The Resilience and Robustness (and Hopefully Responsibility) of Linear Algebra Lecture 4

Emily J. King

Mathematics Department and Data Science Research Institute
Colorado State University

IK Interdisciplinary College
March 2024

Course Outline

- Lecture 1
  - Signals and vector spaces
  - Linear combinations
- Lecture 2
  - Inner product
  - Inner product as similarity
  - Moving average
  - Convolution and cross correlation
  - **2*D*** convolution and cross correlation
- Lecture 3
  - Cosine similarity
  - Projection
  - Change of basis
  - Graph theory
- **Lecture 4**
  - **Eigenvalues and eigenvectors**
  - **Principal component analysis**
  - **Responsibility vignette**

Lecture Outline

Eigenvalues and eigenvectors

Principal component analysis

Responsibility vignette

What is an eigenvector? What is an eigenvalue?

Assume that **90%** of the time, a sunny day follows another sunny day and **50%** of the time a gray day follows a gray day. We may represent this situation with a transition matrix[1]:

$$\mathbf{P} = \begin{pmatrix} \mathbf{9/10} & \mathbf{1/2} \\ \mathbf{1/10} & \mathbf{1/2} \end{pmatrix}.$$

[1] also, left stochastic matrix, Markov matrix

Assume that **90%** of the time, a sunny day follows another sunny day and **50%** of the time a gray day follows a gray day. We may represent this situation with a transition matrix[1]:

$$\mathbf{P} = \begin{pmatrix} \mathbf{9/10} & \mathbf{1/2} \\ \mathbf{1/10} & \mathbf{1/2} \end{pmatrix}.$$

Today is sunny, $\begin{pmatrix} \mathbf{1} \\ \mathbf{0} \end{pmatrix}$. So,

[1] also, left stochastic matrix, Markov matrix

Assume that **90%** of the time, a sunny day follows another sunny day and **50%** of the time a gray day follows a gray day. We may represent this situation with a transition matrix[1]:

$$\mathbf{P} = \begin{pmatrix} \mathbf{9/10} & \mathbf{1/2} \\ \mathbf{1/10} & \mathbf{1/2} \end{pmatrix}.$$

Today is sunny, $\begin{pmatrix} \mathbf{1} \\ \mathbf{0} \end{pmatrix}$. So,

$$\mathbf{P}\begin{pmatrix} \mathbf{1} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{9/10} \\ \mathbf{1/10} \end{pmatrix}$$

encodes the weather probabilities for tomorrow.

[1] also, left stochastic matrix, Markov matrix

Note that

$$\mathbf{P}^2 \begin{pmatrix} \mathbf{1} \\ \mathbf{0} \end{pmatrix} = \mathbf{P} \begin{pmatrix} \mathbf{9/10} \\ \mathbf{1/10} \end{pmatrix} = \begin{pmatrix} \mathbf{0.86} \\ \mathbf{0.14} \end{pmatrix}$$

are the probabilities for two days from now.

Note that

$$\mathbf{P}^2 \begin{pmatrix} \mathbf{1} \\ \mathbf{0} \end{pmatrix} = \mathbf{P} \begin{pmatrix} \mathbf{9/10} \\ \mathbf{1/10} \end{pmatrix} = \begin{pmatrix} \mathbf{0.86} \\ \mathbf{0.14} \end{pmatrix}$$

are the probabilities for two days from now.

Extending this,

$$\mathbf{P}^{\boldsymbol{n}} \begin{pmatrix} \mathbf{1} \\ \mathbf{0} \end{pmatrix} = \text{ probabilities for } \boldsymbol{n} \text{ days from now.}$$

Note that

$$\mathbf{P}^2 \begin{pmatrix} \mathbf{1} \\ \mathbf{0} \end{pmatrix} = \mathbf{P} \begin{pmatrix} \mathbf{9/10} \\ \mathbf{1/10} \end{pmatrix} = \begin{pmatrix} \mathbf{0.86} \\ \mathbf{0.14} \end{pmatrix}$$

are the probabilities for two days from now.

Extending this,

$$\mathbf{P}^{\boldsymbol{n}} \begin{pmatrix} \mathbf{1} \\ \mathbf{0} \end{pmatrix} = \text{ probabilities for } \boldsymbol{n} \text{ days from now.}$$

What’s the long term behavior?

Similarly, if today were gray, the probabilities for tomorrow would be

$$\mathbf{P}\begin{pmatrix} \mathbf{0} \\ \mathbf{1} \end{pmatrix} = \begin{pmatrix} \mathbf{1/2} \\ \mathbf{1/2} \end{pmatrix}$$

and for n days from now

$$\mathbf{P}^n \begin{pmatrix} \mathbf{0} \\ \mathbf{1} \end{pmatrix}.$$

An oracle gives you special vectors and tells you to multiply them by $\mathbf{P}$:

$$\mathbf{P}\begin{pmatrix} \mathbf{5/6} \\ \mathbf{1/6} \end{pmatrix} = \begin{pmatrix} \mathbf{5/6} \\ \mathbf{1/6} \end{pmatrix} \quad \text{and}$$

$$\mathbf{P}\begin{pmatrix} \mathbf{1/6} \\ \mathbf{-1/6} \end{pmatrix} = \mathbf{(2/5)}\begin{pmatrix} \mathbf{1/6} \\ \mathbf{-1/6} \end{pmatrix}.$$

Also note that

$$\vec{x}_0 := \begin{pmatrix} \mathbf{1} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{5/6} \\ \mathbf{1/6} \end{pmatrix} + \begin{pmatrix} \mathbf{1/6} \\ \mathbf{-1/6} \end{pmatrix}.$$

We keep computing

$$\mathbf{P}\vec{x}_0 = \mathbf{P}\begin{pmatrix} \mathbf{5/6} \\ \mathbf{1/6} \end{pmatrix} + \mathbf{P}\begin{pmatrix} \mathbf{1/6} \\ -\mathbf{1/6} \end{pmatrix} = \begin{pmatrix} \mathbf{5/6} \\ \mathbf{1/6} \end{pmatrix} + (\mathbf{2/5})\begin{pmatrix} \mathbf{1/6} \\ -\mathbf{1/6} \end{pmatrix}$$

We keep computing

$$\mathbf{P}\vec{x}_0 = \mathbf{P}\begin{pmatrix} 5/6 \\ 1/6 \end{pmatrix} + \mathbf{P}\begin{pmatrix} 1/6 \\ -1/6 \end{pmatrix} = \begin{pmatrix} 5/6 \\ 1/6 \end{pmatrix} + (2/5)\begin{pmatrix} 1/6 \\ -1/6 \end{pmatrix}$$

$$\mathbf{P}^2\vec{x}_0 = \mathbf{P}^2\begin{pmatrix} 5/6 \\ 1/6 \end{pmatrix} + \mathbf{P}^2\begin{pmatrix} 1/6 \\ -1/6 \end{pmatrix} = \mathbf{P}\begin{pmatrix} 5/6 \\ 1/6 \end{pmatrix} + \mathbf{P}(2/5)\begin{pmatrix} 1/6 \\ -1/6 \end{pmatrix}$$

We keep computing

$$\mathbf{P}\vec{x}_0 = \mathbf{P}\begin{pmatrix} \mathbf{5/6} \\ \mathbf{1/6} \end{pmatrix} + \mathbf{P}\begin{pmatrix} \mathbf{1/6} \\ -\mathbf{1/6} \end{pmatrix} = \begin{pmatrix} \mathbf{5/6} \\ \mathbf{1/6} \end{pmatrix} + (\mathbf{2/5})\begin{pmatrix} \mathbf{1/6} \\ -\mathbf{1/6} \end{pmatrix}$$

$$\begin{aligned}\mathbf{P}^2\vec{x}_0 &= \mathbf{P}^2\begin{pmatrix} \mathbf{5/6} \\ \mathbf{1/6} \end{pmatrix} + \mathbf{P}^2\begin{pmatrix} \mathbf{1/6} \\ -\mathbf{1/6} \end{pmatrix} = \mathbf{P}\begin{pmatrix} \mathbf{5/6} \\ \mathbf{1/6} \end{pmatrix} + \mathbf{P}(\mathbf{2/5})\begin{pmatrix} \mathbf{1/6} \\ -\mathbf{1/6} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{5/6} \\ \mathbf{1/6} \end{pmatrix} + (\mathbf{2/5})^2\begin{pmatrix} \mathbf{1/6} \\ -\mathbf{1/6} \end{pmatrix}\end{aligned}$$

We keep computing

$$\mathbf{P}\vec{x}_0 = \mathbf{P}\begin{pmatrix} 5/6 \\ 1/6 \end{pmatrix} + \mathbf{P}\begin{pmatrix} 1/6 \\ -1/6 \end{pmatrix} = \begin{pmatrix} 5/6 \\ 1/6 \end{pmatrix} + (2/5)\begin{pmatrix} 1/6 \\ -1/6 \end{pmatrix}$$

$$\begin{aligned}
\mathbf{P}^2\vec{x}_0 &= \mathbf{P}^2\begin{pmatrix} 5/6 \\ 1/6 \end{pmatrix} + \mathbf{P}^2\begin{pmatrix} 1/6 \\ -1/6 \end{pmatrix} = \mathbf{P}\begin{pmatrix} 5/6 \\ 1/6 \end{pmatrix} + \mathbf{P}(2/5)\begin{pmatrix} 1/6 \\ -1/6 \end{pmatrix} \\
&= \begin{pmatrix} 5/6 \\ 1/6 \end{pmatrix} + (2/5)^2\begin{pmatrix} 1/6 \\ -1/6 \end{pmatrix} \\
&\vdots \\
\mathbf{P}^n\vec{x}_0 &= \mathbf{P}^n\begin{pmatrix} 5/6 \\ 1/6 \end{pmatrix} + \mathbf{P}^n\begin{pmatrix} 1/6 \\ -1/6 \end{pmatrix} = \begin{pmatrix} 5/6 \\ 1/6 \end{pmatrix} + (2/5)^n\begin{pmatrix} 1/6 \\ -1/6 \end{pmatrix}
\end{aligned}$$

We keep computing

$$\mathbf{P}\vec{x}_0 = \mathbf{P}\begin{pmatrix} \mathbf{5/6} \\ \mathbf{1/6} \end{pmatrix} + \mathbf{P}\begin{pmatrix} \mathbf{1/6} \\ \mathbf{-1/6} \end{pmatrix} = \begin{pmatrix} \mathbf{5/6} \\ \mathbf{1/6} \end{pmatrix} + (\mathbf{2/5})\begin{pmatrix} \mathbf{1/6} \\ \mathbf{-1/6} \end{pmatrix}$$

$$\mathbf{P}^2\vec{x}_0 = \mathbf{P}^2\begin{pmatrix} \mathbf{5/6} \\ \mathbf{1/6} \end{pmatrix} + \mathbf{P}^2\begin{pmatrix} \mathbf{1/6} \\ \mathbf{-1/6} \end{pmatrix} = \mathbf{P}\begin{pmatrix} \mathbf{5/6} \\ \mathbf{1/6} \end{pmatrix} + \mathbf{P}(\mathbf{2/5})\begin{pmatrix} \mathbf{1/6} \\ \mathbf{-1/6} \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{5/6} \\ \mathbf{1/6} \end{pmatrix} + (\mathbf{2/5})^2\begin{pmatrix} \mathbf{1/6} \\ \mathbf{-1/6} \end{pmatrix}$$

$$\vdots$$

$$\mathbf{P}^n\vec{x}_0 = \mathbf{P}^n\begin{pmatrix} \mathbf{5/6} \\ \mathbf{1/6} \end{pmatrix} + \mathbf{P}^n\begin{pmatrix} \mathbf{1/6} \\ \mathbf{-1/6} \end{pmatrix} = \begin{pmatrix} \mathbf{5/6} \\ \mathbf{1/6} \end{pmatrix} + (\mathbf{2/5})^n\begin{pmatrix} \mathbf{1/6} \\ \mathbf{-1/6} \end{pmatrix}$$

As n gets bigger, $(\mathbf{2/5})^n$ goes to zero.

We keep computing

$$\mathbf{P}\vec{x}_0 = \mathbf{P}\begin{pmatrix} 5/6 \\ 1/6 \end{pmatrix} + \mathbf{P}\begin{pmatrix} 1/6 \\ -1/6 \end{pmatrix} = \begin{pmatrix} 5/6 \\ 1/6 \end{pmatrix} + (2/5)\begin{pmatrix} 1/6 \\ -1/6 \end{pmatrix}$$

$$\mathbf{P}^2\vec{x}_0 = \mathbf{P}^2\begin{pmatrix} 5/6 \\ 1/6 \end{pmatrix} + \mathbf{P}^2\begin{pmatrix} 1/6 \\ -1/6 \end{pmatrix} = \mathbf{P}\begin{pmatrix} 5/6 \\ 1/6 \end{pmatrix} + \mathbf{P}(2/5)\begin{pmatrix} 1/6 \\ -1/6 \end{pmatrix}$$

$$= \begin{pmatrix} 5/6 \\ 1/6 \end{pmatrix} + (2/5)^2\begin{pmatrix} 1/6 \\ -1/6 \end{pmatrix}$$

$$\vdots$$

$$\mathbf{P}^n\vec{x}_0 = \mathbf{P}^n\begin{pmatrix} 5/6 \\ 1/6 \end{pmatrix} + \mathbf{P}^n\begin{pmatrix} 1/6 \\ -1/6 \end{pmatrix} = \begin{pmatrix} 5/6 \\ 1/6 \end{pmatrix} + (2/5)^n\begin{pmatrix} 1/6 \\ -1/6 \end{pmatrix}$$

As n gets bigger, $(2/5)^n$ goes to zero.

Thus, without explicitly defining what this limit is, it seems reasonable to say that

$$\lim_{n\to\infty} \mathbf{P}^n\vec{x}_0 = \begin{pmatrix} 5/6 \\ 1/6 \end{pmatrix}.$$

But what if we started on a gray day?

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 5/6 \\ 1/6 \end{pmatrix} - 5 \begin{pmatrix} 1/6 \\ -1/6 \end{pmatrix}.$$

So,

$$\mathbf{P}^n \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 5/6 \\ 1/6 \end{pmatrix} - 5(2/5)^n \begin{pmatrix} 1/6 \\ -1/6 \end{pmatrix},$$

yielding

$$\lim_{n\to\infty} \mathbf{P}^n \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 5/6 \\ 1/6 \end{pmatrix}.$$

But what if we started on a gray day?

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 5/6 \\ 1/6 \end{pmatrix} - 5 \begin{pmatrix} 1/6 \\ -1/6 \end{pmatrix}.$$

So,

$$\mathbf{P}^n \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 5/6 \\ 1/6 \end{pmatrix} - 5(2/5)^n \begin{pmatrix} 1/6 \\ -1/6 \end{pmatrix},$$

yielding

$$\lim_{n \to \infty} \mathbf{P}^n \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 5/6 \\ 1/6 \end{pmatrix}.$$

$$\begin{pmatrix} 5/6 \\ 1/6 \end{pmatrix}$$

is the steady state vector of the stochastic matrix **P**.

Those special vectors that helped us compute are eigenvectors.

For $\boldsymbol{n} \times \boldsymbol{n}$ matrix $\mathbf{A}$ and $\lambda \in \mathbb{R}$, define

$$E_\lambda(\mathbf{A}) = \{\vec{x} \in \mathbb{R}^n \mid \mathbf{A}\vec{x} = \lambda\vec{x}\}\,.$$

If $E_\lambda(\mathbf{A})$ contains more than the zero vector, then

- λ is an eigenvalue of $\mathbf{A}$,
- all $\vec{y} \in E_\lambda(\mathbf{A})$ is an eigenvector of $\mathbf{A}$ for eigenvalue λ, and
- $E_\lambda(\mathbf{A})$ is the eigenspace of $\mathbf{A}$ for eigenvalue λ.

The number of such λ can't exceed n.

For

$$\mathbf{P} = \begin{pmatrix} \mathbf{9/10} & \mathbf{1/2} \\ \mathbf{1/10} & \mathbf{1/2} \end{pmatrix},$$

$$E_1(\mathbf{P}) = \left\{ a \begin{pmatrix} \mathbf{5/6} \\ \mathbf{1/6} \end{pmatrix} \,\middle|\, a \in \mathbb{R} \right\} \text{ and}$$

$$E_{2/5}(\mathbf{P}) = \left\{ a \begin{pmatrix} \mathbf{1/6} \\ \mathbf{-1/6} \end{pmatrix} \,\middle|\, a \in \mathbb{R} \right\}.$$

For

$$\mathbf{P} = \begin{pmatrix} \mathbf{9/10} & \mathbf{1/2} \\ \mathbf{1/10} & \mathbf{1/2} \end{pmatrix},$$

$$E_1(\mathbf{P}) = \left\{ a \begin{pmatrix} \mathbf{5/6} \\ \mathbf{1/6} \end{pmatrix} \;\middle|\; a \in \mathbb{R} \right\} \text{ and}$$

$$E_{2/5}(\mathbf{P}) = \left\{ a \begin{pmatrix} \mathbf{1/6} \\ -\mathbf{1/6} \end{pmatrix} \;\middle|\; a \in \mathbb{R} \right\}.$$

A stochastic matrix always has an probability vector which is an eigenvector with eigenvalue $\mathbf{1}$ and all other eigenvalues are less than $\mathbf{1}$ in absolute value.

Go to Geogebra:

`https://www.geogebra.org/m/atavkxaz`

`https://www.geogebra.org/m/sqG26hQj`

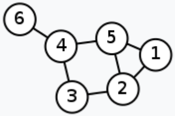
For the $\mathbf{100,001} \times \mathbf{100,001}$ matrix we multiplied by to "throw away" low frequencies,

there were two eigenvalues:

$\mathbf{1}$, with eigenvectors high frequency vectors in $\mathbb{R}^{\mathbf{100,001}}$, and

$\mathbf{0}$, with eigenvectors low frequency vectors in $\mathbb{R}^{\mathbf{100,001}}$.

More graph matrices

Labeled graph	Degree matrix	Adjacency matrix	Laplacian matrix
	$\begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2 & -1 & 0 & 0 & -1 & 0 \\ -1 & 3 & -1 & 0 & -1 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 3 & -1 & -1 \\ -1 & -1 & 0 & -1 & 3 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{pmatrix}$

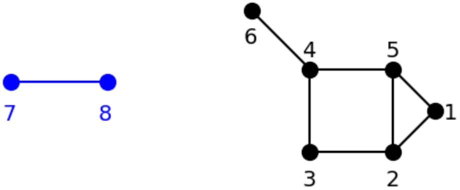
Left to right: a labeled graph, its degree matrix $\mathbf{D}$, its adjacency matrix $\mathbf{A}$, and its graph Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{A}$. Source: `https://en.wikipedia.org/wiki/Laplacian_matrix`

$$\mathbf{L}\begin{pmatrix}\mathbf{1}\\\mathbf{1}\\\vdots\\\mathbf{1}\end{pmatrix}=\mathbf{D}\begin{pmatrix}\mathbf{1}\\\mathbf{1}\\\vdots\\\mathbf{1}\end{pmatrix}-\mathbf{A}\begin{pmatrix}\mathbf{1}\\\mathbf{1}\\\vdots\\\mathbf{1}\end{pmatrix}$$

$$\mathbf{L}\begin{pmatrix}\mathbf{1}\\\mathbf{1}\\\vdots\\\mathbf{1}\end{pmatrix}=\mathbf{D}\begin{pmatrix}\mathbf{1}\\\mathbf{1}\\\vdots\\\mathbf{1}\end{pmatrix}-\mathbf{A}\begin{pmatrix}\mathbf{1}\\\mathbf{1}\\\vdots\\\mathbf{1}\end{pmatrix}$$

$$=\begin{pmatrix}\mathbf{2}\\\mathbf{3}\\\mathbf{2}\\\mathbf{3}\\\mathbf{3}\\\mathbf{1}\end{pmatrix}-\begin{pmatrix}\mathbf{2}\\\mathbf{3}\\\mathbf{2}\\\mathbf{3}\\\mathbf{3}\\\mathbf{1}\end{pmatrix}$$

$$=\begin{pmatrix}\mathbf{0}\\\mathbf{0}\\\vdots\\\mathbf{0}\end{pmatrix}=\mathbf{0}\begin{pmatrix}\mathbf{1}\\\mathbf{1}\\\vdots\\\mathbf{1}\end{pmatrix}$$



A graph with two connected components

$$\mathbf{L} = \begin{pmatrix} 2 & -1 & 0 & 0 & -1 & 0 & 0 & 0 \\ -1 & 3 & -1 & 0 & -1 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 3 & -1 & -1 & 0 & 0 \\ -1 & -1 & 0 & -1 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}$$

$$E_0\left(\begin{pmatrix} 2 & -1 & 0 & 0 & -1 & 0 & 0 & 0 \\ -1 & 3 & -1 & 0 & -1 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 3 & -1 & -1 & 0 & 0 \\ -1 & -1 & 0 & -1 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}\right) = \left\{ a\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + b\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} \;\middle|\; a, b \in \mathbb{R} \right\}$$

So, the eigenvectors for **0** are somehow related to connectivity.

Given a graph Laplacian $\mathbf{L}$,

$\mathbf{L}$ has only non-negative real eigenvalues

the elements of $E_{\mathbf{0}}(\mathbf{L})$ are constant on the connected components of the graph.

The same holds for a graph Laplacian of a weighted graph (but not a directed graph).

Let $\mathbf{L}$ be a graph Laplacian of an undirected weighted or unweighted graph.

The Fiedler eigenvalue is the smallest positive eigenvalue.

A Fiedler eigenvector is a corresponding eigenvector.

Where the eigenvector is positive / negative gives you information on clustering.

This is called algebraic connectivity.

(Go to Matlab.)

More eigenvalues and graphs

Perron-Frobenius theory guarantees that the largest eigenvalue of an (undirected) adjacency matrix has an eigenvector with all non-negative entries.

The relative sizes of the entries tells you the eigenvector centrality of the corresponding vertex.

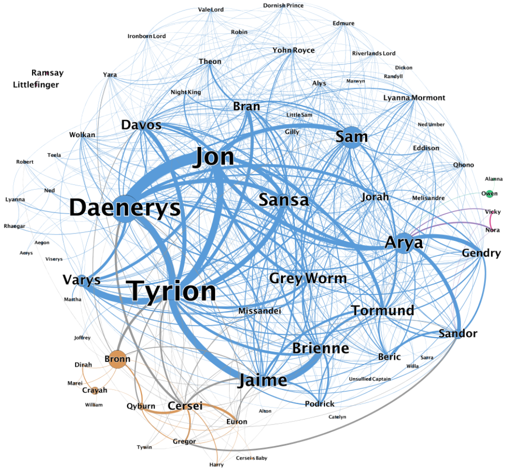


Image source: `networkofthrones.wordpress.com`

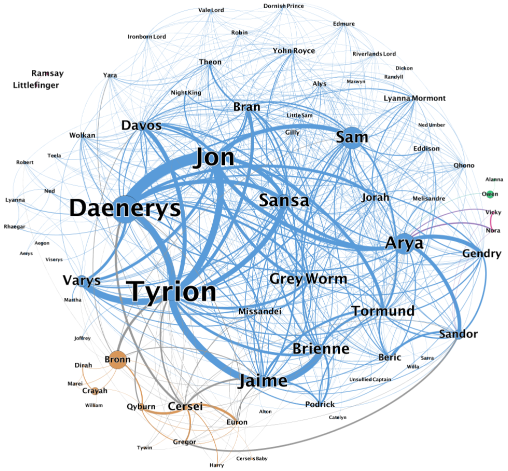


Image source: `networkofthrones.wordpress.com`

Tyrion has the highest eigenvector centrality.

(Go to Matlab)

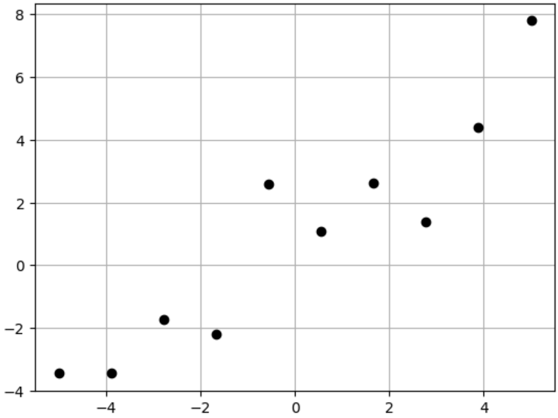
Lecture Outline

Eigenvalues and eigenvectors

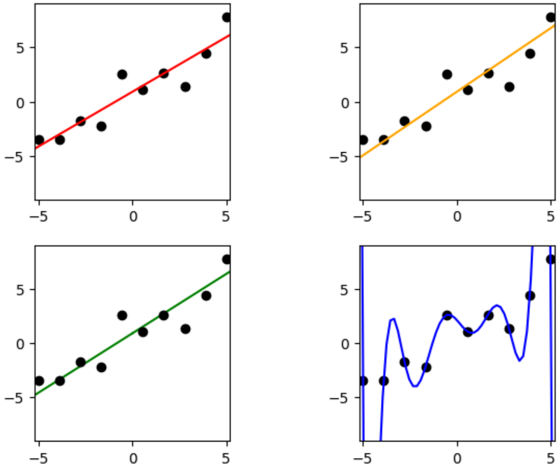
Principal component analysis

Responsibility vignette

BEST MODEL FOR THIS DATA?



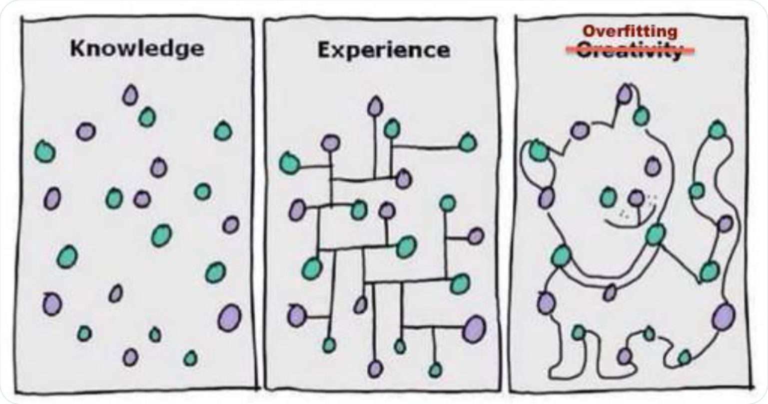
Best Model for This Data?



Overfitting

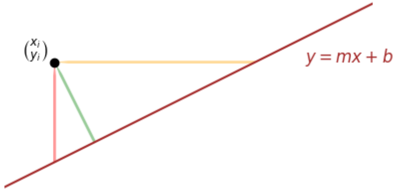
@deliprao@mastodon.social
@deliprao

overfitting



2:02 AM · Jul 23, 2020 · Twitter Web App

For a data point $\begin{pmatrix} x_i \\ y_i \end{pmatrix}$ and an equation for a line $y = mx + b$,



Vertical distance $d_{v,i}$ from point to line is $|y_i - (mx_i + b)|$

Horizontal distance $d_{h,i}$ from point to line is $|x_i - ((y_i - b)/m)|$

Perpendicular distance $d_{p,i}$ from point to line is $\frac{|mx_i - y_i + b|}{\sqrt{m^2+1}}$

Least squares line fitting

For a data set $\{x_i\}_i^N$ the solution to

$$\min_{m,b} \sum_{i=1}^{N} d_i^2$$

is

Least squares line fitting

For a data set $\{x_i\}_i^N$ the solution to

$$\min_{m,b} \sum_{i=1}^{N} d_i^2$$

is

- linear regression of y onto x when $d = d_v$

Least squares line fitting

For a data set $\{x_i\}_i^N$ the solution to

$$\min_{m,b} \sum_{i=1}^{N} d_i^2$$

is

- linear regression of y onto x when $d = d_v$
- linear regression of x onto y when $d = d_h$

Least squares line fitting

For a data set $\{x_i\}_i^N$ the solution to

$$\min_{m,b} \sum_{i=1}^{N} d_i^2$$

is

- linear regression of y onto x when $d = d_v$
- linear regression of x onto y when $d = d_h$
- (one-dimensional) principal component analysis of x and y when $d = d_p$

Linear Regression via Least Squares, I

Given data $\{(t_i, f_i)\}_{i=1}^n$ and variables a, b, define

$$\vec{y} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} t_1 & 1 \\ t_2 & 1 \\ \vdots & \vdots \\ t_n & 1 \end{pmatrix}, \quad \vec{x} = \begin{pmatrix} a \\ b \end{pmatrix}.$$

Linear Regression via Least Squares, I

Given data $\{(t_i, f_i)\}_{i=1}^n$ and variables a, b, define

$$\vec{y} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} t_1 & 1 \\ t_2 & 1 \\ \vdots & \vdots \\ t_n & 1 \end{pmatrix}, \quad \vec{x} = \begin{pmatrix} a \\ b \end{pmatrix}.$$

Then

$$\mathbf{A}\vec{x} = \begin{pmatrix} t_1 & 1 \\ t_2 & 1 \\ \vdots & \vdots \\ t_n & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} at_1 + b \\ at_2 + b \\ \vdots \\ at_n + b \end{pmatrix}$$

Linear Regression via Least Squares, I

Given data $\{(t_i, f_i)\}_{i=1}^n$ and variables a, b, define

$$\vec{y} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} t_1 & 1 \\ t_2 & 1 \\ \vdots & \vdots \\ t_n & 1 \end{pmatrix}, \quad \vec{x} = \begin{pmatrix} a \\ b \end{pmatrix}.$$

Then

$$\mathbf{A}\vec{x} = \begin{pmatrix} t_1 & 1 \\ t_2 & 1 \\ \vdots & \vdots \\ t_n & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} at_1 + b \\ at_2 + b \\ \vdots \\ at_n + b \end{pmatrix}$$

and

$$\vec{y} - \mathbf{A}\vec{x} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{pmatrix} - \begin{pmatrix} at_1 + b \\ at_2 + b \\ \vdots \\ at_n + b \end{pmatrix} = \begin{pmatrix} f_1 - (at_1 + b) \\ f_2 - (at_2 + b) \\ \vdots \\ f_n - (at_n + b) \end{pmatrix}$$

LINEAR REGRESSION VIA LEAST SQUARES, II

Given data $\{(t_i, f_i)\}_{i=1}^n$, define

$$\vec{y} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{pmatrix}, \quad A = \begin{pmatrix} t_1 & 1 \\ t_2 & 1 \\ \vdots & \vdots \\ t_n & 1 \end{pmatrix}, \quad x = \begin{pmatrix} a \\ b \end{pmatrix}.$$

Then the slope a and intercept b of the linear regression line is the entries of (the) solution to

$$\hat{\vec{x}} = \mathbf{argmin}_{\vec{x}} \| \vec{y} - \mathbf{A}\vec{x} \|^2 .$$

If $\mathbf{A}^\top \mathbf{A}$ is invertible, the *normal equations* yield

$$\hat{\vec{x}} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \vec{y}.$$

Many other problems may be formulated as a least squares problem,

where $\mathbf{A}$ is a $d \times n$ matrix encoding your model,

$\vec{y}$ are your d measurements,

and $\vec{x}$ are the n model parameters that you want to find, using

$$\hat{\vec{x}} = \mathbf{argmin}_{\vec{x}} \| \vec{y} - \mathbf{A}\vec{x} \|^2 .$$

The solutions $\hat{\vec{x}}$ are precisely solutions to the normal equations

$$\mathbf{A}^\top \mathbf{A} \hat{\vec{x}} = \mathbf{A}^\top \vec{y}.$$

When should one use linear regression?

When should one use linear regression?

When you assume that there is an approximately linear relationship between a scalar response/dependent variable and one or more explanatory variables/independent variables.

Singular value decomposition (SVD)

Every $d \times n$ matrix $\mathbf{A}$ has

$\min\{d, n\}$ non-negative singular values σ_i

d left singular vectors $\{\vec{u}_j\}_{j=1}^{d}$

n right singular vectors $\{\vec{v}_k\}_{k=1}^{n}$

Singular value decomposition (SVD)

Every $d \times n$ matrix $\mathbf{A}$ has

$\min\{d, n\}$ non-negative singular values σ_i

d left singular vectors $\{\vec{u}_j\}_{j=1}^{d}$

n right singular vectors $\{\vec{v}_k\}_{k=1}^{n}$

They are like eigenvalues.
In fact, they arise from finding eigenvalues of an associated matrix.

Singular value decomposition (SVD)

Every $d \times n$ matrix $\mathbf{A}$ has

$\min\{d, n\}$ non-negative singular values σ_i

d left singular vectors $\{\vec{u}_j\}_{j=1}^{d}$

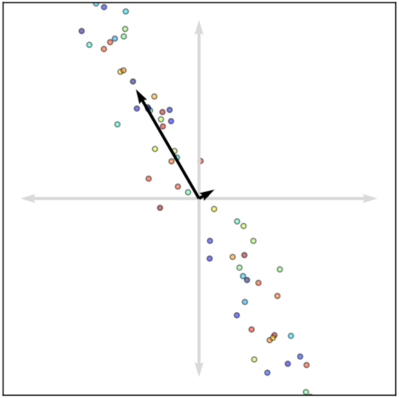
n right singular vectors $\{\vec{v}_k\}_{k=1}^{n}$

They are like eigenvalues.
In fact, they arise from finding eigenvalues of an associated matrix.

The Schmidt-Eckart-Young-Mirsky Theorem guarantees the awesomeness of the singular vectors in approximating data.

Principal Component Analysis

- Center the data and enter as columns[a] of a matrix.
- Perform SVD.
- Large singular values indicate importance. The corresponding singular left singular vectors correspond to directions of data trends.



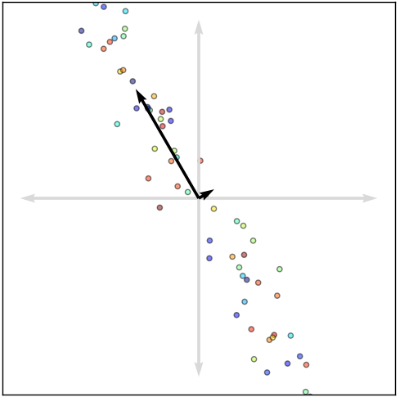
[a]some weirdos use rows

Principal Component Analysis

Should be used in many cases that linear regression is used!

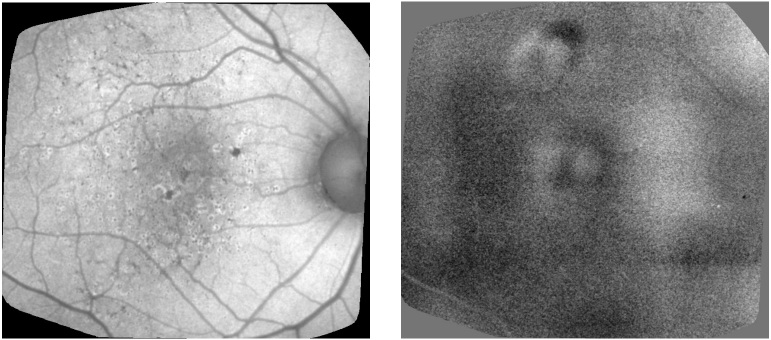
- Center the data and enter as columns[a] of a matrix.
- Perform SVD.
- Large singular values indicate importance. The corresponding singular left singular vectors correspond to directions of data trends.

[a]some weirdos use rows



(Switch to Matlab)

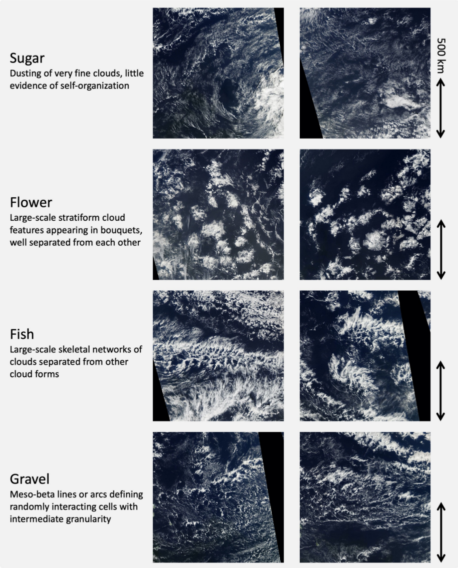
Retinal images



The first and second principal components of a stack of four fundus camera images of the retina of a patient with age-related macular degeneration.

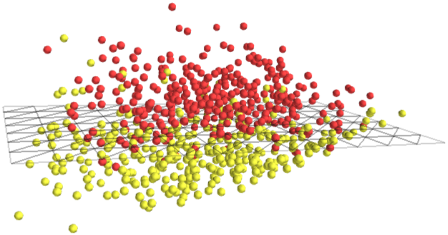
Ehler, Majumdar, K, et al. 2010; original data source: National Eye Institute

Mesoscale cloud organization



Examples of the four cloud types from the Sugar, Fish, Flowers, and Gravel dataset, Rasp et al. 2019

PCA AND CLOUDS



After mapping to $\mathbb{R}^{2000}$ to capture geometric traits, PCA performed down to $\mathbb{R}^3$ – capturing about **90%** of the variation – for separation.

ver Hoef, Adams, King, Ebert-Uphoff 2023

Lecture Outline

Eigenvalues and eigenvectors

Principal component analysis

Responsibility vignette

The importance of taking responsibility when assigning mathematical certainty to any model, not just one trained on potentially biased data should not be dismissed.

Go forth,

and responsibly use what you’ve learned here!