Eigenvalues and Eigenvectors

Thomas Siskos

Numerical Introductory Seminar
Humboldt–University Berlin
siskosth@student.hu-berlin.de
http://github.com/thsis/NIS18

Agenda

- Motivation
- Key Idea & Definitions
  - Characteristic Polynomial & Diagonal Matrices
  - Similarity Transformations
    - Householder Reflections
    - Givens Rotations
- Algorithms
  - Jacobi-Method
  - QR-Method
    - Basic Variant
    - Hessenberg Variant
    - Accelerated Variant
- Analysis
  - Accuracy
  - Efficiency

PCA

- The iris dataset is already linearly separable.
- With various techniques we can show this in even more detail.

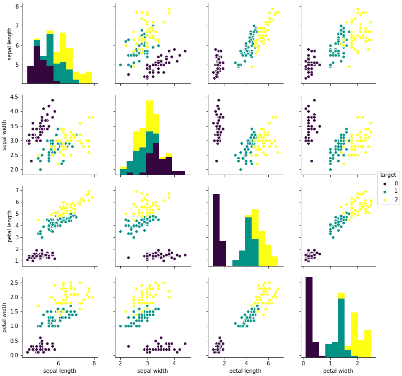


Figure 1: Iris Pairplot

PCA

- objective:

$$max\ \delta' Var(X)\,\delta\ s.t.\ \sum \delta_i^2 = 1. \quad (1)$$

  where $X \in \mathbb{R}^{n\times m}$; $m, n \in \mathbb{N}$; $\delta \in \mathbb{R}^m$

- solution ▸ Proof :

$$Y = \Gamma'(X - \mu) \quad (2)$$

  where $Y \in \mathbb{R}^{n\times m}$ is the matrix of rotations, $\Gamma \in \mathbb{R}^{m\times m}$ is the matrix of eigenvectors, $\mu \in \mathbb{R}^m$ is the vector of sample means.

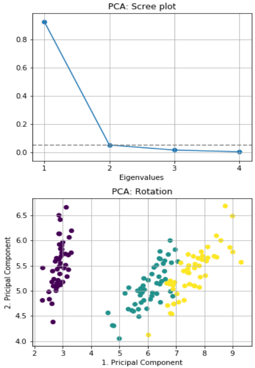


Figure 2: Iris PCA

LDA

- objective:

$$max \; \frac{w' S_B w}{w' S_W w}, \tag{3}$$

where

$$S_B = \sum_{c}^{C} (\mu_c - \mu)(\mu_c - \mu)',$$

$$S_W = \sum_{c}^{C} \sum_{i=1}^{n} (x_i - \mu_c)(x_i - \mu_c)'$$

and $x_i \in \mathbb{R}^m$, μ_c is the vector of class means.

LDA

- solution ▸ Proof:

$$S_B^{-\frac{1}{2}} S_W^{-1} S_B^{-\frac{1}{2}} w = \lambda w \qquad (4)$$

where this is again an Eigenvalue problem and it's solution will provide the rotation that ensures the largest possible (linear) separability.

- Now how do we get the Eigenvalues?

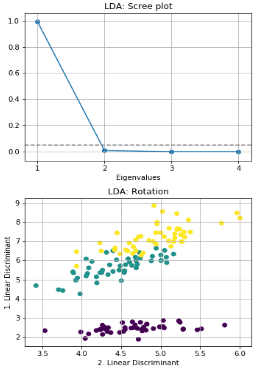


Figure 3: Iris LDA

Eigenvalue

If A is an $n \times n$ matrix, v is a non-zero vector and λ is a scalar, such that

$$Av = \lambda v \tag{5}$$

then v is called an *eigenvector* and λ is called an *eigenvalue* of the matrix A. An eigenvalue of A is a root of the characteristic equation,

$$det\,(A - \lambda I) = 0 \tag{6}$$

The characteristic polynomial

Consider the polynomial

$$f(\lambda) = \lambda^p + a_{p-1}\lambda^{p-1} + \cdots + a_1\lambda + a_0 \tag{7}$$

We now construct a matrix $A \in \mathbb{R}^{n\times n}$ such that the eigenvalues of A are the roots of the polynomial $f(\lambda)$ Example:

$$A = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ & & & \ddots & \\ 0 & 0 & 0 & \dots & 1 \\ -a_0 & -a_1 & -a_2 & \dots & -a_{p-1} \end{bmatrix} \tag{8}$$

General Idea

What are the eigenvalues of X_1 and X_2?

$$X_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix} \quad X_2 = \begin{bmatrix} 2.297 & -0.461 & -0.459 & 0.225 \\ -0.461 & 1.4 & -0.097 & -0.829 \\ -0.459 & -0.097 & 2.672 & 0.224 \\ 0.225 & -0.829 & 0.224 & 3.631 \end{bmatrix}$$

Similarity Transformations

Two $n \times n$ matrices A and B, are said to be *similar* if there exists a nonsingular matrix P such that

$$A = P^{-1}BP \tag{9}$$

If the two matrices A and B are similar, then they also share the same eigenvalues. Proof

Householder Reflections

Let u and v be orthonormal vectors and let x be a vector in the space spanned by u and v, such that

$$x = c_1 u + c_2 + v$$

for some scalars c_1 and c_2. The vector

$$\tilde{x} = -c_1 u + c_2 v$$

is a *reflection* of x through the line difined by the vector u. Now consider the matrix

$$Q = I - 2uu'. \tag{10}$$

Note that Proof:

$$Qx = \tilde{x}$$

Householder Reflections

We will use Householder-Reflections to tranform a vector

$$a = (a_1, \dots, a_n)$$

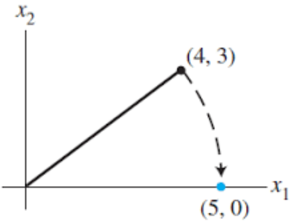
into

$$\hat{a} = (\hat{a}_1, 0, \dots, 0)$$

Givens Rotations

Using orthogonal transformations we can also rotate a vector in such a way that a specified element becomes 0 and only one other element in the vector is changed.

$$Q = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$



A Givens rotation in $\mathbb{R}^2$.

Givens Rotations

$$V_{pq}(\theta) = \begin{bmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ & & & \cos\theta & & \sin\theta & & & \\ & & & & \ddots & & & & \\ & & & -\sin\theta & & \cos\theta & & & \\ & & & & & & 1 & & \\ & & & & & & & \ddots & \\ & & & & & & & & 1 \end{bmatrix} \tag{11}$$

where $\cos\theta = \frac{x_p}{||x||}$ and $\sin\theta = \frac{x_q}{||x||}$

Jacobi Method

The Jacobi method for determining the eigenvalues of a symmetric matrix A uses a sequence of orthogonal similarity transformations that result in the transformation:

$$A = P \Lambda P^{-1}$$

or rather:

$$\Lambda = P^{-1} A P$$

where we use Givens Rotations to obtain P. The Jacobi iteration is:

$$A^k = V_{p_k q_k}(\theta_k) A^{k-1} V_{p_k q_k}(\theta_k) \tag{12}$$

The Jacobi Method is of $O(n^3)$

Jacobi-Method

Algorithm 1 `jacobi`

Require: symmetric matrix A

Ensure: $0 < precision < 1$

initialize: $L \leftarrow A$; $U \leftarrow I$; $L_{max} \leftarrow 1$

while $L_{max} > precision$ **do**

Find indices i, j of largest value in lower triangle of $abs(L)$

$L_{max} \leftarrow L_{i,j}$

$\alpha \leftarrow \frac{1}{2} \cdot \arctan(\frac{2A_{i,j}}{A_{i,i} - A_{j,j}})$

$V \leftarrow I$

$V_{i,i}, V_{j,j} \leftarrow \cos\alpha$; $V_{i,j}, V_{j,i} \leftarrow -\sin\alpha, \sin\alpha$

$A \leftarrow V'AV$; $U \leftarrow UV$

return $diag(A), U$

QR-Method

The QR-Method is the most common algorithm for obtaining eigenvalues and eigenvectors of a matrix A. It relies on the so called QR-Factorization:

$$A = QR, \tag{13}$$

where Q is an orthogonal and R is an upper triangular matrix. The QR iteration is:

$$A^k = Q'_{k-1} A_{k-1} Q_{k-1} = R_{k-1} Q_{k-1} \tag{14}$$

The QR Method is of $O(n^3)$

Basic QR-Method

Algorithm 2 `QRM1`

Require: square matrix A
  initialize: $\mathit{conv} \leftarrow \mathit{False}$
  while not conv **do**
    $Q, R \leftarrow$ QR-Factorization of A
    $A \leftarrow RQ$
    if A is diagonal **then**
      $\mathit{conv} \leftarrow$ `True`

  return $\mathit{diag}\,(A)\,, Q$

Refined QR-Method

For faster convergence it is common to convert the matrix first into a so called upper Hessenberg form.

$$\begin{bmatrix} X & X & X & X & X & X & X \\ X & X & X & X & X & X & X \\ 0 & X & X & X & X & X & X \\ 0 & 0 & X & X & X & X & X \\ 0 & 0 & 0 & X & X & X & X \\ 0 & 0 & 0 & 0 & X & X & X \\ 0 & 0 & 0 & 0 & 0 & X & X \end{bmatrix}$$

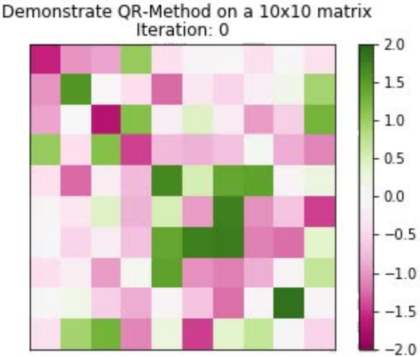
Algorithm 3 `QRM2`

Require: square matrix A

$A \leftarrow \texttt{hessenberg}(A)$

continue with: QRM1(A)

QRM2 Visualized



QR-Method

Accelerated QR-Method

We can accelerate the QR-Method by creating an artificial zero on the main diagonal of A^ks Hessenberg form T at an iteration step k:

$$
\begin{aligned}
T^\star &= T - t_{p-1,p-1} I \\
T^\star &= QR \\
T &= T^\star + t_{p-1,p-1} I
\end{aligned}
$$

Accelerated QR-Method

Require: square matrix $A \in \mathbb{R}^{p \times p}$

$T \leftarrow \texttt{hessenberg}(A),\ \ conv \leftarrow False$

while not *conv* **do**

$Q, R \leftarrow$ QR-Factorization of $T - t_{p-1,p-1} I$

$T \leftarrow RQ + t_{p-1,p-1} I$

if T is diagonal **then**

$conv \leftarrow True$

return $diag\,(T)\,, Q$

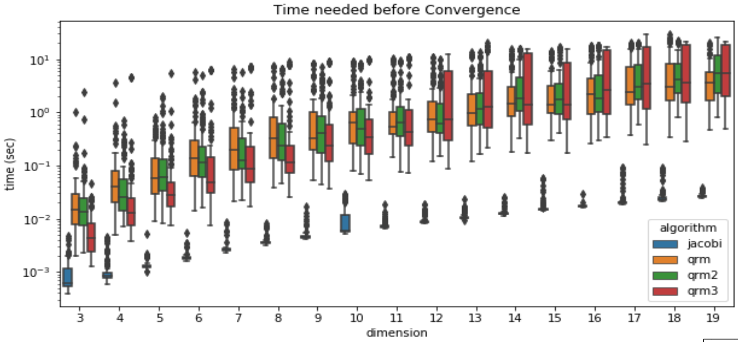
Unit tests - Idea

1. Construct a $p \times p$ matrix, with known Eigenvalues $\lambda_{true} \in \mathbb{R}^p$. To do this we can invert the spectral decomposition.
2. Run the implemented algorithm on it, obtain the computed Eigenvalues $\lambda_{algo} \in \mathbb{R}^p$.
3. Assess L_1-Norm: $|\lambda_{true} - \lambda_{algo}|$, pass the test if it is smaller than a threshold ϵ

Repeat the procedure 1000 times.

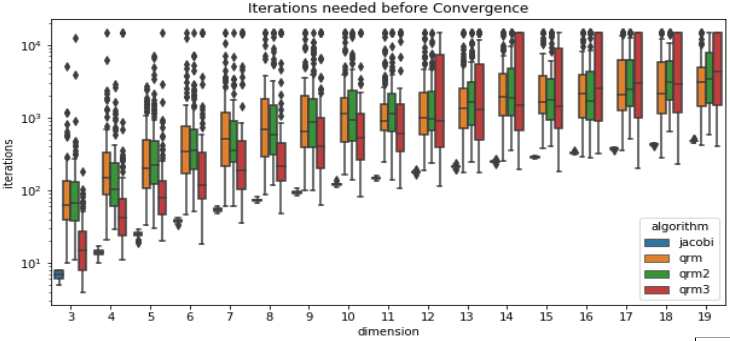
Time taken

Figure 4: Unit-tests: Time



Iterations needed

Figure 5: Unit-tests: Iterations



PCA: proof

The objective of PCA:

$$max\ \delta' Var(X)\,\delta\ s.t.\ \sum \delta_i^2 = 1$$

Corresponding Lagrangean:

$$\mathcal{L}(Var(X), \delta, \lambda) = \delta' Var(X)\,\delta - \lambda\left(\delta'\delta - 1\right),$$

where $\lambda \in \mathbb{R}^m$
First order condition:

$$\frac{\partial \mathcal{L}}{\partial \delta} \stackrel{!}{=} 0$$

$$2Var(X)\delta - 2\lambda_k\delta \stackrel{!}{=} 0$$

$$Var(X)\delta = \lambda_k\delta$$

Which is now reduced to a common Eigenvalue problem.

LDA: proof

The objective of LDA:

$$max \ \frac{w' S_B w}{w' S_W w},$$

Which we can reformulate to:

$$max w' S_B w \ s.t. w' S_W w = 1.$$

Corresponding Lagrangean:

$$\mathcal{L}(w, S_B, S_W, \lambda) = w' S_B w - \lambda \left(w' S_W w - 1 \right)$$

LDA: proof

First order condition:

$$\frac{\partial \mathcal{L}}{\partial w} \overset{!}{=} 0$$

$$2S_B w - 2\lambda S_W w \overset{!}{=} 0$$

$$S_W^{-1} S_B w = \lambda w,$$

which is known as a generalized Eigenvalue problem. We can redefine

$$S_B = S_B^{\frac{1}{2}} S_B^{\frac{1}{2}}$$

$$v = S_B^{\frac{1}{2}} w$$

LDA: proof

We then get:

$$S_W^{-1} S_B w = \lambda w$$

$$S_W^{-1} S_B^{\frac{1}{2}} \underbrace{S_B^{\frac{1}{2}} w}_{v} = \lambda w$$

$$S_B^{\frac{1}{2}} S_W^{-1} S_B^{\frac{1}{2}} v = \lambda \underbrace{S_B^{\frac{1}{2}} w}_{v}$$

We can also rewrite this as:

$$S_B^{-\frac{1}{2}} S_W^{-1} S_B^{-\frac{1}{2}} v = \lambda v$$

Which now an Eigenvalue problem of a symmetric, positive semidefinite matrix ▸ back.

Eigenvalues of similar matrices

From the definition in (9) it follows immediately that a matrix A with Eigenvalues $\lambda_1, \ldots, \lambda_n$ is similar to the matrix $diag(\lambda_1, \ldots, \lambda_n)$.
If A and B are similar, as in (9), it holds:

$$\begin{aligned} B - \lambda I &= P^{-1}BP - \lambda P^{-1}IP \\ &= A - \lambda I. \end{aligned}$$

Hence A and B have the same eigenvalues. Additionally, important transformations are based around orthogonal matrices. If Q is orthogonal and

$$A = Q'BQ,$$

A and B are said to be *orthogonally similar* back.

Companion-Matrix: Example

Demonstrate that the companion matrix back:

1. corresponds to a polynomial.
2. has eigenvalues equal to the roots of the polynomial.

$$A = \begin{bmatrix} 0 & 1 \\ -a_0 & -a_1 \end{bmatrix}$$

$$\begin{aligned} det\left(A - \lambda I\right) &= \begin{bmatrix} 0 - \lambda & 1 \\ -a_0 & -a_1 - \lambda \end{bmatrix} \\ &= -\lambda\left(-a_1 - \lambda\right) + a_0 \\ &= \lambda^2 + a_1\lambda + a_0 \end{aligned}$$

Householder Reflections: proof

back Remember:

- $Q = I - 2uu'$.
- u v are orthonormal.

$$
\begin{aligned}
Qx &= c_1 u + c_2 v - 2c_1 uuu' - 2c_2 vuu' \\
&= c_1 u + c_2 v - 2c_1 u'uu - 2c_2 u'vu \\
&= -c_1 u + c_2 v \\
&= \tilde{x}
\end{aligned}
$$

Sources

- Seffen Börm and Christian Mehl. Numerical Methods for Eigenvalue Problems. Walter de Gruyter GmbH & Co.KG, Berlin/Boston, 2012.
- James E. Gentle. Numerical Linear Algebra for Applications in Statistics. Springer Science + Business Media, New York, 2003.
- Wolfgang K. Härdle and Léopold Simar. Applied Multivariate Statistical Analysis. Springer-Verlag Gmbh, Berlin, Heidelberg, 2015.