edX

Course › Unit 2: ... › 4 Eigen... › 17. Ap...

# 17. Application: the Pagerank Algorithm

While we've studied a lot of physical systems in this class, there are many non-physical systems that use linear algebra. One rich source of applications is computational science. We'll give one example here, the **pagerank algorithm** used by Google and other search engines to determine the order of results in a query. To define this rigorously, we'll use the tools of linear algebra. What we will do here will not be the exact pagerank algorithm, but will get across the general idea while also providing some linear algebra practice.

Some Pagerank History

Show

Working through the problems below will both provide excellent linear algebra practice as well as teach you about the pagerank algorithm.

**Definitions and Problems**

In order to show how pagerank works, we'll work through a series of examples. First imagine we have three webpages: Wikipedia, Facebook, and Buzzfeed. Let's assume this is the entire internet, and the percentage of people on each page at a time is $W$, $F$, and $B$ respectively. Now we need to find the number of links between each page. We'll store this information in a matrix, calling this matrix $\mathbf{P}$. Starting with no links, we write

$$\mathbf{P} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Now to start filling in this matrix. Suppose Buzzfeed has six links to Facebook, and four links to Wikipedia. In our matrix, the column represents where the link is coming **from** , and the row represents where the link is going **to** . So for the links from Buzzfeed to

Facebook, the matrix becomes

$$\mathbf{P} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 6 \\ 0 & 0 & 0 \end{pmatrix}$$

and then adding the links to Wikipedia, we get

$$\mathbf{P} = \begin{pmatrix} 0 & 0 & 4 \\ 0 & 0 & 6 \\ 0 & 0 & 0 \end{pmatrix}$$

The matrix we are constructing is called the adjacency matrix. Each entry represents a connection from one node to another in a graph. This is a useful way of storing information dealing with different objects and directed connections between them.

**Problem 17.1** Fill in the remainder of the adjacency matrix. Suppose that Wikipedia has one link to each Facebook and Buzzfeed, and Facebook has one link to Buzzfeed and two links to Wikipedia. What is our resultant matrix?

**Solution**

$$\mathbf{P} \quad = \quad \begin{pmatrix} 0 & 2 & 4 \\ 1 & 0 & 6 \\ 1 & 1 & 0 \end{pmatrix}$$

Hide

Now that we have an adjacency matrix constructed, we can move forward and begin to attempt a ranking. To do this, we have to find the eigenvalues.

**Problem 17.2** Find the eigenvalues of the matrix $\mathbf{P}$. Which eigenvalue has the largest absolute value?

**Solution**

The eigenvalues are the solutions of the equation

$$0 = \det(\mathbf{P} - \lambda \mathbf{I})$$

Plugging in numbers we get

$$0 = \det \begin{pmatrix} -\lambda & 2 & 4 \\ 1 & -\lambda & 6 \\ 1 & 1 & -\lambda \end{pmatrix}$$

Computing the determinant gives

$$0 = -\lambda^3 + 4 + 12 + 6\lambda + 2\lambda + 4\lambda.$$

Simplifying we see that

$$0 = \lambda^3 - 12\lambda - 16.$$

We search for a root and find one, $\lambda_1 = -2$, and then factoring gives

$$0 = (\lambda + 2)(\lambda^2 - 2\lambda - 8).$$

Now factoring the last term we get

$$0 = (\lambda + 2)(\lambda + 2)(\lambda - 4),$$

so our last two eigenvalues are

$$\begin{aligned} \lambda_2 &= -2, \\ \lambda_3 &= 4. \end{aligned}$$

Now we have found three eigenvalues, and the eigenvalue with the largest absolute value is $4$.

Hide

Why do we care about the eigenvalue that has the largest absolute value? Qualitatively, if we watch the path of a user as they browse the web, this is like applying the matrix to an initial state vector repeatedly. The eigenvector with the largest eigenvalue will dominate.

So this eigenvector will give us the best ranking.

**Problem 17.3** Find the eigenvector corresponding to the eigenvalue of largest absolute value.

**Solution**

Finding the eigenvector is equivalent to finding the null space of the matrix

$$P - 4I = \begin{pmatrix} -4 & 2 & 4 \\ 1 & -4 & 6 \\ 1 & 1 & -4 \end{pmatrix}.$$

To find the null space we use Gauss-Jordan elimination. We divide the first row by $-4$ and then subtract it from the second and third rows getting

$$\begin{pmatrix} 1 & -\frac{1}{2} & -1 \\ 0 & -\frac{7}{2} & 7 \\ 0 & \frac{3}{2} & -3 \end{pmatrix}.$$

Now we divide the second row by $-7/2$ to get

$$\begin{pmatrix} 1 & -\frac{1}{2} & -1 \\ 0 & 1 & -2 \\ 0 & \frac{3}{2} & -3 \end{pmatrix}.$$

We now subtract $\frac{3}{2}(\mathbf{row\ 2})$ from the $\mathbf{row\ 3}$, and add $\frac{1}{2}(\mathbf{row\ 2})$ to the $\mathbf{row\ 1}$, to get

$$\begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{pmatrix}$$

Now from inspection we see that a vector in the nullspace is

$$\mathbf{v} = \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}.$$

Any vector in the nullspace is an eigenvector, so $\mathbf{v}$ is an eigenvector of $P$. Now we can use this vector as a rough ranking of our websites; Wikipedia and Facebook each have a rank of $2$, and Buzzfeed has a rank of $1$.

Hide

## 17. Application: the Pagerank Algorithm

Hide Discussion

**Topic:** Unit 2: Linear Algebra, Part 2 / 17. Application: the Pagerank Algorithm

**Add a Post**

Show all posts ▼ by recent activity ▼

There are no posts in this topic yet.

✖

Learn About Verified Certificates