Knowledge Discovery and Data Mining 1 (VO) (706.701)

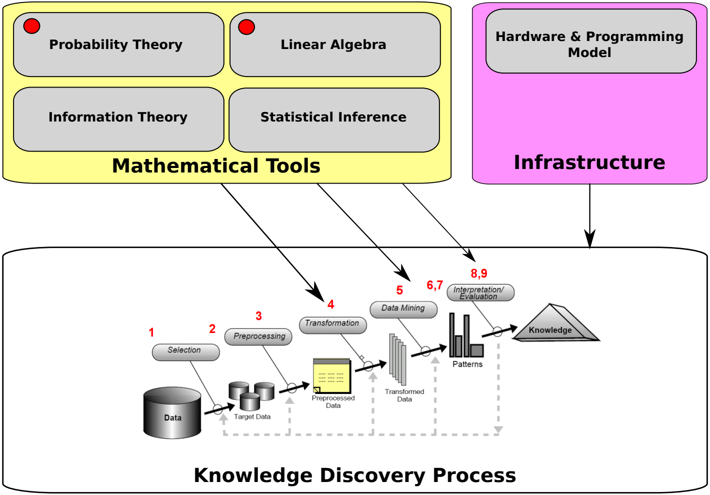
Principal Component Analysis

Denis Helic

ISDS, TU Graz

May 3, 2020

Big picture: KDDM



Outline

Recap — Linear Algebra

Eigenvalues and eigenvectors

Given a square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, we say that $\lambda \in \mathbb{C}$ is an eigenvalue of $\mathbf{A}$ and $\mathbf{x} \in \mathbb{C}^n$ is the corresponding eigenvector if

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{x}, \mathbf{x} \neq \mathbf{0}$$

Recap — Linear Algebra

- Symmetric matrices: all eigenvalues are real, eigenvectors are orthonormal and form an eigenbasis
- We can decompose symmetric $\mathbf{A}$ as $\mathbf{Q\Lambda Q}^T$
- $\mathbf{Q}$ is an orthonormal matrix $(\mathbf{QQ}^T = \mathbf{Q}^T\mathbf{Q} = \mathbf{I})$ where colums are the eigenvectors
- $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n)$
- If we can factorize symmetric $\mathbf{A} = \mathbf{B}^T\mathbf{B}$ then $\mathbf{A}$ is positive semi-definite and all eigenvalues are non-negative

Eigenvalues and eigenvectors

Eigenvalues and eigenvectors

Given a square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, we say that $\lambda \in \mathbb{C}$ is an eigenvalue of $\mathbf{A}$ and $\mathbf{x} \in \mathbb{C}^n$ is the corresponding eigenvector if

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}, \mathbf{x} \neq \mathbf{0}$$

Interpretation of eigenvalues and eigenvectors

Example

Suppose we have a Web-based question-answer system, where standard users (S) post questions about certain topics expert (E) and other users answer those questions. For example, Stackoverflow is an example of such a system. The questions are about programming, software development, and so on. We observe the following developments with our system:

- The system is very successful and we observe increasing numbers of both standard (S) as well as expert users (E). Each month we have 4x more (S) users and 2x more (E) users.
- The numbers of these two different types of users seem to be *decoupled* because we can not observe that they affect each other.

Interpretation of eigenvalues and eigenvectors

Example

This *decoupling* may arise in the following situation:

- The standard users (S) talk only to each other: “Hey, there is this great system where you get all the answers from experts very quickly”.
- The expert users (E) also talk only to each other: “Hey, there is this great system where you can show what you know and become popular very quickly”.
- (S) and (E) do not talk to each other and do not influence each other.

Note

Such a system is completely hypothetical and is not realistic, but illustrates very well the educational point.

Interpretation of eigenvalues and eigenvectors

Question

How will our system develop? How many standard users (S) we will have in e.g. two, three, .. years. How many expert users (E), and hence how many users in total.

Model

Let us denote the number of standard users (S) with x and the number of expert users (E) with y, or in matrix notation:

$$\mathbf{u} = \begin{pmatrix} x \\ y \end{pmatrix}$$

Interpretation of eigenvalues and eigenvectors

Model

We know from the data that x increases four times each month and y increases two times each month:

$$\begin{aligned} x(t+1) &= 4x(t) \\ y(t+1) &= 2y(t), \end{aligned}$$

where $x(t)$ is the number of e.g. standard users in month t and $x(t+1)$ is the number of standard users next month.

Interpretation of eigenvalues and eigenvectors

- How does the system evolve?

$$\begin{aligned} x(1) &= 4 \cdot x(0) \\ y(1) &= 2 \cdot y(0) \end{aligned}$$

$$\begin{aligned} x(2) &= 4 \cdot x(1) = 4 \cdot 4 \cdot x(0) = 4^2 \cdot x(0) \\ y(2) &= 2 \cdot y(1) = 2 \cdot 2 \cdot y(0) = 2^2 \cdot y(0) \end{aligned}$$

Interpretation of eigenvalues and eigenvectors

$$\begin{aligned} x(3) &= 4 \cdot x(2) = 4 \cdot 4^2 \cdot x(0) = 4^3 \cdot x(0) \\ y(3) &= 2 \cdot y(2) = 2 \cdot 2^2 \cdot y(0) = 2^3 \cdot y(0) \end{aligned}$$

$$\begin{aligned} x(t+1) &= 4 \cdot x(t) = 4 \cdot 4^t \cdot x(0) = 4^{t+1} \cdot x(0) \\ y(t+1) &= 2 \cdot y(t) = 2 \cdot 2^t \cdot y(0) = 2^{t+1} \cdot y(0) \end{aligned}$$

Interpretation of eigenvalues and eigenvectors

- How does the system evolve?
- Suppose we have the following initial conditions: we have a single standard user (S) and a single expert user (E) in the beginning.
- How the numbers of standard and expert users compare after e.g. five years?

$$
\begin{aligned}
x(0) &= 1 \\
y(0) &= 1
\end{aligned}
$$

$$
\begin{aligned}
x(60) &= 4^{60} \cdot x(0) = 4^{60} \\
y(60) &= 2^{60} \cdot y(0) = 2^{60}
\end{aligned}
$$

Interpretation of eigenvalues and eigenvectors

- The ratio: $\frac{x}{y}$

$$\frac{x^{60}}{y^{60}} = \frac{4^{60}}{2^{60}} = 2^{60}$$

- 2^{60} is a huge number, i.e. standard users (S) completely dominate the expert users (E)
- The total number of users is approx. equal to the number of standard users (S)

Interpretation of eigenvalues and eigenvectors

- But, how is this related to the eigenvalues and eigenvectors?

Model

We can express the previous equations in the matrix form:

$$\mathbf{u}(t+1) = \mathbf{A}\mathbf{u}(t)$$

Interpretation of eigenvalues and eigenvectors

$$\mathbf{A} = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}$$

$$\mathbf{u} = \begin{pmatrix} x \\ y \end{pmatrix}$$

Interpretation of eigenvalues and eigenvectors

$$\mathbf{A} = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}$$

$$det(\lambda \mathbf{I} - \mathbf{A}) = det(\begin{pmatrix} \lambda - 4 & 0 \\ 0 & \lambda - 2 \end{pmatrix}) = (\lambda - 4)(\lambda - 2)$$

- Thus, $\lambda_1 = 4$, and $\lambda_2 = 2$ are eigenvalues of $\mathbf{A}$
- **The eigenvalues of a diagonal matrix are equal to its diagonal entries**
- We now solve $\mathbf{Ax} = \lambda \mathbf{x}$ for each eigenvalue to find the corresponding eigenvectors

Interpretation of eigenvalues and eigenvectors

- For $\lambda_1 = 4$

$$\begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 4 \begin{pmatrix} x \\ y \end{pmatrix}$$

$$\begin{aligned} 4x &= 4x \\ y &= 4y \end{aligned}$$

- Thus, $y = 0$ and we might pick $x = 1$, i.e. $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

Interpretation of eigenvalues and eigenvectors

- For $\lambda_1 = 2$

$$\begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 2 \begin{pmatrix} x \\ y \end{pmatrix}$$

$$\begin{aligned} 4x &= x \\ y &= y \end{aligned}$$

- Thus, $x = 0$ and we might pick $y = 1$, i.e. $\mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$
- **The eigenvectors of a diagonal matrix form a standard basis for a Euclidean space**

Interpretation of eigenvalues and eigenvectors

- But, how is this related to the eigenvalues and eigenvectors?

Model

We can express the previous equations in the matrix form:

$$\mathbf{u}(t+1) = \mathbf{A}\mathbf{u}(t)$$

Interpretation of eigenvalues and eigenvectors

- How does the system evolve (in matrix form)?

$$\mathbf{u}(1) = \mathbf{A}\mathbf{u}(0)$$

$$\mathbf{u}(2) = \mathbf{A}\mathbf{u}(1) = \mathbf{A}\mathbf{A}\mathbf{u}(0) = \mathbf{A}^2\mathbf{u}(0)$$

$$\mathbf{u}(3) = \mathbf{A}\mathbf{u}(2) = \mathbf{A}\mathbf{A}^2\mathbf{u}(0) = \mathbf{A}^3\mathbf{u}(0)$$

$$\mathbf{u}(t+1) = \mathbf{A}\mathbf{u}(t) = \mathbf{A}\mathbf{A}^t\mathbf{u}(0) = \mathbf{A}^{t+1}\mathbf{u}(0)$$

Interpretation of eigenvalues and eigenvectors

- Since the eigenvectors form a basis for the space
- We can write $\mathbf{u}(0)$ as a linear combination of the eigenvectors $\mathbf{v}_i$ of the matrix (for appropriate choice of constants c_i)
- $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$
- $\mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

$$\mathbf{u}(0) = c_1\mathbf{v}_1 + c_2\mathbf{v}_1$$

Interpretation of eigenvalues and eigenvectors

- We have the following initial conditions: we have a single standard user (S) and a single expert user (E) in the beginning.

$$\mathbf{u}(0) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$\mathbf{u}(0) = \mathbf{v}_1 + \mathbf{v}_2$$

Interpretation of eigenvalues and eigenvectors

- We know from before:

$$\mathbf{u}(t+1) = \mathbf{A}^{t+1}\mathbf{u}(0)$$

$$\mathbf{u}(t+1) = \mathbf{A}^{t+1}(\mathbf{v}_1 + \mathbf{v}_2) = \mathbf{A}^{t+1}\mathbf{v}_1 + \mathbf{A}^{t+1}\mathbf{v}_2$$

Interpretation of eigenvalues and eigenvectors

- We have: $\mathbf{A}\mathbf{v}_1 = \lambda_1\mathbf{v}_1$ and $\mathbf{A}\mathbf{v}_2 = \lambda_1\mathbf{v}_2$
- By substituting:

$$\mathbf{u}(t+1) = \mathbf{A}^{t+1}\mathbf{v}_1 + \mathbf{A}^{t+1}\mathbf{v}_2 = \lambda_1^{t+1}\mathbf{v}_1 + \lambda_2^{t+1}\mathbf{v}_2$$

$$\mathbf{u}(t+1) = 4^{t+1}\begin{pmatrix}1\\0\end{pmatrix} + 2^{t+1}\begin{pmatrix}0\\1\end{pmatrix}$$

Interpretation of eigenvalues and eigenvectors

- After five years:

$$\mathbf{u}(60) = 4^{60}\begin{pmatrix}1\\0\end{pmatrix} + 2^{60}\begin{pmatrix}0\\1\end{pmatrix}$$

- 4^{60} is a much larger number than 2^{60}
- The first term dominates the second
- Standard users (S) completely dominate the expert users (E)

Interpretation of eigenvalues and eigenvectors

- The system evolves along the directions of the matrix eigenvectors
- The eigenvalues determine the speed of the development
- Larger eigenvalues represent quicker development and dominant behavior
- In the long run the system develops in the direction of the eigenvector corresponding to the largest eigenvalue
- This is the leading eigenvector and leading eigenvalue

Interpretation of eigenvalues and eigenvectors

- Another example: a *coupled* system (more realistic)
- The number of standard users (S) is influenced also by the number of expert users (E)
- I.e. a famous expert joins the system and as a consequence a lot of standard users joins as well
- Also, the number of expert users is influenced by the number of standard users
- Experts like to have an audience and are therefore attracted to the system

Interpretation of eigenvalues and eigenvectors

Model

Let say that we know from the data how x and y increase each month. For example:

$$\begin{aligned} x(t+1) &= 3x(t) + 6y(t) \\ y(t+1) &= x(t) + 4y(t), \end{aligned}$$

where $x(t)$ is the number of e.g. standard users in month t and $x(t+1)$ is the number of standard users next month.

Interpretation of eigenvalues and eigenvectors

- Again, in matrix form:

$$\mathbf{u}(t+1) = \mathbf{A}\mathbf{u}(t)$$

$$\mathbf{A} = \begin{pmatrix} 3 & 6 \\ 1 & 4 \end{pmatrix}$$

$$\mathbf{u} = \begin{pmatrix} x \\ y \end{pmatrix}$$

Interpretation of eigenvalues and eigenvectors

$$\mathbf{A} = \begin{pmatrix} 3 & 6 \\ 1 & 4 \end{pmatrix}$$

$$\begin{aligned} det(\lambda\mathbf{I} - \mathbf{A}) &= det(\begin{pmatrix} \lambda - 3 & -6 \\ -1 & \lambda - 4 \end{pmatrix}) = (\lambda - 3)(\lambda - 4) - (-1)(-6) \\ &= \lambda^2 - 3\lambda - 4\lambda + 12 - 6 = \lambda^2 - 7\lambda + 6 \\ &= (\lambda - 6)(\lambda - 1) \end{aligned}$$

- Thus, $\lambda_1 = 6$, and $\lambda_2 = 1$ are eigenvalues of $\mathbf{A}$
- We now solve $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ for each eigenvalue to find the corresponding eigenvectors

Interpretation of eigenvalues and eigenvectors

- For $\lambda_1 = 6$

$$\begin{pmatrix} 3 & 6 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 6 \begin{pmatrix} x \\ y \end{pmatrix}$$

$$\begin{aligned} 3x + 6y &= 6x \implies 3x = 6y \implies x = 2y \\ x + 4y &= 6y \implies x = 2y \end{aligned}$$

- Thus, we might pick $y = 1$ and then $x = 2$, i.e. $\mathbf{v}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$

Interpretation of eigenvalues and eigenvectors

- For $\lambda_1 = 1$

$$\begin{pmatrix} 3 & 6 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}$$

$$\begin{aligned} 3x + 6y &= x \implies 2x = -6y \implies x = -3y \\ x + 4y &= y \implies x = -3y \end{aligned}$$

- Thus, we might pick $y = 1$ and then $x = -3$, i.e. $\mathbf{v}_2 = \begin{pmatrix} -3 \\ 1 \end{pmatrix}$

Interpretation of eigenvalues and eigenvectors

- We know that he system evolves along the directions of the matrix eigenvectors
- The eigenvalues determine the speed of the development
- Larger eigenvalues represent quicker development and dominant behavior
- In the long run the system develops in the direction of the eigenvector corresponding to the largest eigenvalue
- This is the leading eigenvector and leading eigenvalue

Interpretation of eigenvalues and eigenvectors

- We can write $\mathbf{u}(0)$ as a linear combination of the eigenvectors $\mathbf{v}_i$ of the matrix (for appropriate choice of constants c_i)
- $\mathbf{v}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$
- $\mathbf{v}_2 = \begin{pmatrix} -3 \\ 1 \end{pmatrix}$

$$\mathbf{u}(0) = c_1\mathbf{v}_1 + c_2\mathbf{v}_1$$

Interpretation of eigenvalues and eigenvectors

- We have the following initial conditions: we have a single standard user (S) and a single expert user (E) in the beginning.

$$\mathbf{u}(0) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{4}{5}\begin{pmatrix} 2 \\ 1 \end{pmatrix} + \frac{1}{5}\begin{pmatrix} -3 \\ 1 \end{pmatrix}$$

$$\mathbf{u}(0) = \frac{4}{5}\mathbf{v}_1 + \frac{1}{5}\mathbf{v}_2$$

Interpretation of eigenvalues and eigenvectors

- As before we have:

$$\mathbf{u}(t+1) = \frac{4}{5}\mathbf{A}^{t+1}\mathbf{v}_1 + \frac{1}{5}\mathbf{A}^{t+1}\mathbf{v}_2 = \frac{4}{5}\lambda_1^{t+1}\mathbf{v}_1 + \frac{1}{5}\lambda_2^{t+1}\mathbf{v}_2$$

$$\mathbf{u}(t+1) = \frac{4}{5}6^{t+1}\begin{pmatrix}2\\1\end{pmatrix} + \frac{1}{5}\begin{pmatrix}-3\\1\end{pmatrix}$$

Interpretation of eigenvalues and eigenvectors

- After five years:

$$\mathbf{u}(60) = \frac{4}{5}6^{60}\begin{pmatrix} 2 \\ 1 \end{pmatrix} + \frac{1}{5}\begin{pmatrix} -3 \\ 1 \end{pmatrix}$$

- The first term dominates the second
- There are twice as many standard users (S) as expert users (E)

Advanced: Power Method

- Since the eigenvectors form a basis for the vector space (e.g. in case of a symmetric matrix)
- We can write $\mathbf{u}(0)$ as a linear combination of the eigenvectors $\mathbf{v}_i$ of the matrix (for appropriate choice of constants c_i)

$$\mathbf{u}(0) = \sum_i c_i \mathbf{v}_i$$

$$\mathbf{u}(t) = \mathbf{A}^t \sum_i c_i \mathbf{v}_i = \sum_i c_i \mathbf{A}^t \mathbf{v}_i = \sum_i c_i \lambda_i^t \mathbf{v}_i = \lambda_1^t \sum_i c_i \left[\frac{\lambda_i}{\lambda_1} \right]^t \mathbf{v}_i$$

Advanced: Power Method

$$\mathbf{u}(t) = \lambda_1^t \sum_i c_i \left[\frac{\lambda_i}{\lambda_1}\right]^t \mathbf{v}_i$$

- λ_i are eigenvalues, and λ_1 is the largest of themselves
- $\frac{\lambda_i}{\lambda_1} < 1$ for all $i > 1$
- When $t \to \infty$ $\frac{\lambda_i}{\lambda_1} \to 0$, for all $i > 1$
- When $t \to \infty$ $\mathbf{u}(t) \to c_1 \lambda_1^t \mathbf{v}_1$

Advanced: Power Method

- In other words, the limiting behavior of the system is proportional to the leading eigenvector of the matrix
- The system evolves in the direction of the eigenvectors
- However, the leading eigenvector dominates all other eigenvectors
- The limiting behavior is therefore dominated by the leading eigenvalue and the leading eigenvector

Advanced: Power Method

- Typically, we would calculate the leading eigenvector and leading eigenvalue iteratively
- A standard approach is the power method
- We make an initial guess about the eigenvector $\mathbf{x}^0$
- Then we iteratively calculate $\mathbf{x}^t$ (which converges to the leading eigenvector)

$$\mathbf{x}^t = \frac{\mathbf{A}\mathbf{x}^{(t-1)}}{||\mathbf{A}\mathbf{x}^{(t-1)}||_2}$$

Advanced: Power Method

- In other words, the limiting vector is approximately equal the leading eigenvector of the matrix
- At the end of the iteration the leading (principal) eigenvalue can be calculated as:

$$\lambda_1 = \mathbf{x}^T \mathbf{A} \mathbf{x}$$

Advanced: Power Method

- To find the second eigenpair we create a new matrix $\mathbf{A}^* = \mathbf{A} - \lambda_1 \mathbf{x}\mathbf{x}^T$
- We then again use the power iteration to calculate the leading eigenpair of $\mathbf{A}^*$
- This leading eigenpair corresponds to the second largest eigenpair of the original matrix $\mathbf{A}$
- Intuitively, we have eliminated the influence of a given eigenvector by setting its associated eigenvalue to zero

Advanced: Power Method

- More formally, if $\mathbf{A}^* = \mathbf{A} - \lambda_1 \mathbf{x}\mathbf{x}^T$ where λ_1 is the leading eigenvalue of $\mathbf{A}$ and $\mathbf{x}$ is the leading eigenvector of $\mathbf{A}$ then
  1. $\mathbf{x}$ is also an eigenvector of $\mathbf{A}^*$ where the corresponding eigenvalue is 0.
  2. If $\mathbf{v}$ and λ_v are eigenpair of $\mathbf{A}$ other then the principal eigenpair that they are also an eigenpair of $\mathbf{A}^*$

Advanced: Power Method

Proof.

- We assume that $\mathbf{A}$ is a symmetric matrix

1. $\mathbf{A}^*\mathbf{x} = (\mathbf{A} - \lambda_1 \mathbf{x}\mathbf{x}^T)\mathbf{x} = \mathbf{A}\mathbf{x} - \lambda_1 \mathbf{x}\mathbf{x}^T\mathbf{x} = \mathbf{A}\mathbf{x} - \lambda_1 \mathbf{x} = \mathbf{0} = 0\mathbf{x}$
2. $\mathbf{A}^*\mathbf{v} = (\mathbf{A}^*)^T\mathbf{v} = (\mathbf{A} - \lambda_1 \mathbf{x}\mathbf{x}^T)^T\mathbf{v} = \mathbf{A}^T\mathbf{v} - \lambda_1 \mathbf{x}\mathbf{x}^T\mathbf{v} = \mathbf{A}^T\mathbf{v} = \mathbf{A}\mathbf{v} = \lambda_v\mathbf{v}$

$\square$