expression for $\hat{\beta}$ in terms of U, Σ, and V. Hint: Only square matrices can be invertible.

Begin by substituting in the SVD:

$$\begin{aligned}
\hat{\beta} &= ((U\Sigma V^T)^T U\Sigma V^T)^{-1}(U\Sigma V^T)^T y \\
&= (V\Sigma^T (U^T U)\Sigma V^T)^{-1} V\Sigma^T U^T y \\
&= (V\Sigma^T \Sigma V^T)^{-1} V\Sigma^T U^T y \\
&= V(\Sigma^T\Sigma)^{-1} V^T V \Sigma^T U^T y \\
&= V(\Sigma^T\Sigma)^{-1}\Sigma^T U^T y
\end{aligned}$$

Consider breaking Σ into two pieces by

$$\Sigma = \begin{bmatrix} \tilde{\Sigma} \\ 0 \end{bmatrix}$$

where $\tilde{\Sigma}$ is an $n \times n$ diagonal matrix. Then,

$$\Sigma^T\Sigma = \begin{bmatrix} \tilde{\Sigma}^T & 0 \end{bmatrix} \begin{bmatrix} \tilde{\Sigma} \\ 0 \end{bmatrix} = \tilde{\Sigma}^T\tilde{\Sigma} = \tilde{\Sigma}^2$$

where $\tilde{\Sigma}^2$ is an $n \times n$ diagonal matrix with the diagonal equal to the squared value in Σ. Its inverse is found by taking the reciprocal of each of these. Therefore,

$$\begin{aligned}
\hat{\beta} &= V(\Sigma^T\Sigma)^{-1}\Sigma^T U^T y \\
&= V\tilde{\Sigma}^{-2}\begin{bmatrix} \tilde{\Sigma}^T & 0 \end{bmatrix} U^T y \\
&= V\begin{bmatrix} \tilde{\Sigma}^{-1} & 0 \end{bmatrix} U^T y
\end{aligned}$$

(b) Repeat part (a) using the reduced singular value factorization $X = \tilde{U}\tilde{\Sigma}\tilde{V}^{\mathrm{T}}$.

This one proceeds similarly, except now every matrix is square, which simplifies things:

$$\begin{aligned}
\hat{\beta} &= ((\tilde{U}\tilde{\Sigma}\tilde{V}^T)^T \tilde{U}\tilde{\Sigma}\tilde{V}^T)^{-1}(\tilde{U}\tilde{\Sigma}\tilde{V}^T)^T y \\
&= (\tilde{V}\tilde{\Sigma}(\tilde{U}^T\tilde{U})\tilde{\Sigma}\tilde{V}^T)^{-1}\tilde{V}\tilde{\Sigma}\tilde{U}^T y \\
&= (\tilde{V}\tilde{\Sigma}^2\tilde{V}^T)^{-1}\tilde{V}\tilde{\Sigma}\tilde{U}^T y \\
&= \tilde{V}\tilde{\Sigma}^{-2}(\tilde{V}^T\tilde{V})\tilde{\Sigma}\tilde{U}^T y \\
&= \tilde{V}\tilde{\Sigma}^{-1}(\tilde{\Sigma}^{-1}\tilde{\Sigma})\tilde{U}^T y \\
&= \tilde{V}\tilde{\Sigma}^{-1}\tilde{U}^T y
\end{aligned}$$

5. Let $\tilde{X}$ be an $m \times n$ matrix $(m > n)$ whose columns have sample mean zero, and let $\tilde{X} = \tilde{U}\tilde{\Sigma}\tilde{V}^{\mathrm{T}}$ be a reduced singular value factorization of $\tilde{X}$. The squared *Mahalanobis* distance to the point $\tilde{x}_i^{\mathrm{T}}$ (the i^{th} row of $\tilde{X}$) is

$$d_i^2 = \tilde{x}_i^{\mathrm{T}}\hat{S}^{-1}\tilde{x}_i$$

where $\hat{S} = \frac{1}{m-1}\tilde{X}^{\mathrm{T}}\tilde{X} = \operatorname{cov}(\tilde{X})$. Explain how to compute d_i^2 without inverting a matrix.

Using the reduced SVD,

$$\begin{aligned}
d_i^2 &= \tilde{x}_i^T \left(\frac{1}{m-1}\tilde{X}^T\tilde{X}\right)^{-1} \tilde{x}_i \\
&= (m-1)\tilde{x}_i^T \left((\tilde{U}\tilde{\Sigma}\tilde{V}^T)^T\tilde{U}\tilde{\Sigma}\tilde{V}^T\right)^{-1} \tilde{x}_i \\
&= (m-1)\tilde{x}_i^T \left(\tilde{V}\tilde{\Sigma}(\tilde{U}^T\tilde{U})\tilde{\Sigma}\tilde{V}^T\right)^{-1} \tilde{x}_i \\
&= (m-1)\tilde{x}_i^T \left(\tilde{V}\tilde{\Sigma}^2\tilde{V}^T\right)^{-1} \tilde{x}_i \\
&= (m-1)\tilde{x}_i^T\tilde{V}\tilde{\Sigma}^{-2}\tilde{V}^T\tilde{x}_i \\
&= (m-1)\left(\tilde{V}^T\tilde{x}_i\right)^T\tilde{\Sigma}^{-2}\left(\tilde{V}^T\tilde{x}_i\right)
\end{aligned}$$

Now, no matrices need to be inverted at all.

Alternatively, we can recognize that $\tilde{x}_i^T$ represents the i^{th} row of $\tilde{X}$. Thus, by the row definition of matrix multiplication, $\tilde{x}_i^T = \tilde{U}_i\tilde{\Sigma}\tilde{V}^T$ where $\tilde{U}_i$ is the i^{th} row of $\tilde{U}$. Substituting this in,

$$\begin{aligned}
d_i^2 =&(m-1)\tilde{x}_i^T\tilde{V}\tilde{\Sigma}^{-2}\tilde{V}^T\tilde{x}_i \\
=&(m-1)(\tilde{U}_i\tilde{\Sigma}\tilde{V}^T)\tilde{V}\tilde{\Sigma}^{-2}\tilde{V}^T(\tilde{U}_i\tilde{\Sigma}\tilde{V}^T)^T \\
=&(m-1)\tilde{U}_i\tilde{\Sigma}(\tilde{V}^T\tilde{V})\tilde{\Sigma}^{-2}(\tilde{V}^T\tilde{V})\tilde{\Sigma}\tilde{U}_i^T \\
=&(m-1)\tilde{U}_i(\tilde{\Sigma}\tilde{\Sigma}^{-2}\tilde{\Sigma})\tilde{U}_i^T \\
=&(m-1)\tilde{U}_i\tilde{U}_i^T.
\end{aligned}$$

Thus, the squared Mahalanobis distance is the dot product between the i^{th} row of $\tilde{U}$ and itself. If $\tilde{U}$ was still an orthonormal matrix, this would be one. However, as you saw in problem 3(c), this is not the case. There are no inverses that must be taken to compute this, however.

6. (a) Suppose $A = LU$ where L is lower triangular and U is upper triangular. Explain how you would solve the problem $Ax = b$ using L, U, and the concepts of forward and backward substitution.

   We first replace A by LU:

$$\begin{aligned}
Ax =&b \\
LUx =&b.
\end{aligned}$$

   Now, define $y = Ux$ as another vector. Then, we can solve for y in $Ly = b$ with forward substitution. With this in hand, we could solve for x in $Ux = y$ using back substitution.

   (b) Compute the LU factorization of

$$A = \begin{bmatrix} 1 & 2 & -1 \\ 3 & -3 & 2 \\ -2 & 1 & 1 \end{bmatrix}$$

by hand using elimination matrices.

$$
\begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}
\begin{bmatrix} 1 & 2 & -1 \\ 3 & -3 & 2 \\ -2 & 1 & 1 \end{bmatrix}
=
\begin{bmatrix} 1 & 2 & -1 \\ 0 & -9 & 5 \\ -2 & 1 & 1 \end{bmatrix}
$$

$$
\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}
\begin{bmatrix} 1 & 2 & -1 \\ 0 & -9 & 5 \\ -2 & 1 & 1 \end{bmatrix}
=
\begin{bmatrix} 1 & 2 & -1 \\ 0 & -9 & 5 \\ 0 & 5 & -1 \end{bmatrix}
$$

$$
\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 5/9 & 1 \end{bmatrix}
\begin{bmatrix} 1 & 2 & -1 \\ 0 & -9 & 5 \\ 0 & 5 & -1 \end{bmatrix}
=
\begin{bmatrix} 1 & 2 & -1 \\ 0 & -9 & 5 \\ 0 & 0 & 16/9 \end{bmatrix}
$$

Let's write the three elimination matrices out to the left of A. We can multiply on the left by the inverses of each one after another to leave only A on the left hand side. The inverse of an elimination matrix (i.e. one with zeros above the diagonal, ones along the diagonal, and only one nonzero below the diagonal) is found by multiplying the off-diagonal entry by -1. After doing this to each side, on the right hand side, we'll have a lower triangular matrix multiplied by an upper triangular one, as desired!

$$
\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}
\begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}
\begin{bmatrix} 1 & 2 & -1 \\ 3 & -3 & 2 \\ -2 & 1 & 1 \end{bmatrix}
=
\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -5/9 & 1 \end{bmatrix}
\begin{bmatrix} 1 & 2 & -1 \\ 0 & -9 & 5 \\ 0 & 0 & 16/9 \end{bmatrix}
$$

$$
\begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}
\begin{bmatrix} 1 & 2 & -1 \\ 3 & -3 & 2 \\ -2 & 1 & 1 \end{bmatrix}
=
\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix}
\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -5/9 & 1 \end{bmatrix}
\begin{bmatrix} 1 & 2 & -1 \\ 0 & -9 & 5 \\ 0 & 0 & 16/9 \end{bmatrix}
$$

$$
\begin{bmatrix} 1 & 2 & -1 \\ 3 & -3 & 2 \\ -2 & 1 & 1 \end{bmatrix}
=
\begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}
\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix}
\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -5/9 & 1 \end{bmatrix}
\begin{bmatrix} 1 & 2 & -1 \\ 0 & -9 & 5 \\ 0 & 0 & 16/9 \end{bmatrix}
$$

$$
\underbrace{\begin{bmatrix} 1 & 2 & -1 \\ 3 & -3 & 2 \\ -2 & 1 & 1 \end{bmatrix}}_{A}
=
\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ -2 & -5/9 & 1 \end{bmatrix}}_{L}
\underbrace{\begin{bmatrix} 1 & 2 & -1 \\ 0 & -9 & 5 \\ 0 & 0 & 16/9 \end{bmatrix}}_{U}.
$$

Notice that L consists of the the entries from the elimination process each multiplied by negative one. This happens every time, making it easy to construct L for future computations!