Linear Algebra for Machine Learning-4

M. R. Hasan
Readings

- Advanced Engineering Mathematics (10th edition) by Erwin Kreyszig
- Chapter 7 & 8
What We Will Cover

• Intuition of the Eigenvalue equation
• Matrix eigenvalue problem
• Computing eigenvalues & eigenvectors
• Characteristic equation of a matrix
• Eigenbasis
• Matrix diagonalization
• Eigendecomposition
The study of Linear Algebra can be divided into solving two broad categories of linear systems:

- $Ax = b$
- $Ax = \lambda b$

We have discussed how to solve $Ax = b$
Linear Systems: $Ax = \lambda x$

- The **second major component** of Linear Algebra is to solve the following linear systems:
- $Ax = \lambda x$

- It is called the **eigenvalue equation**.
Ax = λx

Here A is a given square matrix.
λ an unknown scalar, and
x an unknown vector.

This is called the matrix eigenvalue problem.

The vector “x” is called the eigenvector and λ is called the eigenvalue.
Linear Systems: $Ax = \lambda x$

- $Ax = \lambda x$
- Our task is to determine $\lambda$’s and $x$’s that satisfy the equation.
- Since $x = 0$ is always a solution for any $\lambda$ and thus not interesting, we only admit solutions with $x \neq 0$.
- Before solving the equation, let’s try to understand the role of eigenvectors and the eigenvalues.
To motivate the matrix eigenvalue problem, consider multiplying two nonzero vectors by a given square matrix.

Our goal is to see what influence the multiplication of the given matrix has on the vectors.

\[
\begin{bmatrix}
6 & 3 \\
4 & 7 \\
\end{bmatrix}
\begin{bmatrix}
5 \\
1 \\
\end{bmatrix}
= 
\begin{bmatrix}
33 \\
27 \\
\end{bmatrix}
\]

\[
\begin{bmatrix}
6 & 3 \\
4 & 7 \\
\end{bmatrix}
\begin{bmatrix}
3 \\
4 \\
\end{bmatrix}
= 
\begin{bmatrix}
30 \\
40 \\
\end{bmatrix}
\]
The Matrix Eigenvalue Problem

• In the first case, we get a **totally new vector** with a **different direction** and **different length** when compared to the original vector.

• This is what usually happens and is of no interest here.

• In the second case **something interesting happens**.

\[
\begin{bmatrix}
6 & 3 \\
4 & 7
\end{bmatrix}
\begin{bmatrix}
5 \\
1
\end{bmatrix}
= 
\begin{bmatrix}
33 \\
27
\end{bmatrix}
\]

\[
\begin{bmatrix}
6 & 3 \\
4 & 7
\end{bmatrix}
\begin{bmatrix}
3 \\
4
\end{bmatrix}
= 
\begin{bmatrix}
30 \\
40
\end{bmatrix}
\]
The Matrix Eigenvalue Problem

• In the second case, the multiplication produces a vector that has the **same direction** as the original vector.

• The **magnitude or scale** of the vector is **changed by 10**.

• We denote this scale constant by \( \lambda \).

\[
\begin{bmatrix}
6 & 3 \\
4 & 7
\end{bmatrix}
\begin{bmatrix}
3 \\
4
\end{bmatrix}
= 
\begin{bmatrix}
30 \\
40
\end{bmatrix}
\]
The problem of systematically finding such λ’s and nonzero vectors for a given square matrix in a transformation that doesn’t change the vector directions is the essence of the matrix eigenvalue problem.

\[
\begin{bmatrix}
6 & 3 \\
4 & 7
\end{bmatrix}
\begin{bmatrix}
3 \\
4
\end{bmatrix} =
\begin{bmatrix}
30 \\
40
\end{bmatrix}
\]
The Matrix Eigenvalue Problem

• A value of \( \lambda \), for which the equation has a solution \( x \neq 0 \), is called an **eigenvalue or characteristic value** of the matrix \( A \).
• Another term for \( \lambda \) is a **latent root**.
• “Eigen” is a German word which means “proper” or “characteristic”.
• The corresponding solutions \( x \neq 0 \) of the equation are called the **eigenvectors or characteristic vectors** of \( A \) corresponding to that eigenvalue \( \lambda \).
Solving the Eigenvalue Equation

- So, given A, how do we find the eigenvalues and the eigenvectors?
- Write $Ax = \lambda x$ as follows:
  - $(A - I\lambda)x = 0$
  - We admit only nontrivial solution $x \neq 0$.
  - It is possible only if $(A - I\lambda)$ in non-invertible.
  - $(A - I\lambda)$ is non-invertible, if its determinant is zero.
  - This implies that $x$ has a non-trivial solution if $\det(A - \lambda I) = 0$. 
Solving the Eigenvalue Equation

- \( Ax = \lambda x \)
- \( (A - I\lambda)x = 0 \)
- This definition of an eigenvalue, which \textit{does not directly involve the corresponding eigenvector},

is the \textbf{characteristic equation} or characteristic polynomial of \( A \).
Solving the Eigenvalue Equation: Example

• Find the eigenvectors and eigenvalues of the following square matrix.

\[
A = \begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix}
\]

\[
\begin{align*}
A\vec{v} - \lambda \vec{v} &= 0 \\
\Rightarrow \vec{v}(A - \lambda I) &= 0,
\end{align*}
\]

\[
\text{Det}(A - \lambda I) = 0.
\]

\[
\text{Det} \left( \begin{bmatrix} 2 - \lambda & 3 \\ 2 & 1 - \lambda \end{bmatrix} \right) = 0.
\]

\[
(2 - \lambda)(1 - \lambda) - 6 = 0 \\
\Rightarrow 2 - 2\lambda - \lambda - \lambda^2 - 6 = 0 \\
\Rightarrow \lambda^2 - 3\lambda - 4 = 0.
\]
Solving the Eigenvalue Equation: Example

\[(2 - \lambda)(1 - \lambda) - 6 = 0\]
\[\Rightarrow 2 - 2\lambda - \lambda - \lambda^2 - 6 = 0\]
\[\Rightarrow \lambda^2 - 3\lambda - 4 = 0.\]

\[D = b^2 - 4ac = (-3)^2 - 4 \times 1 \times (-4) = 9 + 16 = 25.\]

\[A = \begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix}\]

\[\begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix} = -1 \begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix}\]

\[\lambda_1 = \frac{-b - \sqrt{D}}{2a} = \frac{3 - 5}{2} = -1,\]
\[\lambda_2 = \frac{-b + \sqrt{D}}{2a} = \frac{3 + 5}{2} = 4.\]

\[\left\{\begin{array}{l}
2x_{11} + 3x_{12} = -x_{11} \\
2x_{11} + x_{12} = -x_{12}
\end{array}\right.\]

\[x_{11} = -x_{12}\]

\[\vec{v}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}\]
Solving the Eigenvalue Equation: Example

\[
\begin{bmatrix}
2 & 3 \\
2 & 1 \\
\end{bmatrix}
\begin{bmatrix}
x_{21} \\
x_{22} \\
\end{bmatrix}
= 4 \times
\begin{bmatrix}
x_{21} \\
x_{22} \\
\end{bmatrix}
\]

\[
\begin{aligned}
x_{22} &= \frac{3}{2} x_{21} \\
\vec{v}_2 &= \begin{bmatrix} 3 \\ 2 \end{bmatrix}
\end{aligned}
\]

\[
A = \begin{bmatrix}
2 & 3 \\
2 & 1 \\
\end{bmatrix}
\]

\[
\begin{aligned}
\lambda_1 &= \frac{-b - \sqrt{D}}{2a} = \frac{3 - 5}{2} = -1, \\
\lambda_2 &= \frac{-b + \sqrt{D}}{2a} = \frac{3 + 5}{2} = 4.
\end{aligned}
\]

\[
\vec{v}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}
\]
Let’s gain further insight on the eigenvalue problem.

Recall that an eigenvector is a vector whose direction remains unchanged when a linear transformation is applied to it.

Consider the image below in which three vectors are shown.

The green square is only drawn to illustrate the linear transformation that is applied to each of these three vectors.
Linear Systems: $Ax = \lambda x$

- Eigenvectors (red) **do not change direction** when a linear transformation (e.g., scaling) is applied to them.
- Other vectors (yellow) do.
- Recall that a **transformation is done by the matrix-vector product**.
The transformation in this case is a simple scaling with factor 2 in the horizontal direction and factor 0.5 in the vertical direction.

Hence, we can define the transformation matrix $A$:

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix}.$$
Linear Systems: $Ax = \lambda x$

- A vector $\mathbf{v} = (x, y)$ is then scaled by applying this transformation as $\mathbf{v}' = A\mathbf{v}$.
- The figure shows that the direction of some vectors (shown in red) is not affected by this linear transformation.

These vectors are called **eigenvectors** of the transformation.

They **uniquely define** the square matrix $A$. 
An Important Theorem

- An important relation between the **determinant** and the **eigenvalues** of a matrix:
- The determinant of a matrix is a product of the eigenvalues.

$$\det(A) = \prod_{i=1}^{n} \lambda_i$$

**Useful Fact**: If any one eigenvalue is zero then determinant is also zero so matrix is non-invertible (singular).
The Matrix Eigenvalue Problem

• Observe that constructing matrices with specific eigenvalues and eigenvectors allows us to **stretch space** in desired directions.

Pointer:
We will use this fact to project the **covariance matrix** of sample data onto different bases.
The Matrix Eigenvalue Problem

• An eigenvector of a square matrix $A$ is a non-zero vector $v$ such that multiplication by $A$ alters only the scale of $v$:
  $$Av = \lambda v$$
The Matrix Eigenvalue Problem

• $Av = \lambda v$.
• The scalar $\lambda$ is known as the eigenvalue corresponding to this eigenvector.
• One can also find a left eigenvector such that $v^T A = \lambda v^T$.
• But we are usually concerned with right eigenvectors: $Av = \lambda v$. 
The Matrix Eigenvalue Problem

• $Av = \lambda v$.

• If $v$ is an eigenvector of $A$, then so is any rescaled vector $sv$ for $s \in \mathbb{R}$, $s \neq 0$.

• Moreover, $sv$ still has the same eigenvalue.

• For this reason, we usually only look for unit eigenvectors.
Eigenvectors

• We will now talk about some general properties of eigenvectors.

• Eigenvectors of an $n \times n$ matrix $A$ may (or may not!) form a basis for $\mathbb{R}^n$. 
Eigenvectors

• Recall that the **basis is a combination of vectors** that are **linearly independent**.

• So, if a matrix has $m$ column vectors and if all of them are linearly independent, then it forms basis in $\mathbb{R}^m$.

• We can **change the basis** by a linear combination of the column vectors.
Eigenvectors

- A linear combination of one basis set of vectors (purple) obtains new vectors (red).

The linear combinations relating the first set to the other extend to a linear transformation, called the **change of basis**.
Eigenvectors

• Let’s consider a problem of vector transformation.
• Say that we have an arbitrary vector “x” and we want to transform it by taking its dot product with a matrix “A”:
  \[ y = Ax \]
• When the dimension of these matrices are very large, this kind of multiplication becomes expensive.
• But we can make this transformation efficient by exploiting the matrix eigenvalue equation.
Eigenvectors

- Say that for the $n \times n$ matrix $A$, we have the following eigenvalue equation:
  - $Ax_j = \lambda_j x_j$
- The eigenvectors: $x_1, \ldots, x_n$
- The eigenvalues: $\lambda_1, \ldots, \lambda_n$
- Given an arbitrary vector $x$, we can compute its transformation $y = Ax$ trivially if we can **represent any $x$ in $\mathbb{R}^n$ uniquely as a linear combination** of the eigenvectors $x_1, \ldots, x_n$:

$$x = c_1 x_1 + c_2 x_2 + \cdots + c_n x_n.$$
Eigenvectors

• We can represent any $x$ in $\mathbb{R}^n$ uniquely as a linear combination of the eigenvectors $x_1, \ldots, x_n$, only if:
  • Eigenvectors of an $n \times n$ matrix $A$ form a basis for $\mathbb{R}^n$.
  • We call it “eigenbasis”.
  • In other words, only if ALL eigenvectors are linearly independent.
  • Then, we can represent any $x$ in $\mathbb{R}^n$ uniquely as a linear combination of the eigenvectors $x_1, \ldots, x_n$:

\[ x = c_1x_1 + c_2x_2 + \cdots + c_nx_n. \]
• How does the existence of an “eigenbasis” make the transformation $y = Ax$ trivial?

• Because we can decompose the complicated action of $A$ on an arbitrary vector $x$ into a sum of simple actions (multiplication by scalars) on the eigenvectors of $A$.

$$y = Ax = A(c_1x_1 + \cdots + c_nx_n)$$
$$= c_1Ax_1 + \cdots + c_nAx_n$$
$$= c_1\lambda_1x_1 + \cdots + c_n\lambda_nx_n.$$
Eigenvectors

• Now the *question* is: how do we know that the eigenvectors form an “eigenbasis” (all linearly independent)?

• We state without proof that if *n* eigenvalues are all *different*, we do obtain an “eigenbasis”.
Eigenvectors

• Another theorem that we state without proof.
• If a matrix is symmetric, then it has an orthonormal basis of eigenvectors for \( \mathbb{R}^n \).
• This useful theorem enables to change bases of a symmetric matrix.

Pointer:
In unsupervised dimensionality reduction problem (e.g., PCA) we will apply this theorem on symmetric covariance matrix for the sample data.
Another aspect of symmetric matrix will be useful.
Let’s say $A$ is a symmetric matrix.
Then, $A$ has an orthonormal basis of eigenvectors.
We can take its eigenvectors as column vectors and concatenate to form a matrix $X$.

\[
A = \begin{bmatrix}
1 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{bmatrix}
\]

\[
v_1 = \begin{bmatrix}
1 \\
0 \\
0
\end{bmatrix}
\]

\[
v_2 = \begin{bmatrix}
0 \\
1 \\
0
\end{bmatrix}
\]

\[
v_3 = \begin{bmatrix}
0 \\
0 \\
1
\end{bmatrix}
\]

\[
X = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}
\]

\[
A^T = A
\]
Eigenvectors

• This eigenvector matrix $X$ will be orthogonal.
• Therefore, we can write $X^{-1} = X^T$

$$X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$X^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = X^{-T} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$
Transformation of Eigenmatrix

• We have talked about the transformation of an arbitrary vector \( x \) by a matrix \( A \).
• Sometimes we want to transform a matrix into another matrix.
• One such transformation that we will use in this course is in which eigenvalues are preserved after the transformation.
• This type of transformation is called similarity transformation.

\[
\hat{A} = P^{-1}AP
\]

All matrices are \( n \times n \) and \( P \) is non-singular.
Transformation of Eigenmatrix

• By a suitable similarity transformation we can now transform a matrix \( A \) to a **diagonal matrix** \( D \) whose diagonal entries are the **eigenvalues** of \( A \):

\[
D = X^{-1}AX
\]

• Example:

\[
A = \begin{bmatrix}
7.3 & 0.2 & -3.7 \\
-11.5 & 1.0 & 5.5 \\
17.7 & 1.8 & -9.3
\end{bmatrix}
\]

Solving the eigenvalue equation, we find the **3 eigenvectors**
Transformation of Eigenmatrix

- By stacking the eigenvectors we form $X$:

$$X = \begin{bmatrix} -1 & 1 & 2 \\ 3 & -1 & 1 \\ -1 & 3 & 4 \end{bmatrix}$$

Then we compute $X^{-1}$:

$$X^{-1} = \begin{bmatrix} -0.7 & 0.2 & 0.3 \\ -1.3 & -0.2 & 0.7 \\ 0.8 & 0.2 & -0.2 \end{bmatrix}$$

Finally, we compute the **diagonal matrix** by a similarity transformation $D = X^{-1}AX$:

$$= \begin{bmatrix} -0.7 & 0.2 & 0.3 \\ -1.3 & -0.2 & 0.7 \\ 0.8 & 0.2 & -0.2 \end{bmatrix} \begin{bmatrix} -3 & -4 & 0 \\ 9 & 4 & 0 \\ -3 & -12 & 0 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -4 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$
Transformation of Eigenmatrix

- After we find the diagonal matrix of $A$, we can do another useful thing.
- We can multiply $D$ by $X$ (from left) and $X^{-1}$ (from right).
- $XDX^{-1} = XX^{-1}AXX^{-1} = (XX^{-1})A(XX^{-1}) = A$
- Hence, $A = XDX^{-1}$
- It gives us the matrix $A$ back.

This is useful because now we are able to decompose the matrix $A$ into its eigenvectors and eigenvalues.

It is called *eigendecomposition*!
Eigendecomposition

• We just derived the eigendecomposition of $A = XDX^{-1}$
• Now, what if $A$ is a symmetric matrix?
• Previously we saw that because $A$ is symmetric, its eigenvectors are orthogonal.
Eigendecomposition

- Hence, $X$ is orthogonal: $X^{-1} = X^T$
- We can write the decomposition as, $A = XDX^T$

**Eigendecomposition of $A = XDX^{-1}$**

**Pointer:**
In unsupervised dimensionality reduction (e.g., PCA), we will use this observation to **decompose the symmetric covariance matrix** for the sample data.

\[
A = \begin{bmatrix}
1 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{bmatrix} \quad v_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad v_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad v_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad X = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}
\]

\[
X^{-1} = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix} \quad = \quad X^{-T} = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}
\]
Some useful facts about eigendecomposition.

*Not every matrix* can be decomposed into eigenvalues and eigenvectors.

If a *matrix is not square*, the eigendecomposition is not defined (it will hurt us!)
Eigendecomposition

• Why is eigendecomposition useful?
• Many mathematical objects can be understood better by breaking them into constituent parts.
• Or finding some properties of them that are universal, not caused by the way we choose to represent them.
Eigendecomposition

• For example, integers can be decomposed into **prime factors**.
• The way we **represent the number** 12 will change depending on whether we write it in base ten or in binary.
• But it will always be true that $12 = 2 \times 2 \times 3$.
• From this representation we can **conclude useful properties**.
• Such as that 12 is not divisible by 5, or that any integer multiple of 12 will be divisible by 3.
Eigendecomposition

• We can also decompose matrices in ways that show us information about their functional properties that is not obvious from the representation of the matrix as an array of elements.

• One of the most widely used kinds of matrix decomposition is eigendecomposition, in which we decompose a matrix into a set of eigenvectors and eigenvalues.

\[ A = XDX^T \]