Foundations of AI | **Fall 2024**
Professors Greenwald and Ewing | **Machine Learning**

Eigenvalues and Eigenvectors (Draft)

1 Definitions

An **eigenvector** is a vector whose direction is unchanged by a linear transformation. In other words, given a matrix A, an eigenvector $\boldsymbol{v}$ is one for which $A\boldsymbol{v} = \lambda\boldsymbol{v}$. The value λ is called an **eigenvalue**.

N.B. While in the equation $A\boldsymbol{x} = \boldsymbol{b}$, we generally consider A to be an $m \times n$ matrix, $\boldsymbol{x}$ to be an n-dimensional vector, and $\boldsymbol{b}$ to be an m-dimensional vector, $\boldsymbol{v}$ appears on both the left hand side and the right hand side of this the equation $A\boldsymbol{v} = \lambda\boldsymbol{v}$. Therefore, m must equal n, so that our discussion of eigenvectors and eigenvalues is relevant only to square matrices.

Recall that a linear transformation can **rotate**, **shear**, or **stretch** a vector (or combine these operations). Eigenvectors are vectors that remain on their own span through the origin: i.e., they are *only* stretched (or shrunk). The amount by which they are stretched is called their eigenvalue. If this eigenvalue is negative, the vector's direction is reversed.

Following the German, where "eigen" means (among other things) characteristic, eigenvector is abbreviated "ev" and eigenvalue is abbreviated "ew."

2 Solution Method

The equation $A\boldsymbol{v} = \lambda\boldsymbol{v}$ can be rewritten as $(A - \lambda I)\boldsymbol{v} = \boldsymbol{0}$, by the following reasoning:

$$\begin{aligned} A\boldsymbol{v} &= \lambda\boldsymbol{v} \\ A\boldsymbol{v} &= \lambda I\boldsymbol{v} \\ (A - \lambda I)\boldsymbol{v} &= \boldsymbol{0} \end{aligned}$$

The vector $\boldsymbol{v} = \boldsymbol{0}$ is a solution to this equation, but it is not one of interest. We are interested in non-zero vectors that solve this equation.

N.B. We do not treat $\boldsymbol{0}$ as an eigenvector, as it would not have a corresponding well-defined eigenvalue.

Recall that $\boldsymbol{0}$ is the *only* solution to an equation of the form $B\boldsymbol{v} = \boldsymbol{0}$ iff

- the nullity of the matrix is 0
- the matrix is full rank
- the matrix B is invertible
- the determinant of B is non-zero
- the rows and columns of B are linearly independent

Therefore, we are interested in vectors $\boldsymbol{v}$ for which the determinant is 0! Our strategy for finding such vectors will be to solve for λ s.t. $\det(A - \lambda I) = 0$, and then given such λ, to solve for vectors $\boldsymbol{v} \neq 0$ s.t. $(A - \lambda I)\boldsymbol{v} = \boldsymbol{0}$.

3 Eigenvalues of Markov Chains

Recall that a Markov chain can be expressed as a Markov matrix, meaning a matrix whose values all lie between 0 and 1 (inclusive), and whose rows (or columns) sum to 1. A **stationary distribution** of a Markov matrix M is one for which $M\boldsymbol{x} = \boldsymbol{x}$. In other words, the stationary distribution of a Markov chain is an eigenvector with eigenvalue 1.

Let's compute the eigenvectors and eigenvalues of a few simple Markov chains to learn something about the nature of eigenvalues. First, consider the identity matrix in two dimensions, denoted I_2:

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

The matrix $I_2 - \lambda I_2$ is given by:

$$\begin{bmatrix} 1-\lambda & 0 \\ 0 & 1-\lambda \end{bmatrix}$$

Computing the determinant of $I_2 - \lambda I_2$ and setting it equal to zero yields $(1-\lambda)(1-\lambda) = 0$. Therefore, I_2 has two eigenvalues, both equal to 1.

Next, lets consider anti-diagonal "flip" of I_2,[1] and let's call it A, namely:

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

The matrix $A - \lambda I$ is:

$$\begin{bmatrix} 0-\lambda & 1 \\ 1 & 0-\lambda \end{bmatrix}$$

Computing the determinant of this matrix and setting it equal to zero yields $(0-\lambda)(0-\lambda) - 1 = 0$, or equivalently, $\lambda^2 - 1 = 0$. Therefore, this matrix also has two eigenvalues, $+1$ and -1.

Finally, consider the following matrix, which we dub A':

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

The matrix $A' - \lambda I$ is:

$$\begin{bmatrix} 0-\lambda & 1 \\ 0 & 1-\lambda \end{bmatrix}$$

Computing the determinant of this matrix and setting it equal to zero yields $(0-\lambda)(1-\lambda) = 0$. Therefore, this matrix also has two eigenvalues, 1 and 0.

In these examples, all three Markov chains have 1 as an eigenvalue. Indeed, it is a theorem that all Markov chains have 1 as an eigenvalue. (Equivalently, all Markov chains have a stationary distribution.[2])

Theorem All Markov chains have 1 as an eigenvalue.

Proof Consider the Markov chain M given by

$$\begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{n,n} \end{bmatrix}$$

[1]This terminology is *not* widely accepted!

[2]It is another matter whether that stationary distribution can be found via the power method.

with all columns summing to 1. Now, $M - I$ is given by

$$\begin{bmatrix} a_{1,1}-1 & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2}-1 & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{n,n}-1 \end{bmatrix}$$

Observe that the sum of all the rows in $M - I$ other than the jth row $\boldsymbol{a}_j^T$ equals $-\boldsymbol{a}_j^T$, i.e., $\boldsymbol{a}_{jk} = \sum_{i \neq j} \boldsymbol{a}_{ik}$. Therefore, $M - I$ is not of full rank, and $\det(M - I) = 0$. In other words, 1 is an eigenvalue of M. $\square$

We also observed that eigenvalues can repeat, and they can be negative. The next example (which, by the way, is *not* a Markov chain) demonstrates that not all matrices have real eigenvalues:

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

Computing the determinant of this matrix and setting it equal to zero yields $(0 - \lambda)(0 - \lambda) + 1 = 0$, or equivalently, $\lambda^2 = -1$. Therefore, this matrix has two eigenvalues, $+i$ and $-i$.

4 Finding Eigenvectors

To find the eigenvectors associated with an eigenvalue of a matrix, we simply plug λ into the equation $A\boldsymbol{v} = \lambda\boldsymbol{v}$ and solve for $\boldsymbol{v}$. Equivalently, we can solve $(A - \lambda I)\boldsymbol{v} = 0$.

For example, consider the matrix:

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

The eigenvalues of the matrix are 5 and 2.[3] To compute the corresponding eigenvectors, we solve the equations $(A - 5I)\boldsymbol{v} = 0$ and $(A - 2I)\boldsymbol{v} = 0$.

The first equation $(A - 5I)\boldsymbol{v} = 0$ expands as follows:

$$\begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

This matrix vector multiplication corresponds to the system of equations

$$\begin{aligned} -v_1 + v_2 &= 0 \\ 2v_1 - 2v_2 &= 0 \end{aligned}$$

In other words, $v_1 = v_2$. Therefore, $(1,1)^T$, $(-2,-2)^T$, $({}^{410}/_{1411}, {}^{410}/_{1411})^T$, and so on are *all* eigenvectors with eigenvalue 5. In other words, all non-zero scalar multiples of $(1,1)^T$ are eigenvectors with eigenvalue 5.

The second equation $(A - 2I)\boldsymbol{v} = 0$ expands as follows:

$$\begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

This matrix vector multiplication corresponds to the system of equations

$$\begin{aligned} 2v_1 + v_2 &= 0 \\ 2v_1 + v_2 &= 0 \end{aligned}$$

In other words, $v_1 = -{}^{v_2}/_2$. Therefore, $(-{}^1/_2, 1)^T$, $(1,-2)^T$, and so on are *all* eigenvectors with eigenvalue 2. In other words, all non-zero scalar multiples of $(-{}^1/_2, 1)^T$ are eigenvectors with eigenvalue 2.

[3] according to ChatGPT. It's probably worth checking its work!

5 The Characteristic Equation

The determinant of $A - \lambda I$ is a polynomial function of λ, called the **characteristic polynomial** of A, and the equation $\det(A - \lambda I) = 0$ is called the **characteristic equation**. The eigenvalues are the solutions to the characteristic equation.

The (Leibniz formula for the) determinant of a matrix is a sum over all permutations[4] σ of signed (i.e., ± 1) products of n terms of the form $a_{i\sigma(i)}$: i.e.,

$$\det(A) = \sum_{\text{permutations } \sigma} \left(\operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i\sigma(i)} \right) \tag{1}$$

Since the identity permutation is a valid permutation, this formula involves the product of all diagonal elements of A. In particular, $\det(A - \lambda I)$ involves the product of all diagonal elements of $(A - \lambda I)$, namely $(\lambda_1 - \lambda)(\lambda_2 - \lambda)\ldots(\lambda_n - \lambda)$. It follows that the characteristic polynomial is an nth degree polynomial, and thus has at least 1 and at most n (not necessarily real-valued) roots. Therefore:

Theorem An n-dimensional square matrix has at least 1 and at most n eigenvalues.

[4] A **permutation** σ is a bijective mapping from elements of $\{1, \ldots, n\}$ to itself.