Lec 9: Eigenvalue Decomposition

Ailin Zhang

Agenda

- Review eigenvalue and eigenvector
- Eigenvalue decomposition
- Power Method
- Matrix form of power method
- Principal component analysis

Definitions of Eigenvalue and Eigenvector

Let A be an $n \times n$ matrix. The scalar λ is called an eigenvalue of A if there is a nonzero vector $\mathbf{x}$ such that

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

The vector $\mathbf{x}$ is called an eigenvector of A corresponding to λ.

Eigenvalues and Eigenvectors of a Matrix

Let A be an $n \times n$ matrix.

1. An eigenvalue of A is a scalar λ such that
$$\det(\lambda I - A) = 0.$$
2. The eigenvectors of A corresponding to λ are the nonzero solutions of
$$(\lambda I - A)\mathbf{x} = \mathbf{0}.$$

The equation $\det(\lambda I - A) = 0$ is called the characteristic equation of A.
$$|\lambda I - A| = \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0$$
is called the characteristic polynomial of A. Because the characteristic polynomial of A is of degree n, A can have at most n distinct eigenvalues.

Definition of a Diagonalizable Matrix

An $n \times n$ matrix A is diagonalizable if A is similar to a diagonal matrix. That is, A is diagonalizable if there exists an invertible matrix P such that $P^{-1}AP$ is a diagonal matrix.

Theorem

An $n \times n$ matrix A is diagonalizable if and only if it has n linearly independent eigenvectors.

Proof

First, assume A is diagonalizable. Then there exists an invertible matrix P such that $P^{-1}AP = D$ is diagonal. Letting the main entries of D be $\lambda_1, \lambda_2, \ldots, \lambda_n$ and the column vectors of P be $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$ produces

$$PD = \left[\mathbf{p}_1 \vdots \mathbf{p}_2 \vdots \cdots \vdots \mathbf{p}_n\right] \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

$$= \left[\lambda_1\mathbf{p}_1 \vdots \lambda_2\mathbf{p}_2 \vdots \cdots \vdots \lambda_n\mathbf{p}_n\right].$$

Because $P^{-1}AP = D, AP = PD$, which implies the column vectors $\mathbf{p}_i$ of P are eigenvectors of A

$$\left[A\mathbf{p}_1 \vdots A\mathbf{p}_2 \vdots \cdots \vdots A\mathbf{p}_n\right] = \left[\lambda_1\mathbf{p}_1 \vdots \lambda_2\mathbf{p}_2 \vdots \cdots \vdots \lambda_n\mathbf{p}_n\right]$$

Moreover, because P is invertible, its column vectors are linearly independent. So, A has n linearly independent eigenvectors.

Proof

Conversely, assume A has n linearly independent eigenvectors $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$ with corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Let P be the matrix whose columns are these n eigenvectors. That is, $P = \left[\mathbf{p}_1 \vdots \mathbf{p}_2 \vdots \cdots \vdots \mathbf{p}_n\right]$. Because each $\mathbf{p}_i$ is an eigenvector of A, you have $A\mathbf{p}_i = \lambda_i \mathbf{p}_i$ and

$$AP = A\left[\mathbf{p}_1 \vdots \mathbf{p}_2 \vdots \cdots \vdots \mathbf{p}_n\right] = \left[\lambda_1\mathbf{p}_1 \vdots \lambda_2\mathbf{p}_2 \vdots \cdots \vdots \lambda_n\mathbf{p}_n\right].$$

The right-hand matrix in this equation can be written as the matrix product below.

$$AP = \left[\mathbf{p}_1 \vdots \mathbf{p}_2 \vdots \cdots \mathbf{p}_n\right] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = PD$$

Finally, because the vectors $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$ are linearly independent, P is invertible and you can write the equation $AP = PD$ as $P^{-1}AP = D$, which means that A is diagonalizable.

Sufficient Condition for Diagonalization

Theorem

If an $n \times n$ matrix A has n distinct eigenvalues, then the corresponding eigenvectors are linearly independent and A is diagonalizable.

Remember that the condition in the theorem is sufficient but not necessary for diagonalization. In other words, a diagonalizable matrix need not have distinct eigenvalues. Matrix with distinct eigenvalues must be diagonalizable

Eigenvalues of Symmetric Matrix

If A is an $p \times p$ symmetric matrix, then the following properties are true.

1. A is diagonalizable.
2. All eigenvalues of A are real.
3. If λ is an eigenvalue of A with multiplicity k, then λ has k linearly independent eigenvectors. That is, the eigenspace of λ has dimension k.

- Property of Symmetric Matrices Let A be an $n \times n$ symmetric matrix. If λ_1 and λ_2 are distinct eigenvalues of A, then their corresponding eigenvectors $\mathbf{x}_1$ and $\mathbf{x}_2$ are orthogonal.

Proof

Let λ_1 and λ_2 be distinct eigenvalues of A with corresponding eigenvectors $\mathbf{x}_1$ and $\mathbf{x}_2$. So,

$$A\mathbf{x}_1 = \lambda_1 \mathbf{x}_1 \quad \text{and} \quad A\mathbf{x}_2 = \lambda_2 \mathbf{x}_2.$$

$$\langle \mathbf{x}_1, \mathbf{x}_2 \rangle = \mathbf{x}_1^T \mathbf{x}_2$$

Now you can write

$$\begin{aligned}
\lambda_1 \langle \mathbf{x}_1, \mathbf{x}_2 \rangle &= \langle \lambda_1 \mathbf{x}_1, \mathbf{x}_2 \rangle \\
&= (A\mathbf{x}_1) \cdot \mathbf{x}_2 = (A\mathbf{x}_1)^T \mathbf{x}_2 = \left(\mathbf{x}_1^T A^T \right) \mathbf{x}_2 \\
&= \left(\mathbf{x}_1^T A \right) \mathbf{x}_2 \quad \text{Because } A \text{ is symmetric, } A = A^T. \\
&= \mathbf{x}_1^T (A\mathbf{x}_2) = \mathbf{x}_1^T (\lambda_2 \mathbf{x}_2) \\
&= \langle \mathbf{x}_1, (\lambda_2 \mathbf{x}_2) \rangle = \lambda_2 \langle \mathbf{x}_1, \mathbf{x}_2 \rangle .
\end{aligned}$$

This implies that $(\lambda_1 - \lambda_2) \langle \mathbf{x}_1, \mathbf{x}_2 \rangle = 0$, and because $\lambda_1 \neq \lambda_2$ it follows that $\mathbf{x}_1 \cdot \mathbf{x}_2 = 0$. So, $\mathbf{x}_1$ and $\mathbf{x}_2$ are orthogonal.

Eigenvalue decomposition and diagonalization

A $p \times p$ symmetric matrix Σ can be diagonalized by $\Sigma = Q\Lambda Q^\top$

where Q is an orthogonal matrix, and Λ is a diagonal matrix, $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, ..., \lambda_p)$, where we order λ_j from largest to smallest in magnitude for $j = 1, ..., p$.

$$\Sigma = Q\Lambda Q^\top = Q \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p \end{bmatrix} Q^\top$$

$\Sigma Q = Q\Lambda$, so $\Sigma q_j = \lambda_j q_j$. The column vectors in Q are eigenvectors. The diagonal elements in Λ are eigenvalues of Σ.

Power Method

- Power Method is iterative. First we assume that the matrix A has a dominant eigenvalue with corresponding dominant eigenvectors.
- Then we choose an initial approximation $\mathbf{x}_0$ of one of the dominant eigenvectors of A. This initial approximation must be a nonzero vector in R^n.
- Finally we form the sequence given by

$$\begin{aligned}
\mathbf{x}_1 &= A\mathbf{x}_0 \\
\mathbf{x}_2 = A\mathbf{x}_1 = A\left(A\mathbf{x}_0\right) &= A^2\mathbf{x}_0 \\
\mathbf{x}_3 = A\mathbf{x}_2 = A\left(A^2\mathbf{x}_0\right) &= A^3\mathbf{x}_0 \\
&\vdots \\
\mathbf{x}_k = A\mathbf{x}_{k-1} = A\left(A^{k-1}\mathbf{x}_0\right) &= A^k\mathbf{x}_0.
\end{aligned}$$

Power Method: Example

Complete six iterations of the power method to approximate a dominant eigenvector of

$$A = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}$$

Rayleigh quotient

If $\mathbf{x}$ is an eigenvector of a matrix A, then its corresponding eigenvalue is given by

$$\lambda = \frac{A\mathbf{x} \cdot \mathbf{x}}{\mathbf{x} \cdot \mathbf{x}}.$$

This quotient is called the Rayleigh quotient.

Convergence of the Power Method

If A is an $n \times n$ diagonalizable matrix with a dominant eigenvalue, then there exists a nonzero vector $\mathbf{x}_0$ such that the sequence of vectors given by

$$A\mathbf{x}_0, \quad A^2\mathbf{x}_0, \quad A^3\mathbf{x}_0, \quad A^4\mathbf{x}_0, \ldots, A^k\mathbf{x}_0, \ldots$$

approaches a multiple of the dominant eigenvector of A.

Power Method

For a vector $\vec{v}$, let $\vec{u}$ be its **coordinates** in system Q, i.e. $\vec{v} = Q\vec{u}$

$$\vec{v} = Q\vec{u} = \begin{bmatrix} Q_1, Q_2, \cdots, Q_n \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = u_1 Q_1 + u_2 Q_2 + \cdots + u_n Q_n$$

if we left multiply $Q^\top$ on the two sides of the equation: $\vec{u} = Q^\top \vec{v}$

$u_i =< \vec{v}, Q_i >$

Power Method

$v = Qu \quad \Sigma = Q\Lambda Q^\top$

$$\Sigma v = Q\Lambda Q^\top Qu = Q \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_p \end{bmatrix} = Q \begin{bmatrix} \lambda_1 u_1 \\ \lambda_2 u_2 \\ \vdots \\ \lambda_p u_p \end{bmatrix},$$

which means the vector Σv becomes $(\lambda_1 u_1, \lambda_2 u_2, \cdots, \lambda_p u_p)^\top$ in basis Q, i.e., Σ is Λ in Q.

If we repeat this process n times, then Σ^n is Λ^n in Q,

$$v \xrightarrow{\Sigma^n} (\lambda_1^n u_1, \lambda_2^n u_2, \cdots, \lambda_p^n u_p)^\top.$$

We can keep normalizing $v \leftarrow v/|v|$ to make v a unit vector in this process.

Question: what happens if $n \to \infty$?

Power Method

Suppose λ_1 has the greatest magnitude, this procedure will converge to $u = (1, 0, \cdots, 0)$ in the space of Q, and the corresponding $v = q_1$.

The power method iterates the following two steps:

- Compute normalized vector $\tilde{v} = \dfrac{v}{|v|}$.
- Update $v = \Sigma \tilde{v}$.

Question: How to get q_2?

To get q_2 using this method, we initialize the above procedure with a vector $v \perp q_1$ that is perpendicular to q_1. In Q, the first component of u will always be 0, then the procedure will converge to $u = (0, 1, 0, \cdots, 0)$ in the space of Q, and the corresponding $v = q_2$.

Question: How to get q_3?

Power Method

To get q_3, we initialize the above procedure with v perpendicular to both q_1 and q_2, i.e. $v \perp q_1$ and $v \perp q_2$.

Continue the above procedure, we eventually get all the vectors in Q.

Matrix form of power method

We can parallelize the above sequential method, by starting from p vectors $V = (V_1, ..., V_p)$ and maintain their orthogonality after each multiplication by Σ, by iterating the following two step

- Compute $\tilde{V}$, the orthogonalized V.
- Update $V = \Sigma\tilde{V}$.

R code

```
myeigen <- function(A) {
T <- 1000
p <- nrow(A)
V <- matrix(rnorm(p*p), nrow = p)
for (i in 1:T)
  {
  V = myQR(V)$Q
  V = A %*% V
  }

B = myQR(V)
result <- list(eigen_values = diag(B$R), eigen_vectors = B$Q)
return(result) }
```

Python Code

```
def eigen_qr(A):
    T = 1000
    A_copy = A.copy()
    r, c = A_copy.shape

    V = np.random.random_sample((r, r))

    for i in range(T):
        Q, _ = qr(V)
        V = np.dot(A_copy, Q)
    Q, R = qr(V)
    return R.diagonal(), Q
```

Principle Component Analysis (PCA)

Consider the $n \times p$ data matrix $\mathbf{X}$. Let us assume that all the columns of $\mathbf{X}$ are centralized, i.e., $\sum_{i=1}^{n} x_{ij}/n = 0$.

In other words, let $\mathbf{1}$ be the $n \times 1$ column vector of 1's. Then $\langle \mathbf{X}_j, \mathbf{1} \rangle = 0$, for $j = 1, ..., p$, i.e., $\mathbf{1}^\top \mathbf{X} = 0$.

For each row of $\mathbf{X} = (X_1^\top, ..., X_n^\top)^\top$, we want to represent observation X_i in a new basis system Q, so that $X_i = QZ_i$.

Let $\mathbf{Z} = (Z_1^\top, ..., Z_n^\top)^\top$ be the data matrix in Q. We want the columns of $\mathbf{Z} = (\mathbf{Z}_1, ..., \mathbf{Z}_p)$ to be orthogonal to each other, so that they are uncorrelated.

If you regress any column of $\mathbf{Z}$ on another column of $\mathbf{Z}$, the regression coefficient is 0. Let $\lambda_j = \|\mathbf{Z}_j\|^2/n = \sum_{i=1}^{n} z_{ij}^2/n$, then λ_j is the variance of $\{z_{ij}, i = 1, ..., n\}$, and $\mathbf{Z}^\top \mathbf{Z} = \Lambda = \mathrm{diag}(\lambda_1, ..., \lambda_p)$. Then

$$\mathbf{X}^\top \mathbf{X} = Q\mathbf{Z}^\top \mathbf{Z} Q^\top = Q\Lambda Q^\top.$$