Eigenvalues

Remind me: What are eigenvalues and eigenvectors again?

Given a square matrix $A \in \mathbb{R}^{n \times n}$,

$\vec{x} \neq \vec{0}$ is an eigenvector of A if there is a scalar λ so that

$$A\vec{x} = \lambda \vec{x} .$$

That is: In the direction of x, the effect of matrix-vector multiplication by A is simply scaling (='stretching').

Are there other ways of saying the same thing?

$$A\vec{x} = \lambda \vec{x}$$

$$\Leftrightarrow (A - \lambda I)\vec{x} = \vec{0}$$

$$\Leftrightarrow \vec{x} \in N(A - \lambda I)$$

Are there matrices for which the eigenvalues/vectors are easy to find?

For (both upper and lower) triangular matrices, the eigenvalues can be found on the diagonal.

To find the eigenvectors, find the nullspace N(A-λI).

What happens to eigenvalues if we change the matrix?

Suppose $Ax=\lambda x$. Then what are the eigenvalues of...

$\beta A \rightsquigarrow (\beta A)x = \beta\lambda x$ — Scaling

$A-\sigma I \rightsquigarrow (A-\sigma I)x = Ax - \sigma x = (\lambda-\sigma)x$ — Shift

$A^k \rightsquigarrow A^k x = \underbrace{AA\cdots A}_{k \text{ times}}x = \lambda^k x$ — Matrix Power

$A^{-1} \rightsquigarrow Ax = \lambda x \rightsquigarrow \frac{1}{\lambda}x = A^{-1}x$ — Inverse

All of these leave the eigenvectors x the same, but change the eigenvalues.

Can we change the eigenvectors? (but leave the eigenvalues the same)

Suppose $Ax=\lambda x$. Consider a so-called similarity transform with an invertible matrix T:

$$T^{-1}AT$$

So if x is an eigenvector of A, can we find an eigenvector of $T^{-1}AT$?

$$y = T^{-1}x$$

$$(T^{-1}AT)y = T^{-1}ATT^{-1}x = T^{-1}Ax = \lambda T^{-1}x = \lambda y.$$

Does every n×n matrix have n linearly independent eigenvectors?

First of all, if it does, then we could take those eigenvectors x_i and stick them into a matrix X:

$$X = \begin{pmatrix} | & & | \\ x_1 & \cdots & x_n \\ | & & | \end{pmatrix}$$

Then

$$AX = X \underbrace{\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}}_{D} \quad \Leftrightarrow \quad X^{-1}AX = D$$

So in this case, there is a similarity transform (with the matrix of eigenvectors X) that takes A to a diagonal matrix.

In this case, A is called diagonalizable.

Unfortunately, not every matrix is diagonalizable.

In particular, for repeated eigenvalues, it is not guaranteed that a linearly independent eigenvector exists for each repetition.

Can you give an example?

$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ → Eigenvalue 1 has 'multiplicity' 2.

$A - 1I = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ → dim N(A-1I)=1 → only one eigenvector

How can we estimate an eigenvalue, given an eigenvector?

The so-called Rayleigh quotient allows this:

$$\frac{x \cdot Ax}{x \cdot x}$$

If $Ax=\lambda x$ with $x \neq 0$, then

$$\frac{x \cdot Ax}{x \cdot x} = \frac{x \cdot \lambda x}{x \cdot x} = \lambda$$

For approximate eigenvectors, we can compute an approximate eigenvalue this way.

So how do we actually go about finding eigenvalues/vectors?

Assume A is n×n and diagonalizable with

$$Ax_1 = \lambda_1 x_1 \quad \cdots \quad Ax_n = \lambda_n x_n$$

$$|\lambda_1| \geq |\lambda_2| \cdots \geq |\lambda_n|$$

So any vector x can be written as a linear combination of the x_i.

$$x = \alpha_1 x_1 + \cdots + \alpha_n x_n$$

Now compute Ax:

$$\begin{aligned} Ax &= \alpha_1 Ax_1 + \cdots + \alpha_n Ax_n \\ &= \alpha_1 \lambda_1 x_1 + \cdots + \alpha_n \lambda_n x_n \end{aligned}$$

What if we multiply by A 19,999 more times?

$$A^{20,000} x = AA \cdots Ax$$

$$= \alpha_1 \lambda_1^{20,000} \cdot x_1 + \cdots + \alpha_n \lambda_n^{20,000} x_n$$

So how do we actually go about finding eigenvalues/vectors? (continued)

Divide both sides by $\lambda_1^{20,000}$ to clarify what's going on:

$$\frac{A^{20,000}x}{\lambda_1^{20,000}} = \alpha_1 x_1 + \alpha_2 \left(\frac{\lambda_2}{\lambda_1}\right)^{20,000} x_2 + \cdots + \alpha_n \left(\frac{\lambda_n}{\lambda_1}\right)^{20,000} x_n$$

If $|\lambda_2| \leq |\lambda_1|$, then $\left(\frac{\lambda_2}{\lambda_1}\right)^{20,000}$ is tiny... almost zero.

Lesson: In this case, only the contribution from x_1 survives.

So that could be a recipe to find x_1.

This is called power iteration. (Because we take a high power of A.)

Demo: Power iteration and its variants

Is power iteration bulletproof?

No, not quite.

- If λ_1 is complex-valued, the eigenvector will not converge.

- $|\lambda_1| = |\lambda_2|$ means no guaranteed convergence

- $\alpha_1 = 0$. (Not a problem in practice. Rounding will ensure that a "little bit" of x_1 appears.)

What variants of power iteration are there?

For each variant, you pick an initial, non-zero vector x_0 arbitrarily.

"Plain" power iteration:

$$x_{k+1} = A x_k$$ (overflows quickly if $|\lambda_1| > 1$)

Normalized power iteration:

$$x_{k+1} = \frac{A x_k}{\|A x_k\|}$$ (typically converges to eigenvector for λ_1)

Inverse iteration:

$$x_{k+1} = \frac{A^{-1} x_k}{\|A^{-1} x_k\|}$$ (typically converges to eigenvector for λ_n)

Inverse iteration with shift:

$$x_{k+1} = \frac{(A - \sigma I)^{-1} x_k}{\|(A - \sigma I)^{-1} x_k\|}$$ (typically converges to eigenvector with eigenvalue closest to σ)

↑ greek letter sigma (not zero)

Rayleigh quotient iteration:

Same as Inverse Iteration with Shift, but set σ equal to the Rayleigh quotient computed with x_k.

(Typically converges to one of the eigenvectors, often the one that x_0 is closest to. It does converge very quickly.)

What if we would like to find all eigenvectors at the same time?

Idea: Just iterate multiple vectors at once.

This is called "Simultaneous Iteration" and does not work.

It does not work because all vectors will converge to the eigenvector for the largest (by magnitude) eigenvalue.

It is equivalent to running multiple copies of power iteration at once.

Idea 2: Keep the vectors orthogonal while we're iterating.

How does orthogonal iteration work?

Choose arbitrary $X_0 \in \mathbb{R}^{n \times p}$ $(p \leq n)$

Step 1: $Q_1 R_1 = X_0$ ← Compute QR of X_0

$X_1 = A Q_1$

Step 2: $Q_2 R_2 = X_1$ ← Compute QR of X_1

$X_2 = A Q_1$

Step 3: ⋮

Suppose this "converges" so that $X_{k+1} \approx X_k$.

$\rightsquigarrow$ $Q_{k+1} R_{k+1} = X_k \approx X_{k+1}$

$\rightsquigarrow$ $X_{k+1} = A Q_{k+1}$

$\rightsquigarrow$ $Q_{k+1} R_{k+1} = A Q_{k+1}$

$\rightsquigarrow$ $Q_{k+1} R_{k+1} Q_{k+1}^T = A$

How does that help?

It is a similarity transform to an upper triangular matrix.

That means that the eigenvalues of A appear on the diagonal of R.

$A = Q R Q^T$ with Q orthogonal and R upper triangular is called "Schur form". The eigenvalues of A can be "read off" from it.

Demo: Orthogonal Iteration

Can we find eigen*vectors* from the Schur form as well?

$A = QRQ^T$

(1) Find eigenvector x of R: $Rx = \lambda x$

$\hookrightarrow x \in N(R - \lambda I)$

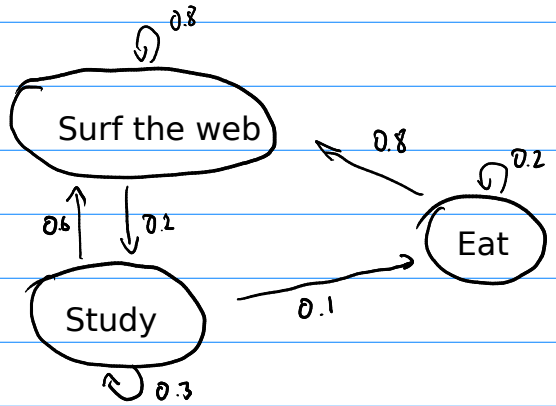
Use nullspace finding.

Make use of R being upper triangular.

(2) y=Qx is then an eigenvector of A:

$Ay = QRQ^T Qx = \lambda y$

In what way can eigenvalue computation be applied to Markov chains?



Important Assumption: Only the most recent state matters to determine the probabilities for the next state.

(This is called the "Markov property.")

Write the transition probabilities into a matrix as before.

from state

	surf	study	eat	
$A =$	.8	.6	.8	surf
	.2	.3	0	study
	0	.1	.2	eat

to state

The columns add up to 1 because in each state the probabilities of the next states must add up to 1.

State transitions are modeled by matrix vector product, where the ith vector entry corresponds to the probability of being in the ith state.

$A \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} .8 \\ .2 \\ 0 \end{pmatrix}$ surf

$A \begin{pmatrix} .5 \\ .5 \\ 0 \end{pmatrix} = \begin{pmatrix} .4 + .3 \\ .1 + .15 \\ .05 \end{pmatrix}$

"Equilibrium" means that the previous state equals the next state, i.e. the probabilities from one state to the next do not change:

$Ax = \lambda x$ → eigenvalue problem

Demo: Finding an equilibrium distribution using the power method