Foundations of Data Science

Lecture 5: Eigenvalue and Eigenvector

MING GAO

DaSE@ECNU
(for course related communications)
mgao@sei.ecnu.edu.cn

Oct. 15, 2016

Outline

Outline

1. Eigenvalue and Eigenvector
   - Definition
   - Power Method
   - Diagonalization

Eigenvalue and eigenvector

Definition

Given a matrix $A \in \mathbb{R}^{n \times n}$, and non-zero column vector v, if $Av = \lambda v$, then λ is an eigenvalue of A and v is an eigenvector corresponding to eigenvalue λ.

- Transformation: a matrix A acts on vectors x like a function does, with input x and output Ax. Eigenvectors are vectors for which Ax is parallel to x. In other words: $Ax = \lambda x$.
- If M is Hermitian, then all the eigenvalues of M are real. Note that a real symmetric matrix is always Hermitian, i.e., $A^T = A$.
- The adjacency matrix of an undirected graph is symmetric, and this implies that its eigenvalues are all real.

Outline

How to find eigenvalues and eigenvectors

Method 1

$Av = \lambda v \Leftrightarrow (A - \lambda I)v = 0$. For non-zero vector v, which is equivalent to $det(A - \lambda I) = 0$, where the equation is called characteristic equation.

- For example, $A = \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix}$, we have $det(A - \lambda I) = \lambda^2 - 3\lambda + 2 = 0$. We find eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 2$. We can find eigenvector via solving the linear equation $(A - \lambda_i I)v = 0$. That is $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$.
- This approach is not scalable.

Method 2: the power method

- The power method is an iterative algorithm which has the following basic form for generating a single eigenvalue and eigenvector of A.
  - Initial nonzero vector $x^{(0)} \in \mathbb{R}^n$, such that $\|x^{(0)}\| = 1$.
  - For $k = 0, 1, \cdots$, we let $x^{(k+1)}$ be a nonzero multiple of $Ax^{(k)}$, typically $x^{(k+1)} = Ax^{(k)}/\|Ax^{(k)}\|$.

The power method

Analysis

- Why is it correct? Suppose $x^{(0)}$ is in the subspace generated the eigenvectors, i.e., $x^{(0)} = \sum_{i=1}^{n} c_i v_i$ with $c_1 \neq 0$. Then $x^{(k)}$ converges to the dominant eigenvector v_1 because

$$\begin{aligned}\lim_{k\to\infty} A^k x^{(0)} &= \lim_{k\to\infty} \sum_{i=1}^{n} c_i A^k v_i = \lim_{k\to\infty} \sum_{i=1}^{n} c_i \lambda_i^k v_i \\ &= \lim_{k\to\infty} c_1 \lambda_1^k \Big[v_1 + \sum_{i=2}^{n} \frac{c_i}{c_1} \frac{\lambda_i}{\lambda_1}^k v_i \Big] = c_1 \lambda_1^k v_1\end{aligned}$$

- It will find only one eigenvalue (the one with the greatest absolute value), and it may converge only slowly.
- It can fail if there is not a single largest eigenvalue, i.e., $\lambda_1 = \lambda_2$.

The power method cont.

Extension

- Inverse power method: it operates with A^{-1} rather than A since the eigenvalues of A^{-1} are $\frac{1}{\lambda_i}$. It gives a way of finding the smallest (in absolute value) eigenvalue of a matrix.
- Spectral shift: using the fact that the eigenvalues of $A - \alpha I$ are $\lambda_i - \alpha$. If we find the largest eigenvalue λ_1, we can find the largest in absolute value of $\lambda_i - \lambda_1$. However, it is not clear how it could be implemented in general to find all the eigenvalues of matrix A.
- Symmetric power method: we can repeatedly find all eigenvalues of a symmetric matrix A with distinct eigenvalues because

$$A = \sum_{i=1}^{n} \lambda_i v_i v_i^T,$$

$$A^{(i)} = A^{(i-1)} - \lambda_i v_i v_i^T$$

Characteristic polynomial

Properties

Given a square matrix A, $p_A(\lambda) = det(A - \lambda I)$ is called characteristic polynomial.

- Suppose $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, we have $p_A(\lambda) = (a - \lambda)(d - \lambda) - bc = \lambda^2 - (a + d)\lambda + (ad - bc)$.
- $p_A(\lambda) = \lambda^2 - Trace(A)\lambda + det(A)$.
- If A has two eigenvalues λ_1 and λ_2, then $p_A(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2) = \lambda^2 - (\lambda_1 + \lambda_2)\lambda + \lambda_1\lambda_2$. Hence, $Trace(A) = \lambda_1 + \lambda_2$ and $det(A) = \lambda_1\lambda_2$.
- In general, $det(A - \lambda I)$ has n roots $\lambda_1, \lambda_2, \cdots, \lambda_n$, we have $Trace(A) = \sum_{i=1}^n \lambda_i$ and $det(A) = \Pi_{i=1}^n \lambda_i$.

Outline

Diagonalization

Definition of similar matrices

If matrices $A, B \in \mathbb{R}^{n\times n}$ are said to be similar if there is an invertible $n \times n$ matrix P such that $A = PBP^{-1}$.

- Theorem: If $n \times n$ matrices are similar, then they have the same characteristic of polynomial and hence the same enigenvalues (with the same multiplicities).

Definition of diagonalizable

A square matrix A is said to be diagonalizable if it is similar to a diagonal matrix, i.e., there exists an invertible matrix P and a diagonal matrix D such that $A = PDP^{-1}$.

- Why useful? If A is diagonalizable, then $A^k = PD^kP^{-1}$ for $k > 0$.
- How to find diagonal matrix? If $v_1, \cdots, v_n$ are linearly independent eigenvectors of A and λ_i are their corresponding eignevalues, then $A = PDP^{-1}$, where $P = [v_1 \ \cdots \ v_n]$ and $D = Diag(\lambda_1, \cdots, \lambda_n)$.

Conditions of diagonalizable

Theorem

- A matrix $A \in \mathbb{R}^{n\times n}$ is diagonalizable if and only if A has n linearly independent eigenvectors.
- A matrix $A \in \mathbb{R}^{n\times n}$ with n distinct eigenvalues is diagonalizable.

Example

- $A = \begin{bmatrix} 2 & 4 & 6 \\ 0 & 2 & 2 \\ 0 & 0 & 4 \end{bmatrix}$ with $\lambda_1 = 2$ and $\lambda_1 = 4$.
- While the basis for $\lambda_1 = 2$ is $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ $((A - 2I)\mathbf{v} = 0)$.
- Hence, the matrix is not diagonalizable.

Diagonalization of of symmetric matrix

Definition of orthogonal

- Matrix P is orthogonal if $P^{-1} = P^T$.
- Matrix A is orthogonal diagonalizable if there is a square matrix P such that $A = PDP^T$ where D is a diagonal matrix.

Properties of symmetric matrix

- If A is symmetric then any two eigenvalues from different eigenspaces are orthogonal.
- A has n real eigenvalues if we count multiplicity.
- For each eigenvalue, the dimension of the corresponding eigenspace is equal to the algebraic multiplicity of that eigenvalue.
- The eigenspaces are mutually orthogonal.
- A is orthogonal diagonalizable.

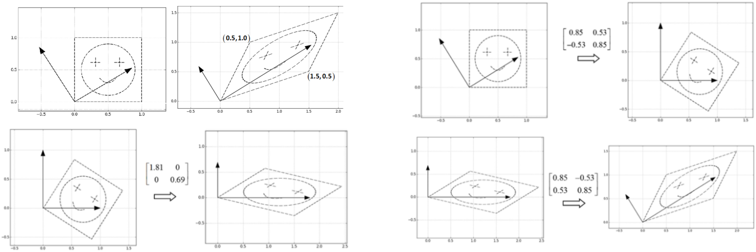
Geometry of diagonalization

Example

Let $A = \begin{bmatrix} 1.5 & 0.5 \\ 0.5 & 1.5 \end{bmatrix}$ associated with eigenvectors $U = \begin{bmatrix} 0.85 & -0.53 \\ 0.53 & 0.85 \end{bmatrix}$ (column vector) and eigenvalues $1.81, 0.69$, then

$$A = U \begin{bmatrix} 1.81 & 0 \\ 0 & 0.69 \end{bmatrix} U^T.$$

Rotation and scaling



Take-home messages

- Eigenvalue and eigenvector
  - Definition
  - Power method
  - Diagonalization