Power Method

Matrix Algebra 4 Statistical Learning

Gaston Sanchez

Power Method

About the Power Method

One of the basic procedures following a successive approximation approach for the dominant eigenvector is precisely the Power Method.

In its simplest form, the Power Method (PM) allows us to find **the largest** eigenvector and its corresponding eigenvalue.

About the Power Method

Choose an arbitrary vector $\mathbf{w_0}$ to which we will apply the symmetric matrix $\mathbf{S}$ repeatedly to form the following sequence:

$$
\begin{aligned}
\mathbf{w_1} &= \mathbf{S}\mathbf{w_0} \\
\mathbf{w_2} &= \mathbf{S}\mathbf{w_1} = \mathbf{S^2}\mathbf{w_0} \\
\mathbf{w_3} &= \mathbf{S}\mathbf{w_2} = \mathbf{S^3}\mathbf{w_0} \\
&\vdots \\
\mathbf{w_k} &= \mathbf{S}\mathbf{w_{k-1}} = \mathbf{S^k}\mathbf{w_0}
\end{aligned}
$$

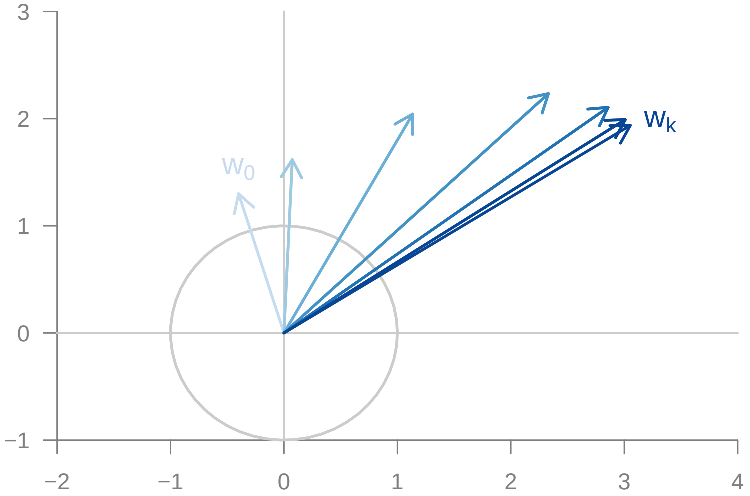
Power Method: Example

Consider a matrix $\mathbf{S}$

$$\mathbf{S} = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$$

and an initial vector $\mathbf{w_0}$

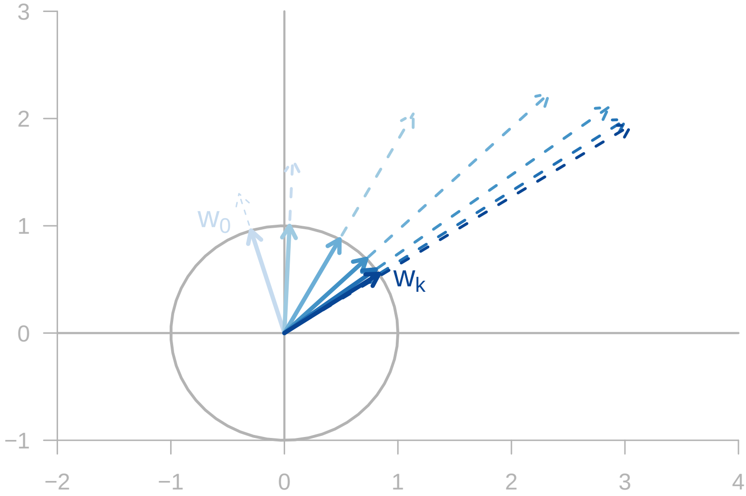
$$\mathbf{w_0} = \begin{bmatrix} -0.4 \\ 1.3 \end{bmatrix}$$



3
2
1
0
−1
−2
−1
0
1
2
3
4
w0
wk

About the Power Method

- In practice, we must rescale the obtained vector $\mathbf{w_k}$ at each step.
- The rescaling will allows us to judge whether the sequence is converging.
- After some iterations, the vector $\mathbf{w_{k-1}}$ and $\mathbf{w_k}$ will be very similar
- Assuming a reasonable scaling strategy, the sequence will usually converge to the dominant eigenvector of $\mathbf{S}$.



3
2
1
0
−1
−2
−1
0
1
2
3
4
w_0
w_k

Dominant Eigenvalue

The obtained vector is the dominant eigenvector. To get the corresponding eigenvalue we calculate the so-called **Rayleigh quotient** given by:

$$\lambda = \frac{\mathbf{w_k}^\mathsf{T}\mathbf{S}\mathbf{w_k}}{\mathbf{w_k}^\mathsf{T}\mathbf{w_k}}$$

Remarks

Conditions for the power method to be succesfully used:

- The matrix must have a *dominant* eigenvalue.
- The starting vector $\mathbf{w_0}$ must be nonzero.
- We need to scale each of the vectors $\mathbf{w_k}$
  otherwise the algorithm will "explode"

PM Pseudocode

Let's consider a more detailed version of the PM algorithm:

1. Start with an arbitraty initial vector $\mathbf{w}$
2. Obtain product $\tilde{\mathbf{w}} = \mathbf{S}\mathbf{w}$
3. Normalize $\tilde{\mathbf{w}}$
$$\text{e.g.} \quad \mathbf{w} = \frac{\tilde{\mathbf{w}}}{\left\|\tilde{\mathbf{w}}\right\|_{p=2}}$$
4. Compare $\mathbf{w}$ with its previous version
5. Repeat steps 2 till 4 until convergence

Why does the PM work?

Assume that the matrix $\mathbf{S}$ has p eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$, and that they are ordered in decreasing way $|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_p|$.

Note that the first eigenvalue is strictly greater than the second one. This is a very important assumption.

In the same way, we'll assume that the matrix $\mathbf{S}$ has p linearly independent vectors $\mathbf{u_1}, \ldots, \mathbf{u_p}$ ordered in such a way that $\mathbf{u_j}$ corresponds to λ_j.

Why does the PM work?

The initial vector $\mathbf{w_0}$ may be expressed as a linear combination of $\mathbf{u_1}, \ldots, \mathbf{u_p}$

$$\mathbf{w_0} = a_1\mathbf{u_1} + \cdots + a_p\mathbf{u_p}$$

At every step of the iterative process the vector $\mathbf{w_k}$ is given by:

$$\mathbf{w_k} = a_1\lambda_1^k\mathbf{u_1} + \cdots + a_p\lambda_p^k\mathbf{u_p}$$

Why does the PM work?

Since λ_1 is the dominant eigenvalue, the component in the direction of $\mathbf{u_1}$ becomes relatively greater than the other components as k increases. If we knew λ_1 in advance, we could rescale at each step by dividing by it to get:

$$\left(\frac{1}{\lambda_1^k}\right)\mathbf{w_k} = a_1\mathbf{u_1} + \cdots + a_p\left(\frac{\lambda_p^k}{\lambda_1^k}\right)\mathbf{u_p}$$

which converges to the eigenvector $a_1\mathbf{u_1}$, provided that a_1 is nonzero.

Why does the PM work?

Of course, in real life this scaling strategy is not possible—we don't know λ_1. Consequenlty, the eigenvector is determined only up to a constant multiple, which is not a concern since the really important thing is the *direction* not the length of the vector.

The speed of the convergence depends on how bigger λ_1 is respect with to λ_2, and on the choice of the initial vector $\mathbf{w_0}$. If λ_1 is not much larger than λ_2, then the convergence will be slow.

More Remarks

- The power method is a sequential method.
- We can obtain $\mathbf{w_1}, \mathbf{w_2}$, and so on, step by step.
- If we only need the first k vectors, we can stop the procedure at the desired stage.

Obtaining more eigenvectors?

For **symmetric** matrices, once we've obtained the first eigenvector $\mathbf{w_1}$ and eigenvalue λ_1, we can compute the second eigenvector by reducing the matrix $\mathbf{S}$ by the amount explained by the first eigenvector.

This operation of reduction is called **deflation** and the residual matrix is obtained as:

$$\mathbf{S_1} = \mathbf{S} - \lambda_1 \mathbf{w_1} \mathbf{w_1}^\top$$

To get the second eigenvalue and its corresponding eigenvector, we operate on $\mathbf{S_1}$ in the same way as the operations on $\mathbf{S}$.

Bibliography

- **Multivariate Analysis** by Maurice Tatsuoka (1988). *Chapter 5: More Matrix Algebra*. Macmillan Publishing.
- **Mathematical Tools for Applied Multivariate Analysis** by J.D. Carroll, P.E. Green, and A. Chaturvedi (1997). *Chapter 5: Decomposition of Matrix Transformations: Eigenstructures and Quadratic Forms*. Academic Press.
- **Hands-on Matrix Algebra using R** by Hrishikesh Vinod (2011). *Chapter 9: Eigenvalues and Eigenvectors*. World Scientific.