Machine Learning

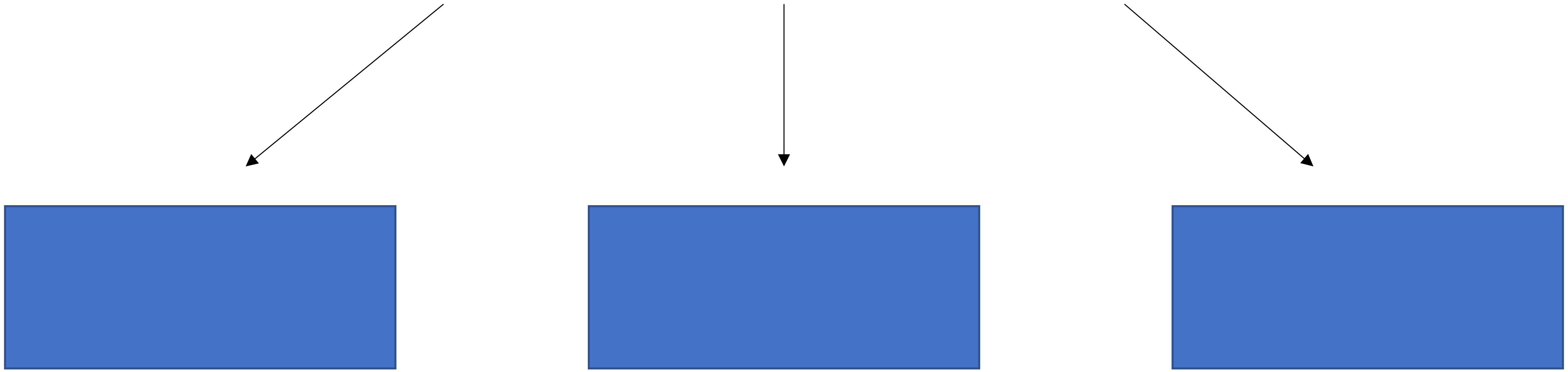
Topic 6. Lecture 6

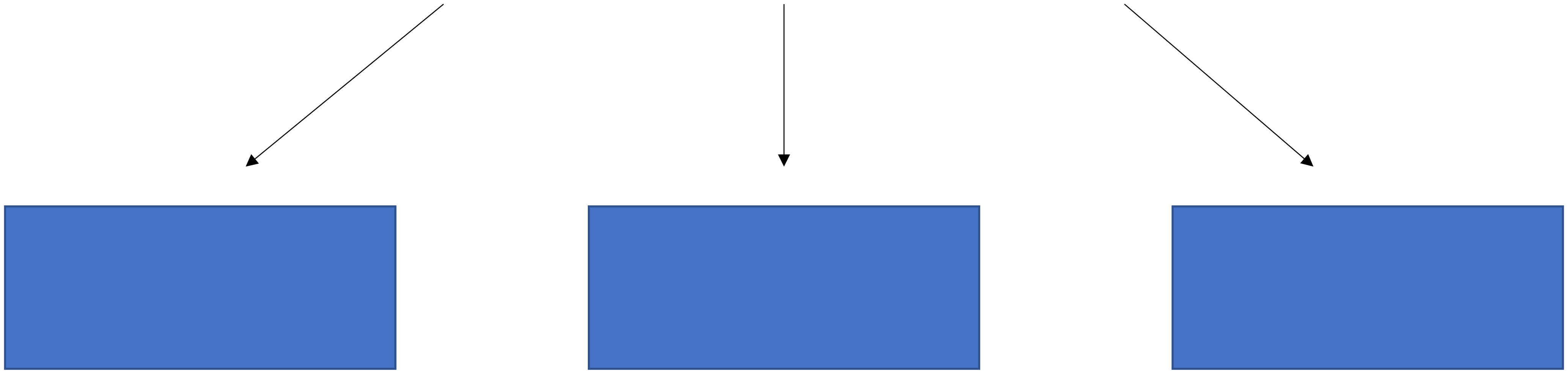
Math for machine learning. Linear algebra

Yury Sanochkin

ysanochkin@hse.ru

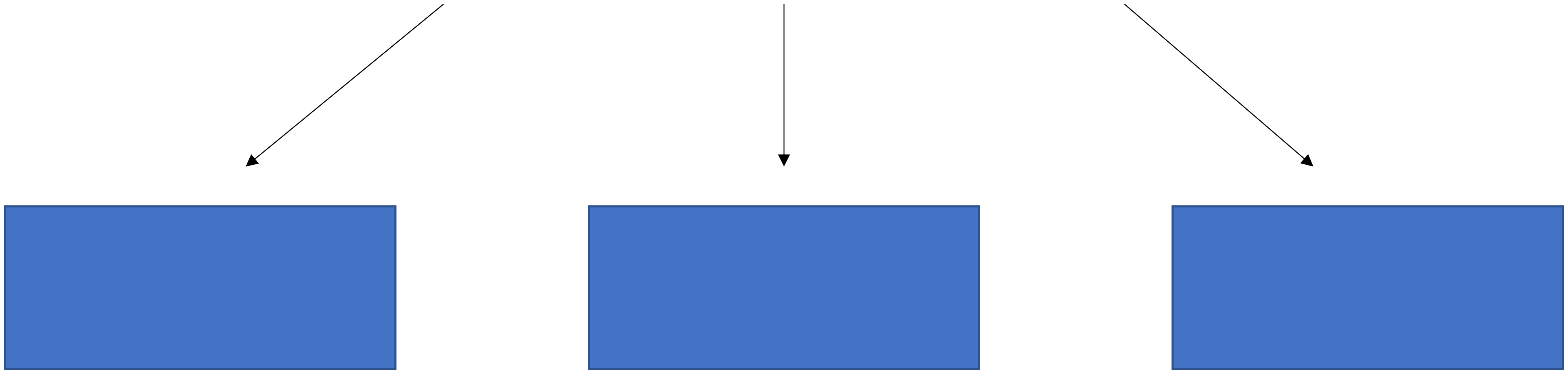
NRU HSE, 2025





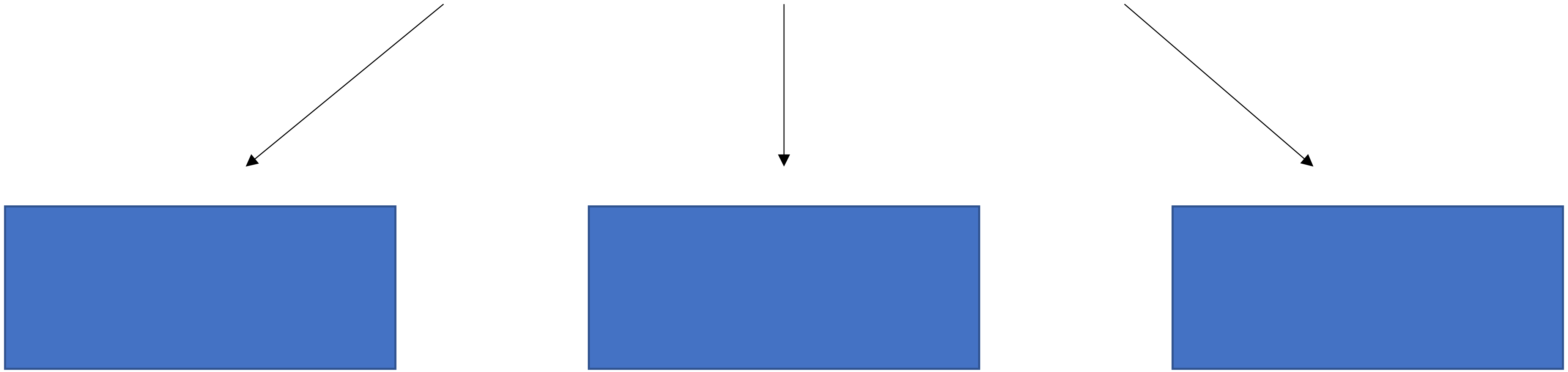
Does this look familiar? :)
Do you think you know what's going to show up on the next slide here? :)

Math for machine learning



I think you're wrong!

Math for machine learning



What parts of mathematics are important in data analysis and for what purpose?

Math for machine learning

- Probability theory and mathematical statistics
- Linear algebra
- Mathematical analysis

Math for machine learning

Probability theory and mathematical statistics

Linear algebra

Mathematical analysis

These are the three whales of the sections of mathematics on which all data analysis is based

Math for machine learning

Probability theory and mathematical statistics

Linear algebra

Mathematical analysis

This we have already recalled and discussed in detail in the previous module!

Math for machine learning

Probability theory and mathematical statistics

Linear algebra

Mathematical analysis

Today our conversation is about this!

Math for machine learning

Probability theory and mathematical statistics

Linear algebra

Mathematical analysis

This is waiting for us ahead!

Linear algebra. Motivation

Linear algebra. Motivation

- Why linear algebra?
- What is this section about and why is it so important for data analysis and specifically for machine learning?

Linear algebra. Motivation

- Over the last 30 years, we've taught machines to
  - Look and see;
  - Listen, hear and answer back;
  - Write poems and songs;
  - Predict stock market quotes;
  - Find cancerous tumors.
- and more....

Linear algebra. Motivation

- Over the last 30 years, we've taught machines to
  - Look and see;
  - Listen, hear and answer back;
  - Write poems and songs;
  - Predict stock market quotes;
  - Find cancerous tumors.
- and more....

- The range of possibilities for machine learning blows human imagination

Linear algebra. Motivation

- The range of possibilities for machine learning blows human imagination

- It seems all the more amazing that an ordinary computer is capable of all this
- In fact - just an advanced calculator

Linear algebra. Motivation

- A computer is a very stupid box!
- It only understands zeros and ones

Linear algebra. Motivation

- A computer is a very stupid box!
- It only understands zeros and ones
- ...but it does it very well!
- How well?

Linear algebra. Motivation

- A computer is a very stupid box!
- It only understands zeros and ones
- ...but it does it very well!
- How well?
- So good that if you make up numbers out of zeros and ones, it can process more numbers in a second than you can in a lifetime!

Linear algebra. Motivation

- What is the conclusion to be drawn from that?
- Photos, videos, the sounds of your voice, Mayakovsky's poems - all need to be expressed in the language of numbers.

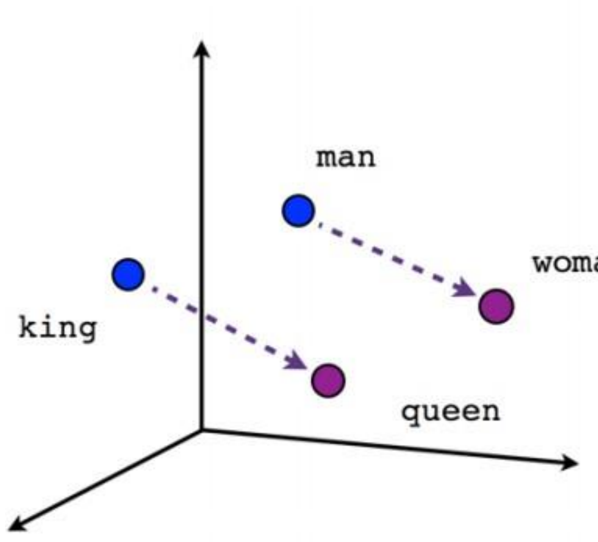
Linear algebra. Motivation

- What is the conclusion to be drawn from that?
- Photos, videos, the sounds of your voice, Mayakovsky's poems - all need to be expressed in the language of numbers.
- And what makes this possible is a section of math called linear algebra.

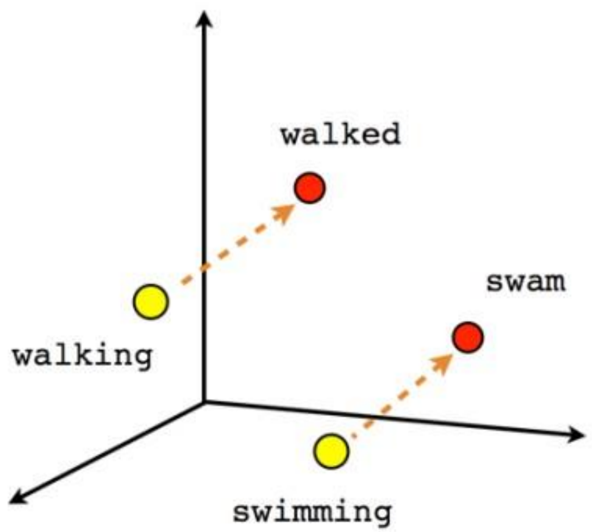
Linear algebra. Motivation

- Linear algebra has opened up unimaginable things to us:

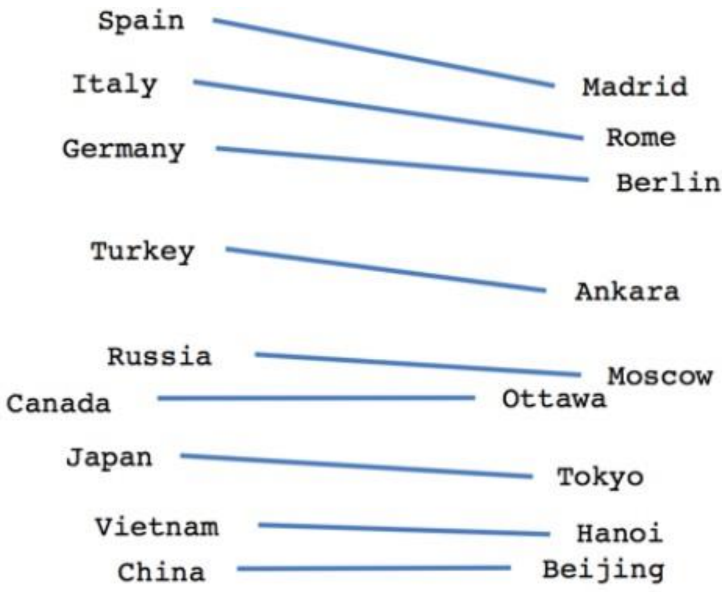
Word arithmetic



Male-Female

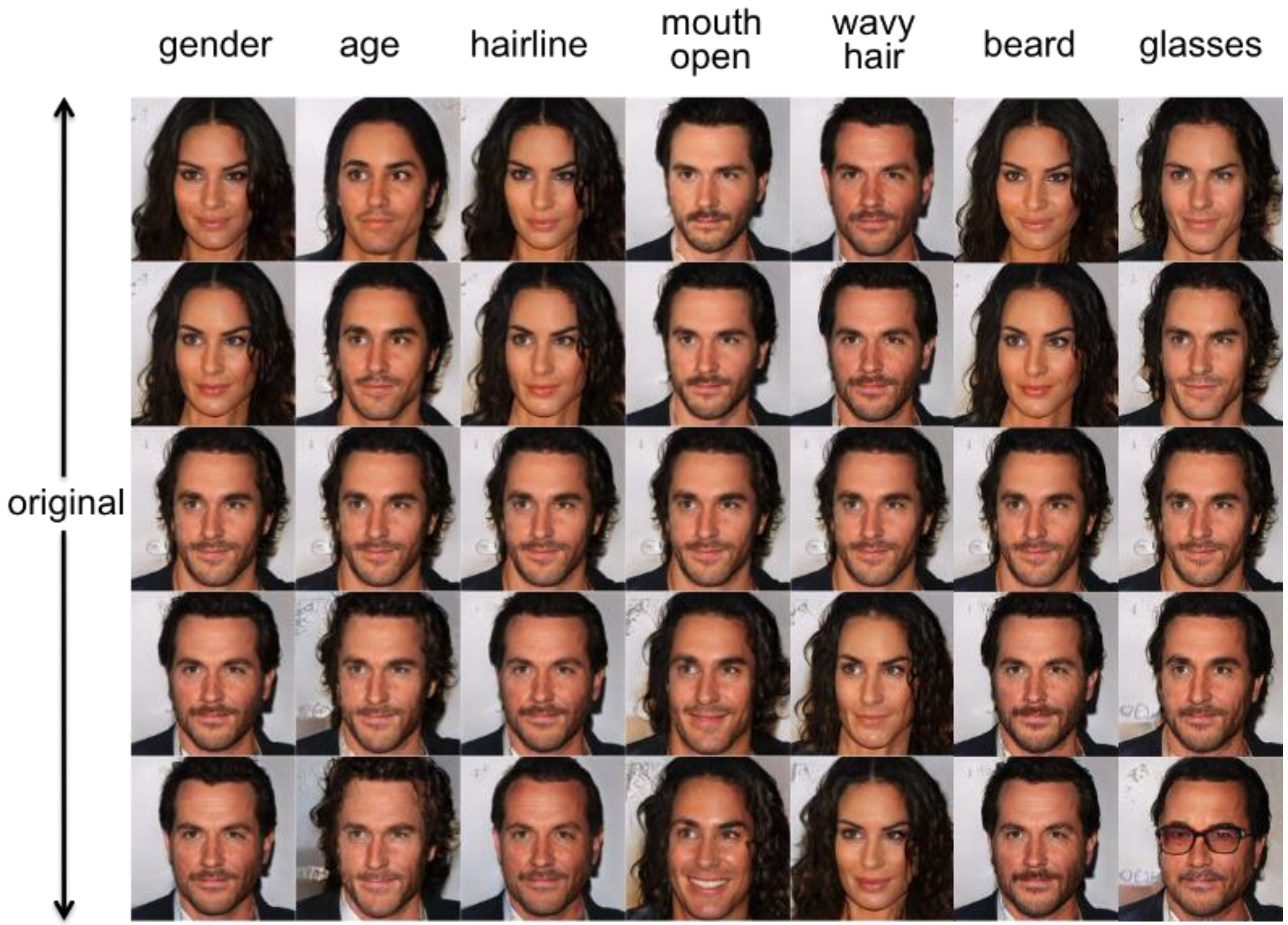


Verb tense



Country-Capital

Generation of faces with specified properties



Linear algebra. Motivation

- …and more

Linear algebra. Motivation

- …and more
- Linear algebra is the language in which the problem is understood by a machine.
- Our job is to become translators from one language to another.

Basic Concepts

Linear space

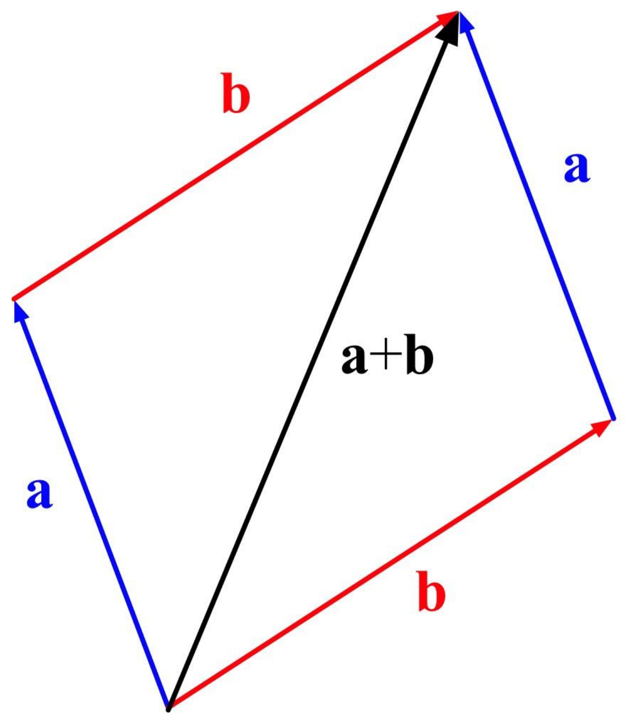
- The fundamental concept of linear algebra is the concept of linear space.
- Let's recall what is it!

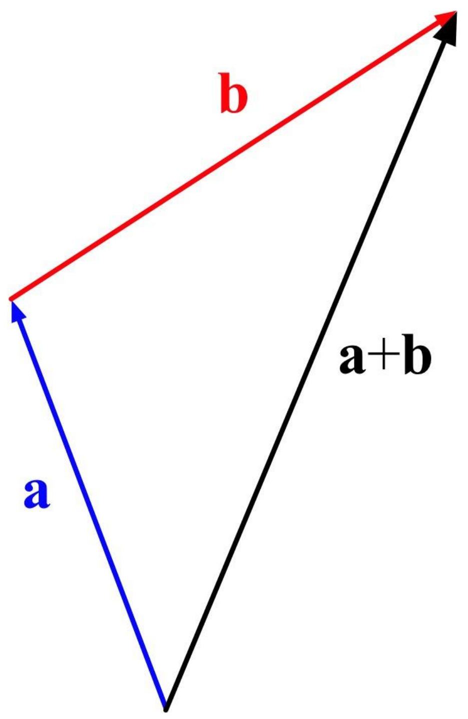
Linear space

- The fundamental concept of linear algebra is the concept of linear space.
- Let's recall what is it!
- A mathematical structure is a set of elements called vectors, for which the operations of addition with each other and multiplication by a number are defined - a scalar

Linear space

- Example: a set of real vectors of n components (e.g. (0, 0, 0, 0, ..., 0))





Linear space

- In practice you will have to deal mostly with real vectors, matrices and tensors.
- Nevertheless, linear spaces do not stop there.

Linear space

- In practice you will have to deal mostly with real vectors, matrices and tensors.
- Nevertheless, linear spaces do not stop there.

In general, a linear space consists of an abelian group of vectors (with the operation of addition), a field of scalars, and the operation of scalar multiplication of a vector, which is consistent with vector addition through the axioms of distributivity.

These conditions are easy to check, thereby determining whether the space is linear (vectorial).

Linear space

- Give examples of other linear spaces!

Linear space

- Give examples of other linear spaces!
  - The space of matrices of the same dimension.
  - The space of tensors ("multidimensional" matrices) of the same dimension.
  - The space of polynomials (ordinary or trigonometric).
  - The space of solutions to a homogeneous system of linear equations, whether algebraic or differential.
  - The space of binary vectors with addition modulo 2 and multiplication operations by 0 and 1.

Linear space

- Great, we recalled linear spaces!
- Nevertheless, in this session we will focus on real vectors, matrices, and the relationships between them, as these are primarily what a Data Scientist works with.

$$
\begin{array}{cl}
\begin{array}{ccccccc} \text{Feature-1} & \text{Feature-2} & \text{Feature-3} & \text{Feature-4} & & & \text{Feature-n} \end{array} & \\
\left[\begin{array}{ccccccc}
x_1^1 & x_2^1 & x_3^1 & x_4^1 & \dots & \dots & x_n^1 \\
x_1^2 & x_2^2 & x_3^2 & x_4^2 & \dots & \dots & x_n^2 \\
x_1^3 & x_2^3 & x_3^3 & x_4^3 & \dots & \dots & x_n^3 \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots \\
x_1^m & x_2^m & x_3^m & x_4^m & \dots & \dots & x_n^m
\end{array}\right] &
\begin{array}{l} \text{Sample-1} \\ \text{Sample-2} \\ \text{Sample-3} \\ \\ \text{Sample-m} \end{array}
\end{array}
$$

Matrices

- Could you please remind me what a matrix even is?

Matrices

- Could you please remind me what a matrix even is?
- Matrix - representation of numbers as a two-dimensional table

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \quad \begin{pmatrix} 1 & 2 & -1 \\ 3 & 4 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}$$

Matrices

- What is a transposed matrix?

Matrices

- What is a transposed matrix?
- A transposed matrix is a matrix mirrored across its main diagonal.

$$\begin{pmatrix} 1 & 2 & -1 \\ 3 & 4 & 0 \end{pmatrix}^T = \begin{pmatrix} 1 & 3 \\ 2 & 4 \\ -1 & 0 \end{pmatrix} \qquad \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}^T = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}^T = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$$

Matrices

- What is a identity matrix?

Matrices

- What is a identity matrix?
- An identity matrix is a square matrix in which all the elements on the main diagonal are ones, and all other elements are zeros. It is typically denoted by I or E.

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Matrices

- What is a diagonal matrix?

Matrices

- What is a diagonal matrix?
- A diagonal matrix is a matrix in which any numbers can be placed on the main diagonal, while all other elements are zeros.

$$\begin{pmatrix} 5 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 10 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \begin{pmatrix} -7 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 70 \end{pmatrix}$$

Matrices

- The principle of matrix-vector multiplication

Matrices

- The principle of matrix-vector multiplication

$$\begin{pmatrix} 1 & 5 \\ 3 & 8 \end{pmatrix} \cdot \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 5 \end{pmatrix} \qquad \begin{pmatrix} 2 & 3 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 5$$

Matrices

- The principle of matrix-vector multiplication

$$\begin{pmatrix} 1 & 5 \\ 3 & 8 \end{pmatrix} \cdot \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 5 \end{pmatrix} \qquad \begin{pmatrix} 2 & 3 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 5$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} =$$

Matrices

- The principle of matrix multiplication

Matrices

- The principle of matrix multiplication

$$\begin{pmatrix} -1 & 3 \\ 4 & 0 \end{pmatrix} \cdot \begin{pmatrix} -1 & 4 \\ 3 & 0 \end{pmatrix} = \begin{pmatrix} 10 & -4 \\ -4 & 16 \end{pmatrix}$$

Matrices

- The principle of matrix multiplication

$$\begin{pmatrix} -1 & 3 \\ 4 & 0 \end{pmatrix} \cdot \begin{pmatrix} -1 & 4 \\ 3 & 0 \end{pmatrix} = \begin{pmatrix} 10 & -4 \\ -4 & 16 \end{pmatrix}$$

$$\begin{pmatrix} -1 & 3 & 1 \\ 4 & 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} =$$

Matrices

- What is an inverse matrix?

Matrices

- What is an inverse matrix?
- An inverse matrix is a matrix that, when multiplied by the original matrix, results in the identity matrix. It is denoted with an superscript of -1.

$$A \cdot A^{-1} = I$$

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \qquad A^{-1} = \begin{pmatrix} -2 & 1 \\ 3/2 & -1/2 \end{pmatrix}$$

Matrices

- What is an inverse matrix?
- An inverse matrix is a matrix that, when multiplied by the original matrix, results in the identity matrix. It is denoted with an superscript of -1.

$$A \cdot A^{-1} = I$$

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \qquad A^{-1} = \begin{pmatrix} -2 & 1 \\ 3/2 & -1/2 \end{pmatrix}$$

In general, finding the inverse matrix computationally is very difficult, and in practice, it is often done approximately.

Basis of a linear space

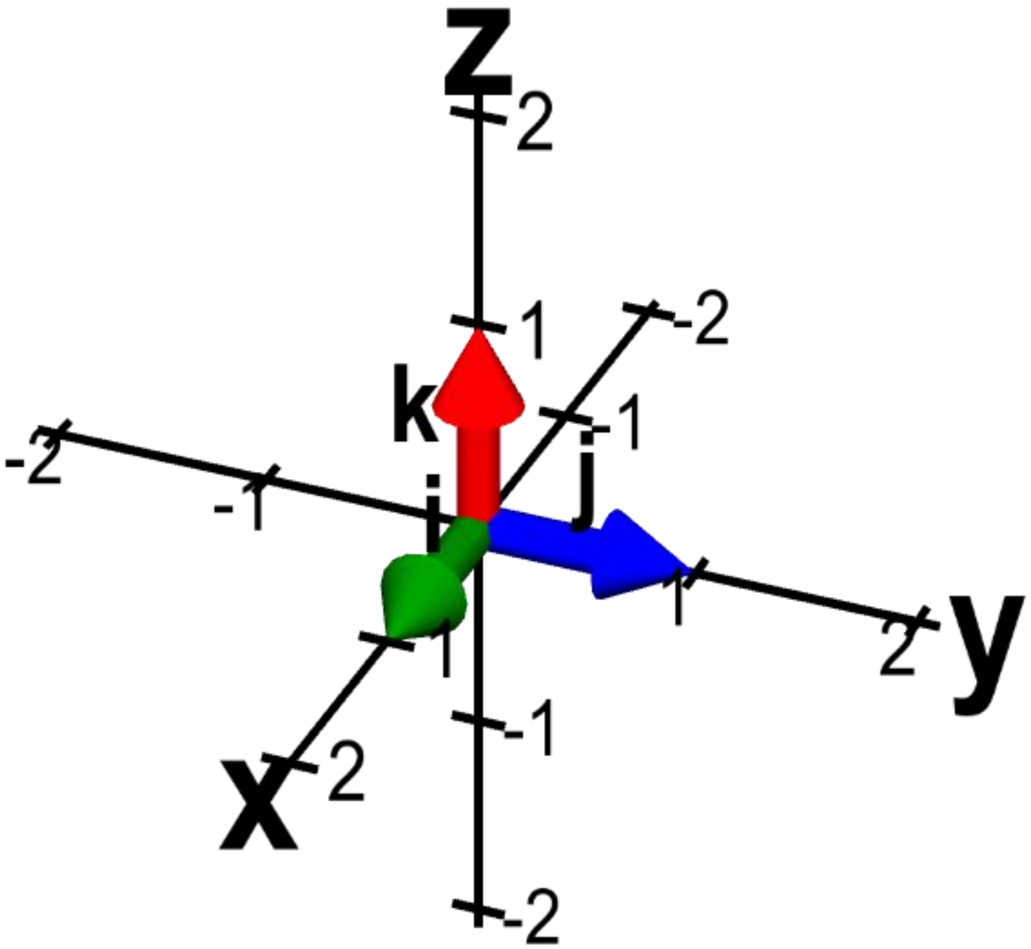
- What is a basis of a linear space?

Basis of a linear space

- What is a basis of a linear space?
- In every linear space, there are several principal vectors called the basis. All other vectors can be expressed through these basis vectors.
- The number of basis vectors in the basis is determined by the dimension of the space.

Basis of a linear space

- Example:
- In the case of ℝ^n, there are exactly n basis vectors.
- They define the coordinate axes.



Basis of a linear space

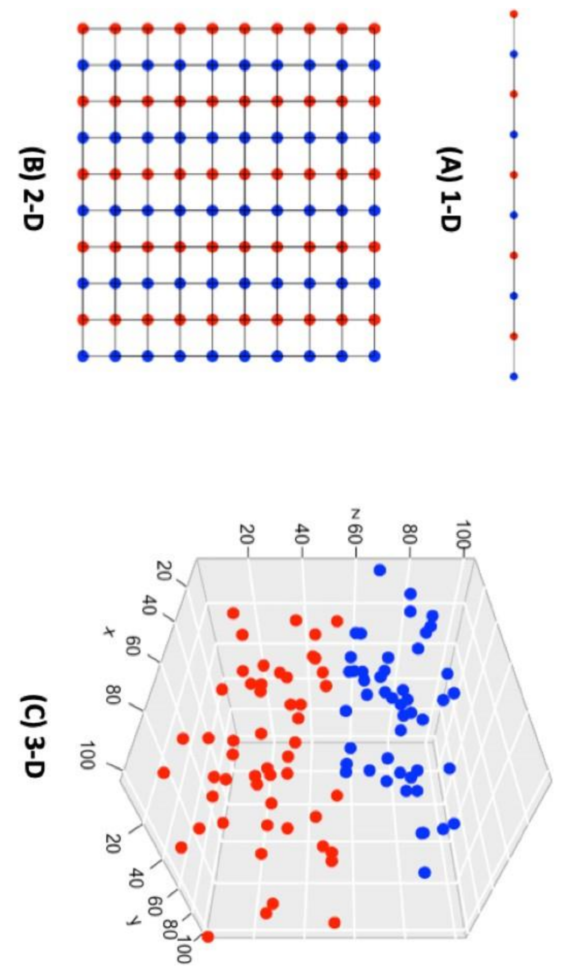
- A basis is a minimal set of vectors such that:
  - None of them can be expressed as a linear combination of the others;
  - Every element of the space can be uniquely represented as a linear combination of a finite set of vectors from this set.

Basis of a linear space

- A basis is a minimal set of vectors such that:
  - None of them can be expressed as a linear combination of the others;
  - Every element of the space can be uniquely represented as a linear combination of a finite set of vectors from this set.
- It is possible to prove that every linear space has a basis using Zorn's Lemma.

Dimension of a linear space

- Dimension is one of the main characteristics of a linear space.
- The dimension precisely corresponds to the number of basis vectors.
- In machine learning, there's even a phenomenon named after it—the "curse of dimensionality."



Matrices

- What is a determinant?

Matrices

- What is a determinant?
- Now this gets more interesting—yes, it is a certain measure for matrices. But what does it show?

Matrices

- What is a determinant?
- Now this gets more interesting—yes, it is a certain measure for matrices. But what does it show?
- The determinant is a numerical characteristic of a matrix, which in a sense describes the compression or expansion of space when transformed by this matrix.

$$det\begin{pmatrix}-1 & -1\\ 1 & 3\end{pmatrix} = (-1)\cdot 3 - (-1)\cdot 1 = -2$$

Matrices

- If we connect the resulting value with its geometric meaning, considering that we are transforming a unit square using the matrix:
  - A negative sign indicates that our square will have an opposite orientation.
  - A value of 2 indicates that after the transformation, our square will have an area twice that of the original square.

$$det \begin{pmatrix} -1 & -1 \\ 1 & 3 \end{pmatrix} = (-1) \cdot 3 - (-1) \cdot 1 = -2$$

Linear transformations

- Now for a serious question: how are matrices related to linear transformations?

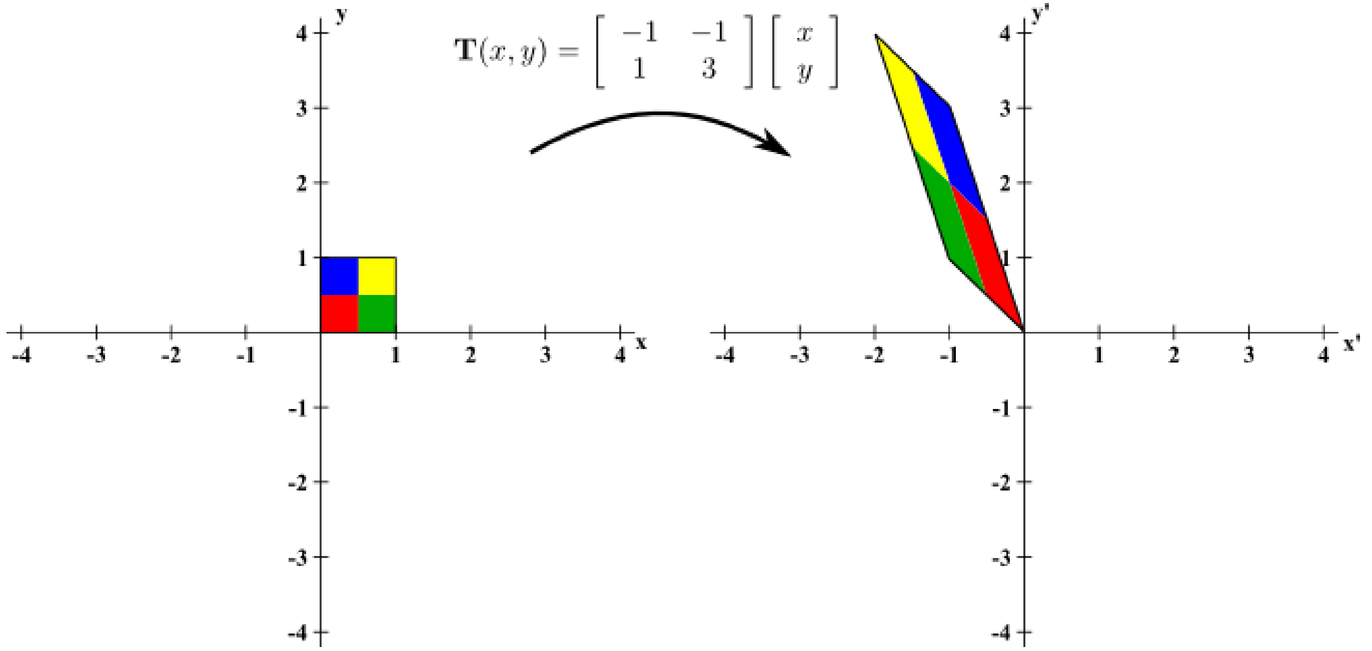
Linear transformations

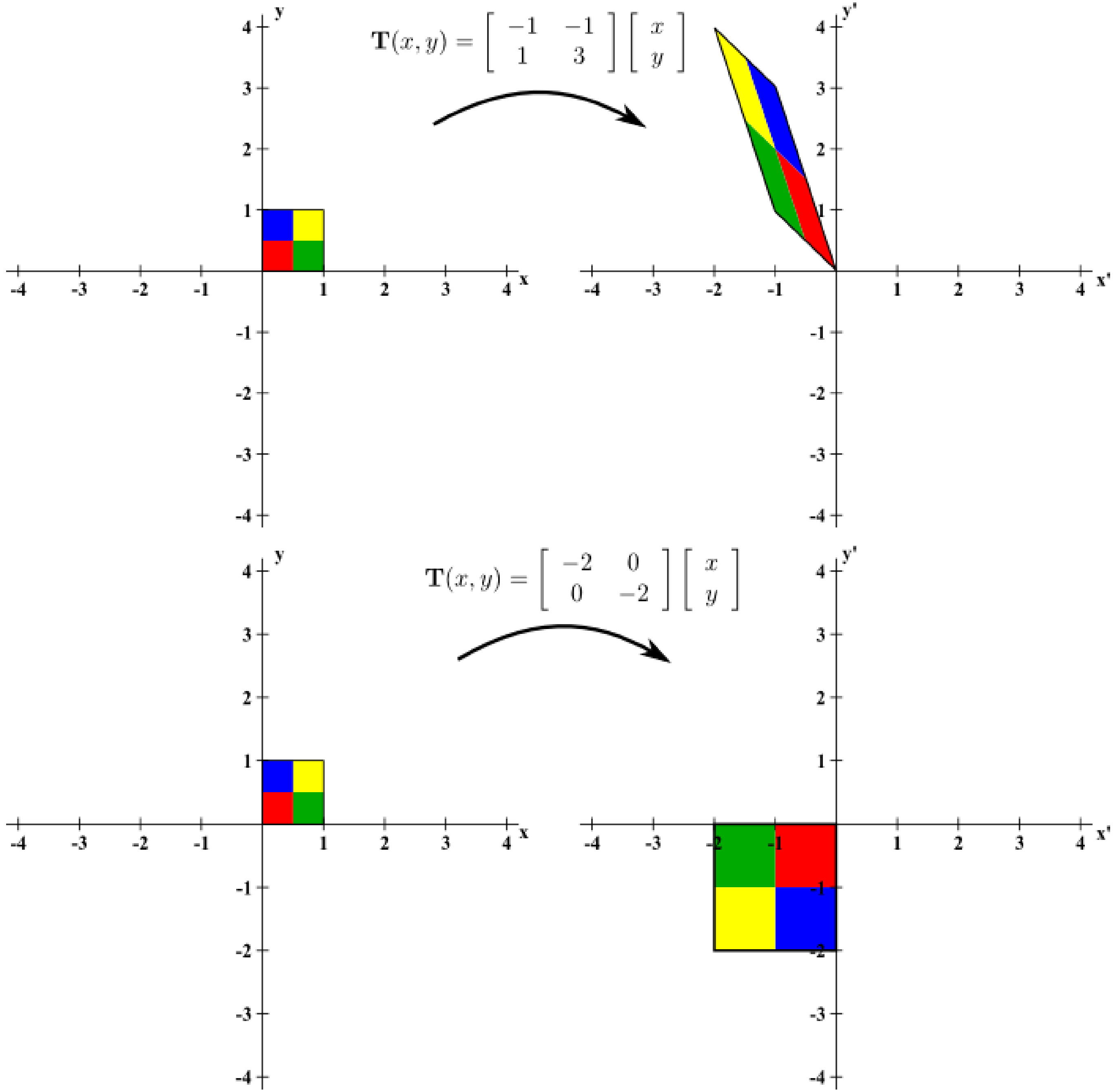
- Now for a serious question: how are matrices related to linear transformations?
- Every matrix encodes a linear transformation of linear spaces in a certain coordinate system: one space is mapped onto a subspace of another.
- All these transformations can be broken down into a combination of projections, rotations, reflections, and scaling along the coordinate axes. There are no other transformations.

Linear transformations

- One transformation can correspond to an infinite number of matrices—just as there are infinite coordinate systems. A matrix by itself is simply a table of numbers!

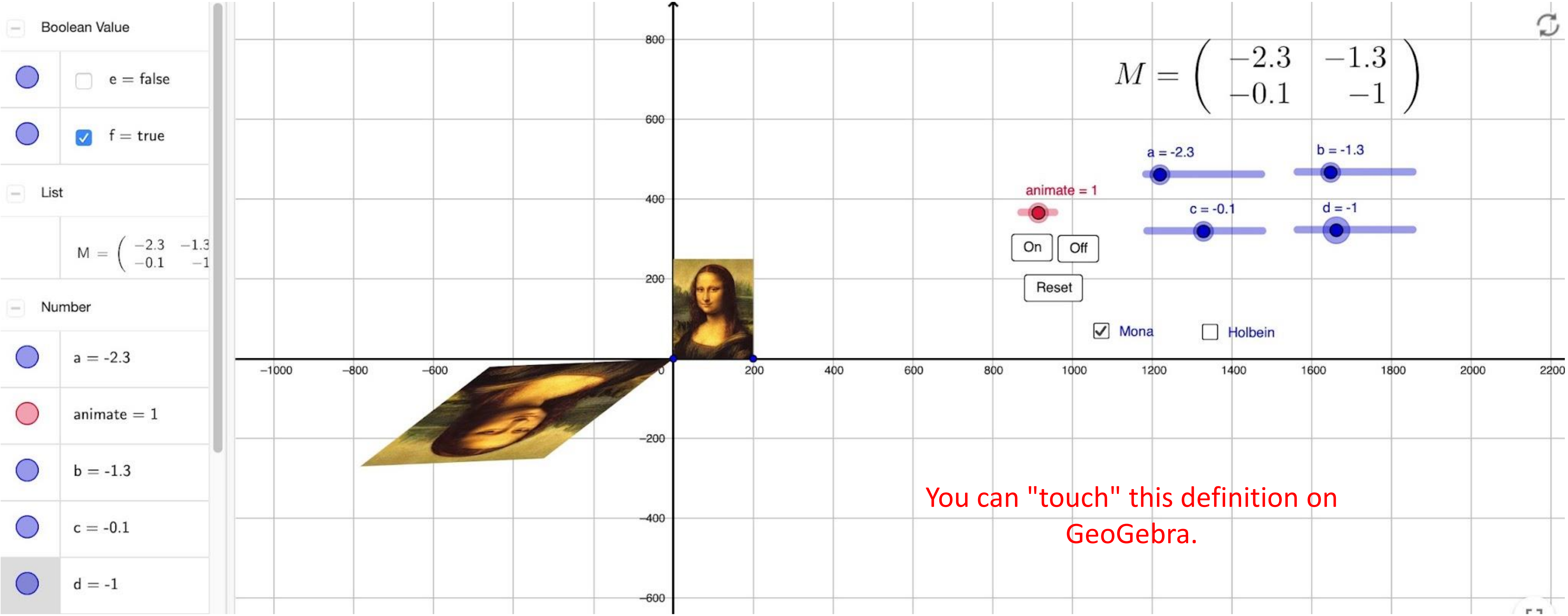
$$\begin{pmatrix} -1 & -1 \\ 1 & 3 \end{pmatrix} \cdot \begin{pmatrix} \quad \\ \quad \end{pmatrix} =$$





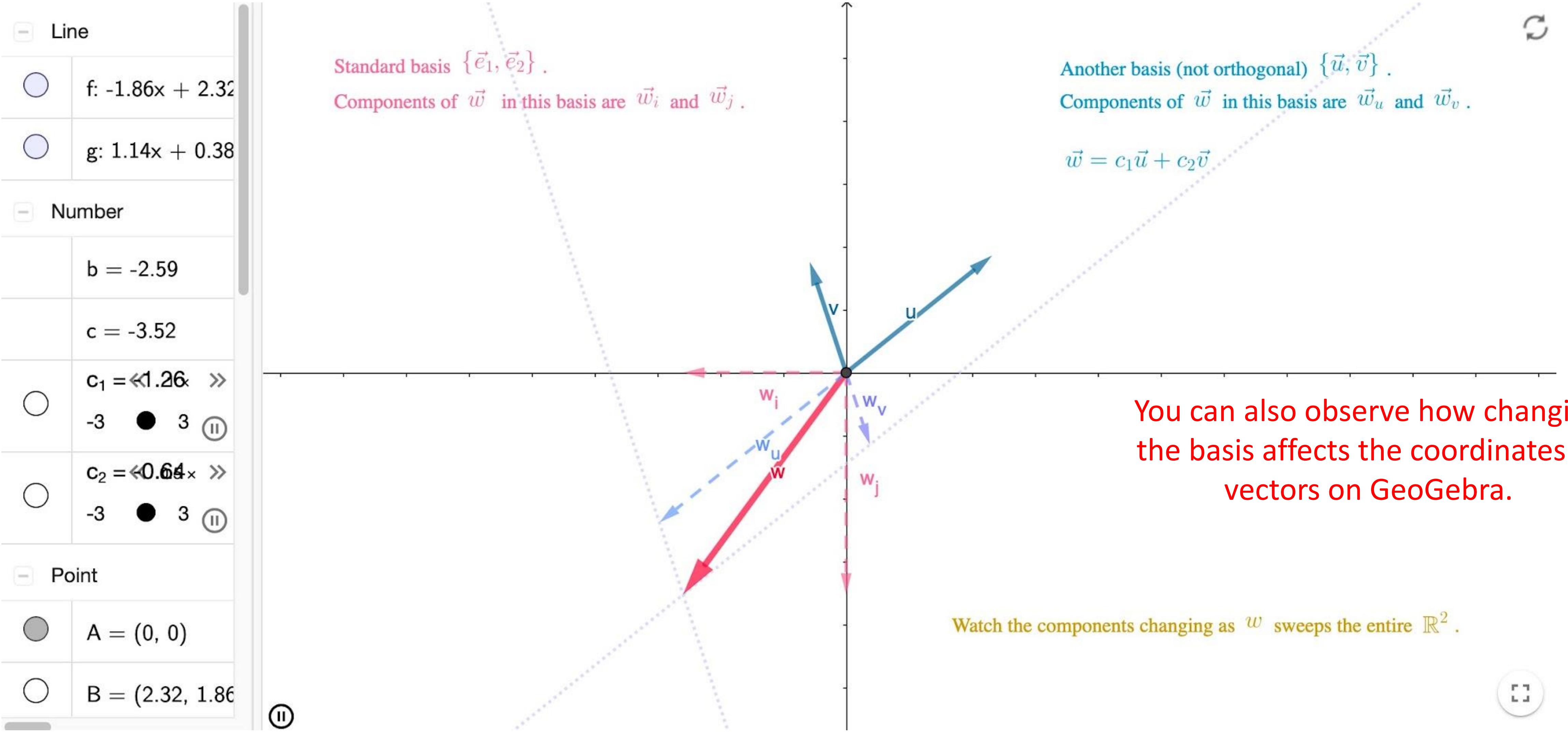
$\mathbf{T}(x,y) = \begin{bmatrix} -1 & -1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$
$\mathbf{T}(x,y) = \begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$
x
y
x'
y'

Linear transformations



You can "touch" this definition on GeoGebra.

Linear transformations



You can also observe how changing the basis affects the coordinates of vectors on GeoGebra.

Eigen directions

- How are all these wonders related to eigen directions and eigenvalues?
- Remember, there was something like that, right? :)

Eigen directions

- A linear transformation $A: V \to V$ has a certain number of eigen directions.
- Along these directions, it does not rotate or reflect but scales!
- These are examples of so-called invariant subspaces: they remain unchanged under the action of A.

- This is mathematically expressed as:
- $Av = \lambda v$
  - λ – eigenvalue
  - v – eigenvector

- For example:

$$A = \begin{pmatrix} 1 & 2 \\ -1 & 4 \end{pmatrix}$$

$$v = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

Eigen directions

- The concept of eigen directions is very useful because it allows us to find "good" directions of our transformation. Moreover, if there are as many eigenvectors as there are basis vectors, we can "transition" to a basis of eigenvectors, in which our matrix A will have a diagonal form, simplifying many calculations.

Eigen directions

- The concept of eigen directions is very useful because it allows us to find "good" directions of our transformation. Moreover, if there are as many eigenvectors as there are basis vectors, we can "transition" to a basis of eigenvectors, in which our matrix A will have a diagonal form, simplifying many calculations.

In eigen directions, for example, there can be the following physical interpretation: in the theory of vibrations, eigenvalues represent the frequencies of the system's oscillations during free motion, while eigenvectors are their "trajectories."

Matrix decompositions

Matrix decompositions

- Now that we have recalled and discussed the basic concepts of linear algebra (which we will need today), we can finally move on to the main subject of our study—matrix decompositions!

Matrix decompositions

- Now that we have recalled and discussed the basic concepts of linear algebra (which we will need today), we can finally move on to the main subject of our study—matrix decompositions!
- Let's take another look at the general form of the feature matrix.

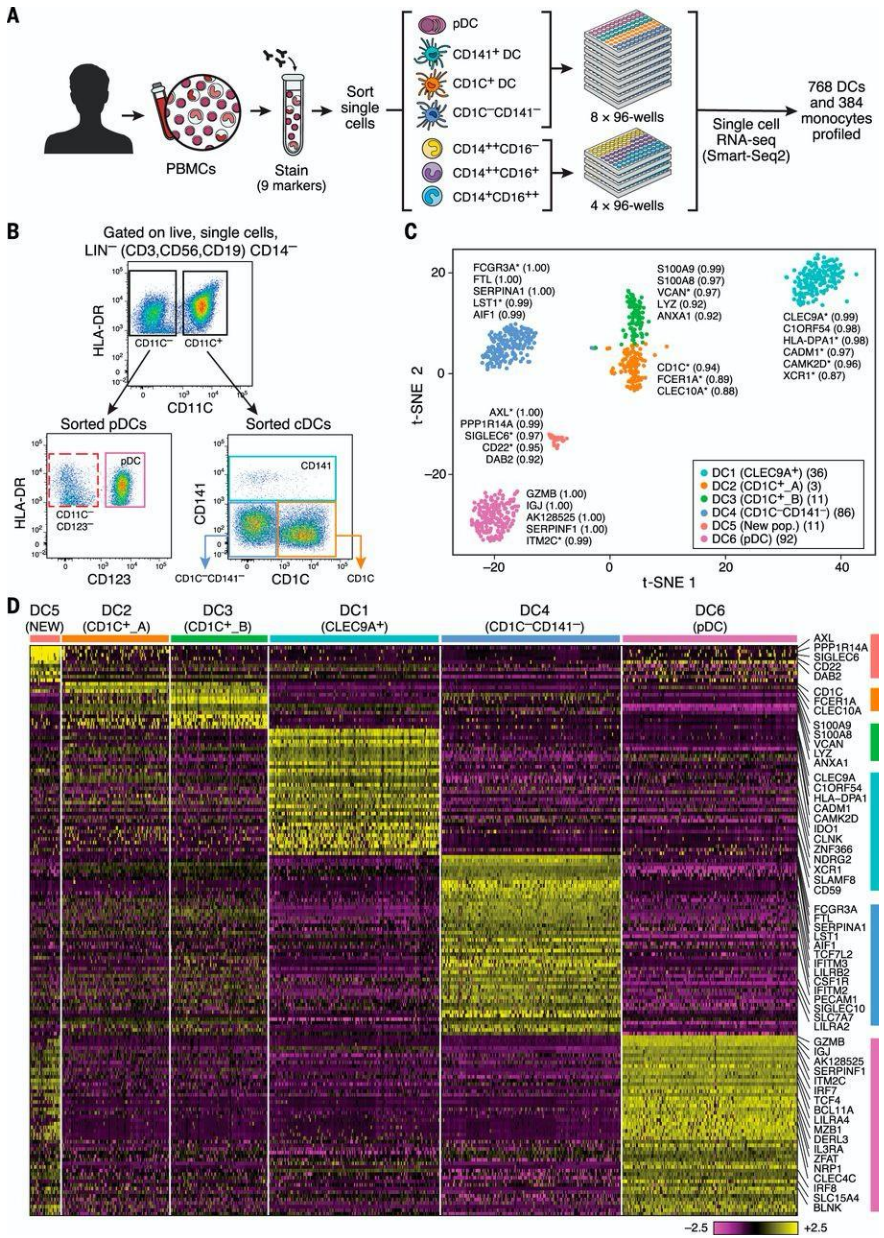
Feature matrix

- Feature matrices are rectangular real matrices of size $m \times n$, where:
  - m is the sample size
  - n is the number of features (and the size of the feature space)

$$
\begin{array}{cc}
\begin{array}{ccccccc} \text{Feature-1} & \text{Feature-2} & \text{Feature-3} & \text{Feature-4} & & & \text{Feature-n} \end{array} & \\
\begin{bmatrix}
x_1^1 & x_2^1 & x_3^1 & x_4^1 & \dots & \dots & x_n^1 \\
x_1^2 & x_2^2 & x_3^2 & x_4^2 & \dots & \dots & x_n^2 \\
x_1^3 & x_2^3 & x_3^3 & x_4^3 & \dots & \dots & x_n^3 \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots \\
x_1^m & x_2^m & x_3^m & x_4^m & \dots & \dots & x_n^m
\end{bmatrix}
&
\begin{array}{l} \text{Sample-1} \\ \text{Sample-2} \\ \text{Sample-3} \\ \\ \text{Sample-m} \end{array}
\end{array}
$$

Feature matrix

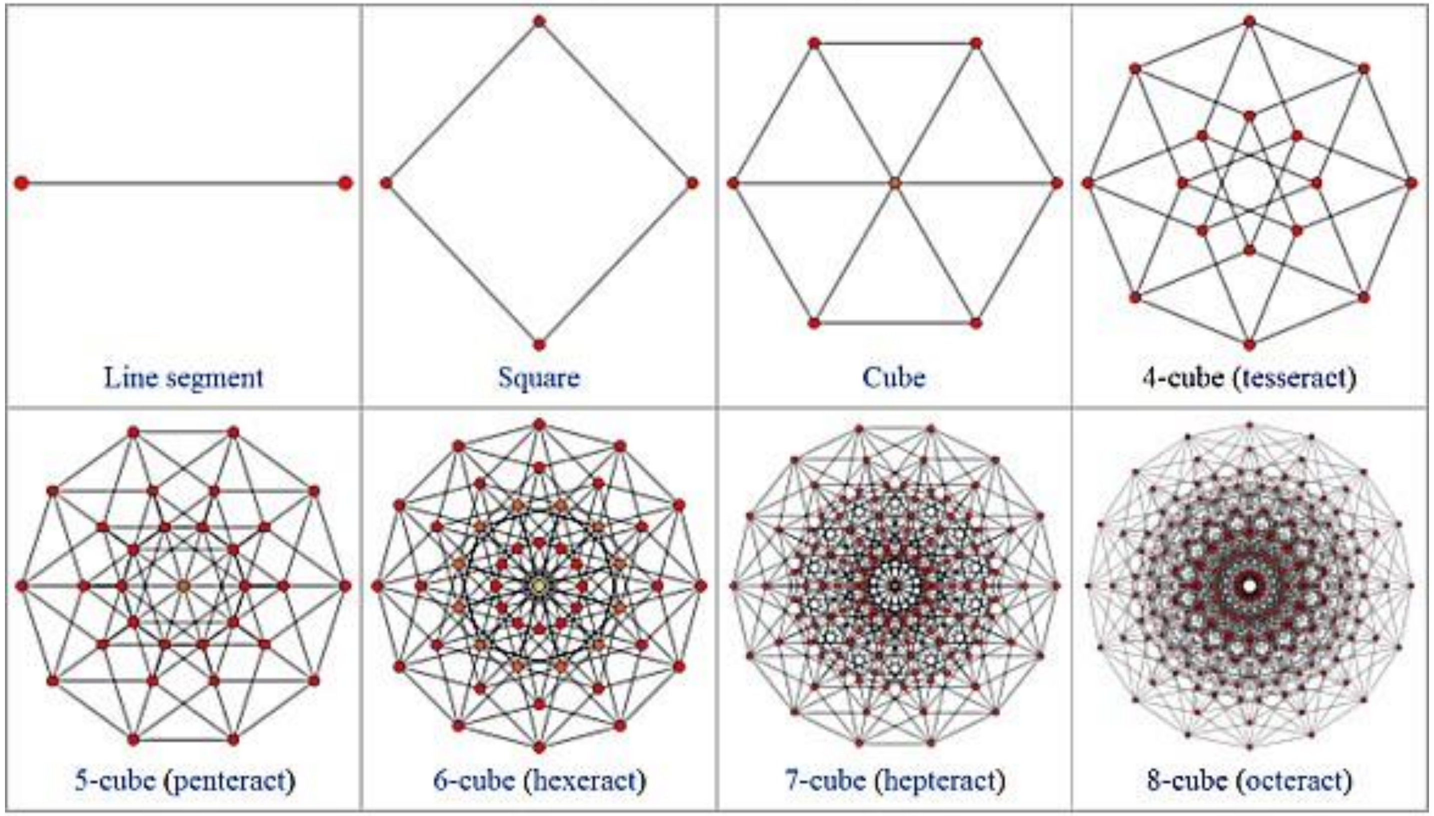
- Linear algebra is a tool that allows you to work with data of immense dimensions.
- Example:
- A typical tumor cell in sequencing data is represented by 14,000 numbers, and a sample usually contains from a hundred to a million cells.



Feature matrix

- Nevertheless, our brain is incapable of perceiving high-dimensional data.
- Everything beyond dimensionality of 7 (length, width, height, three colors, time) is practically beyond our perception.
- Even the simplest geometric objects like a cube become nearly impossible to visualize without losing information.

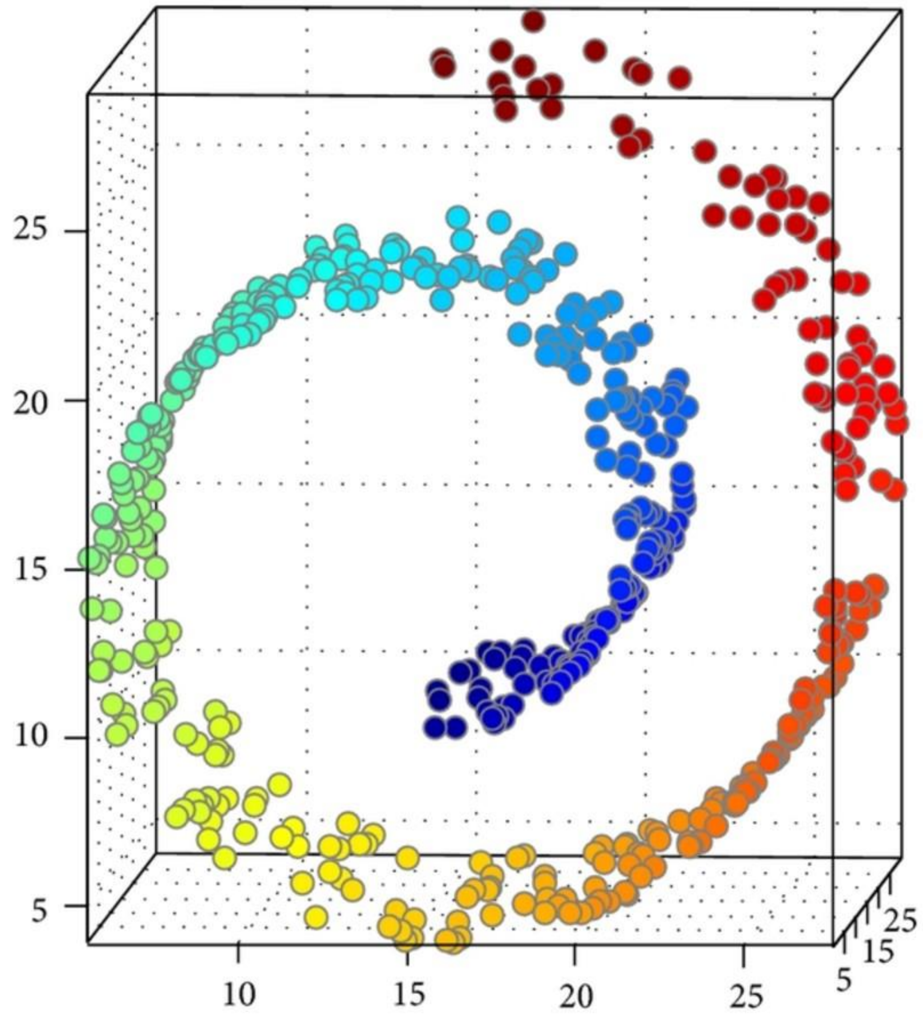
Feature matrix



Laying out multidimensional cubes
(hyper cubes) on a plane

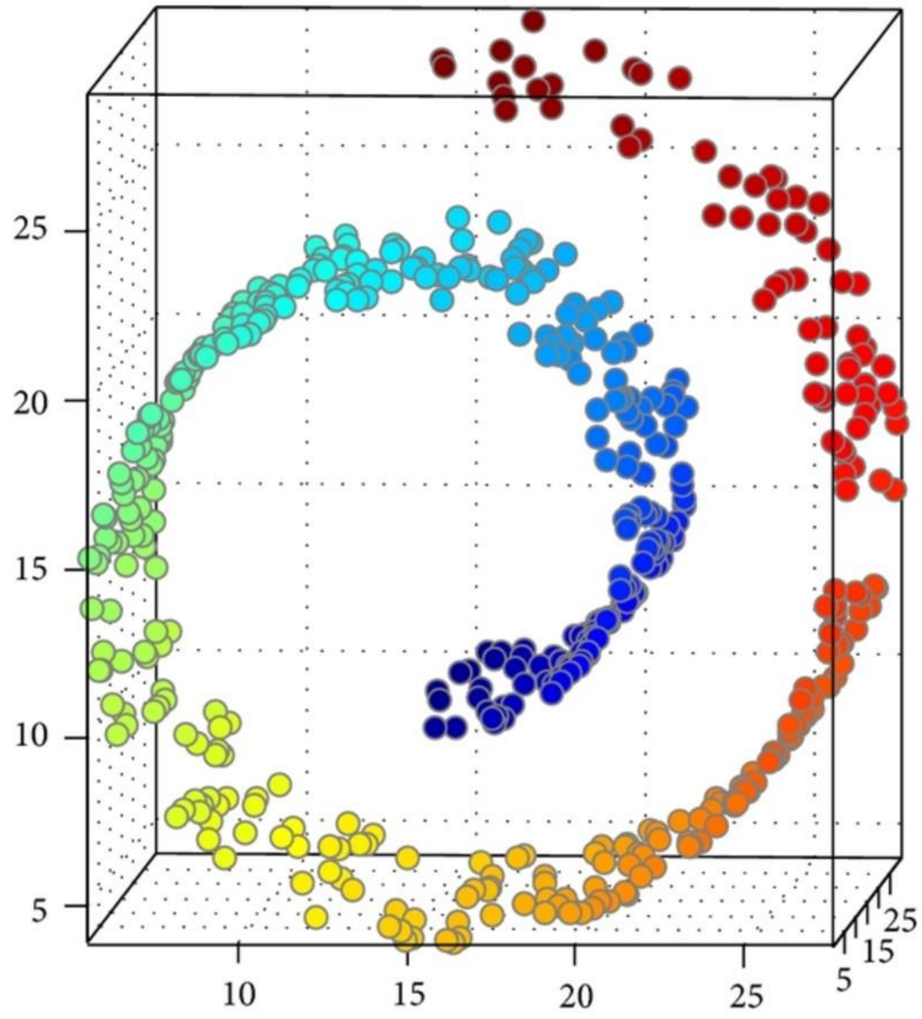
Feature matrix

- Fortunately, real data usually lie on so-called low-dimensional manifolds.
- They are not uniformly distributed across the multidimensional space but reside on surfaces of significantly lower dimensionality.



Feature matrix

- Fortunately, real data usually lie on so-called low-dimensional manifolds.
- They are not uniformly distributed across the multidimensional space but reside on surfaces of significantly lower dimensionality.
- Consider this example: Formally, these are three-dimensional data, but in fact, they almost entirely lie on a two-dimensional strip!

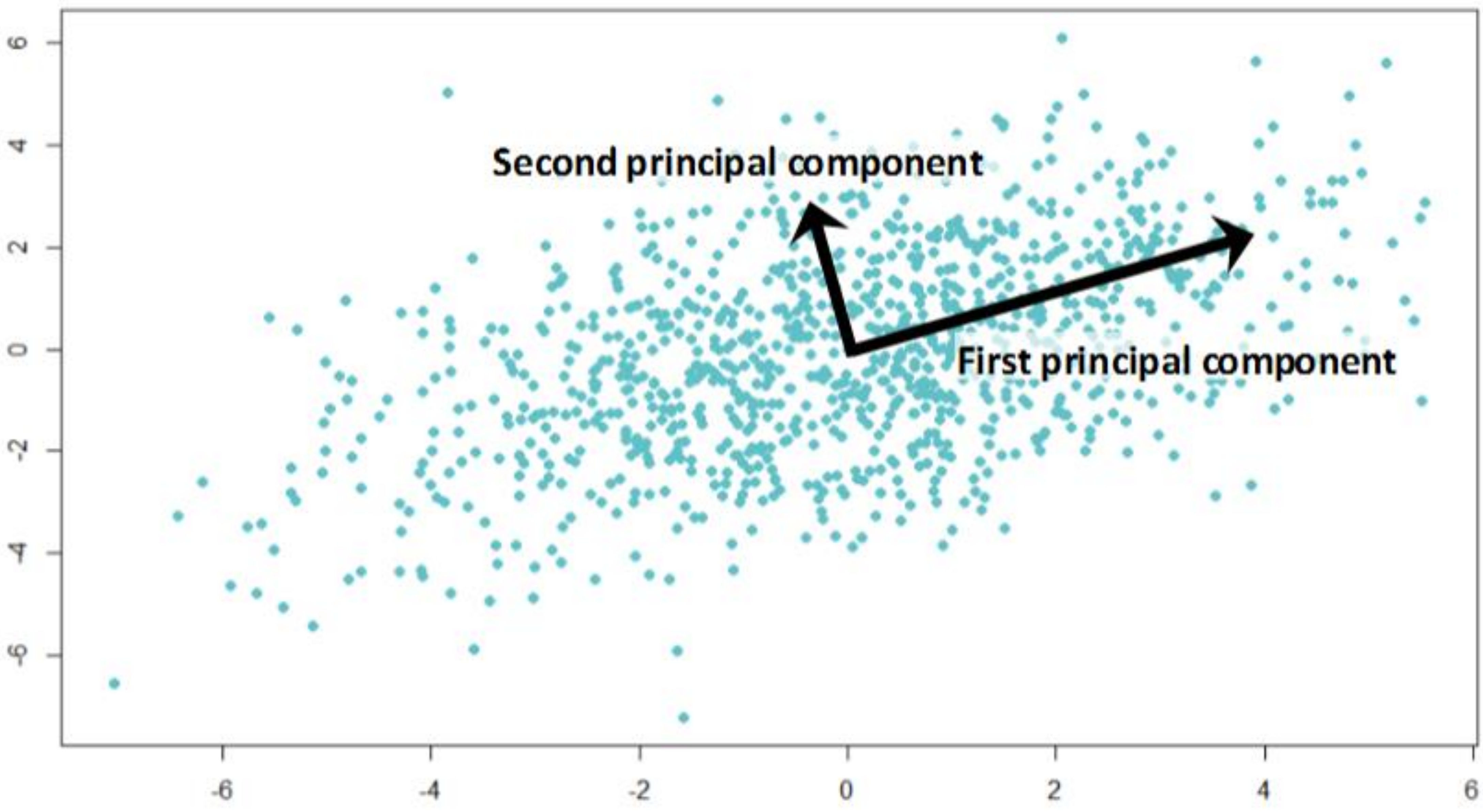


Feature matrix

- The field of machine learning that deals with defining the subtle geometric structure of data is called manifold learning.
- However, it is often possible to reduce dimensionality, for example from 14,000 to 30-100, without significant loss of information using simpler methods.
- For instance, using PCA (Principal Component Analysis), which seeks the optimal dimensionality reduction for your data matrix.

Principal Component Analysis (PCA)

- The PCA algorithm can be described by the following steps:
  1. Find the direction in space along which the variance of the data is maximized.
  2. Among the remaining directions, orthogonal to the previous ones, find the direction along which the variance is maximized.
  3. Repeat step 2 until no more directions can be found.
- The directions obtained are called principal components.


Second principal component
First principal component
6
4
2
0
-2
-4
-6
-6
-4
-2
0
2
4
6

Principal Component Analysis (PCA)

- But how do we find these directions?!

Principal Component Analysis (PCA)

- But how do we find these directions?!
- We could find these directions by exhaustive search, but can you imagine how long that would take, even in a 10-dimensional space?...

Principal Component Analysis (PCA)

- But how do we find these directions?!
- We could find these directions by exhaustive search, but can you imagine how long that would take, even in a 10-dimensional space?...
- Therefore, as true mathematicians, we will use the available mathematical tools to solve this task!

Important concepts in linear algebra

- An orthogonal matrix is a matrix for which the inverse matrix is equal to its transpose.

$$A \cdot A^T = A \cdot A^{-1} = I$$

$$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

Important concepts in linear algebra

- Singular directions and values:

$$Mv = \sigma u$$
$$M^T u = \sigma v$$

- where
  - v is the right singular vector,
  - u is the left singular vector,
  - σ is the singular value.
- A singular value is the square root of an eigenvalue of the matrix $M^T M$.

Important concepts in linear algebra

- Singular directions and values:
- where
  - v is the right singular vector,
  - u is the left singular vector,
  - σ is the singular value.
- A singular value is the square root of an eigenvalue of the matrix $M^T M$.

$$Mv = \sigma u$$
$$M^T u = \sigma v$$

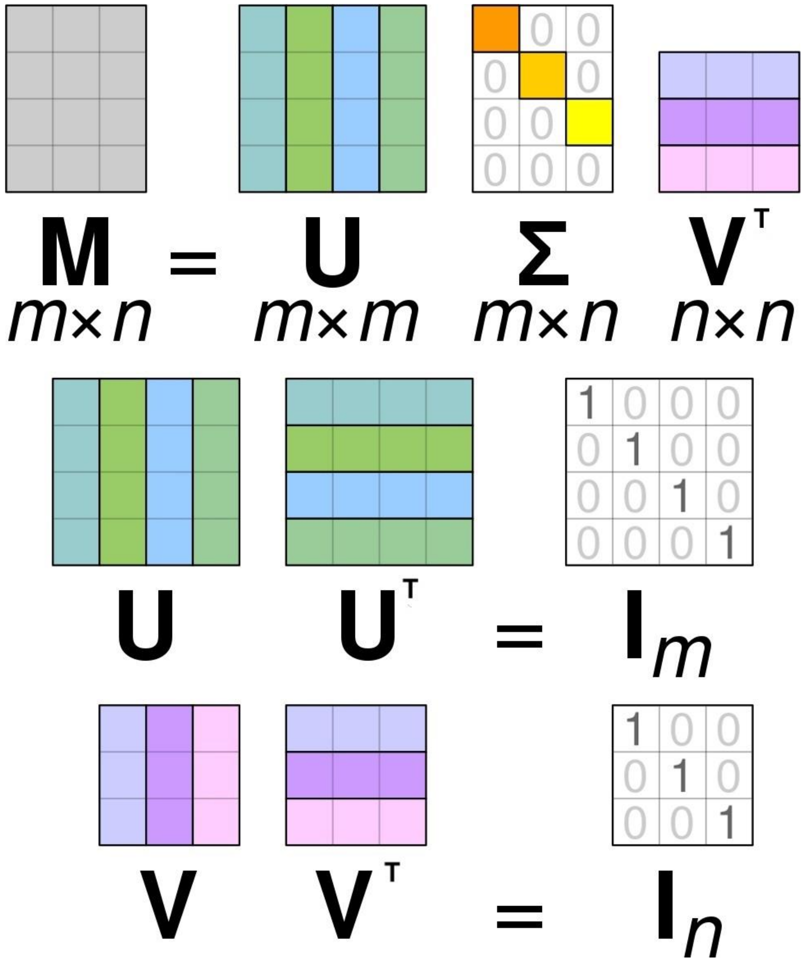
Singular vectors and values are, in a sense, analogues of eigenvectors and eigenvalues for non-square matrices. Moreover, if the matrix is square, the square of the singular value equals the modulus of the corresponding eigenvalue (if it exists).

Singular Value Decomposition (SVD)

- It turns out that any matrix M can be decomposed into the product of three matrices U , Σ and V^T such that:

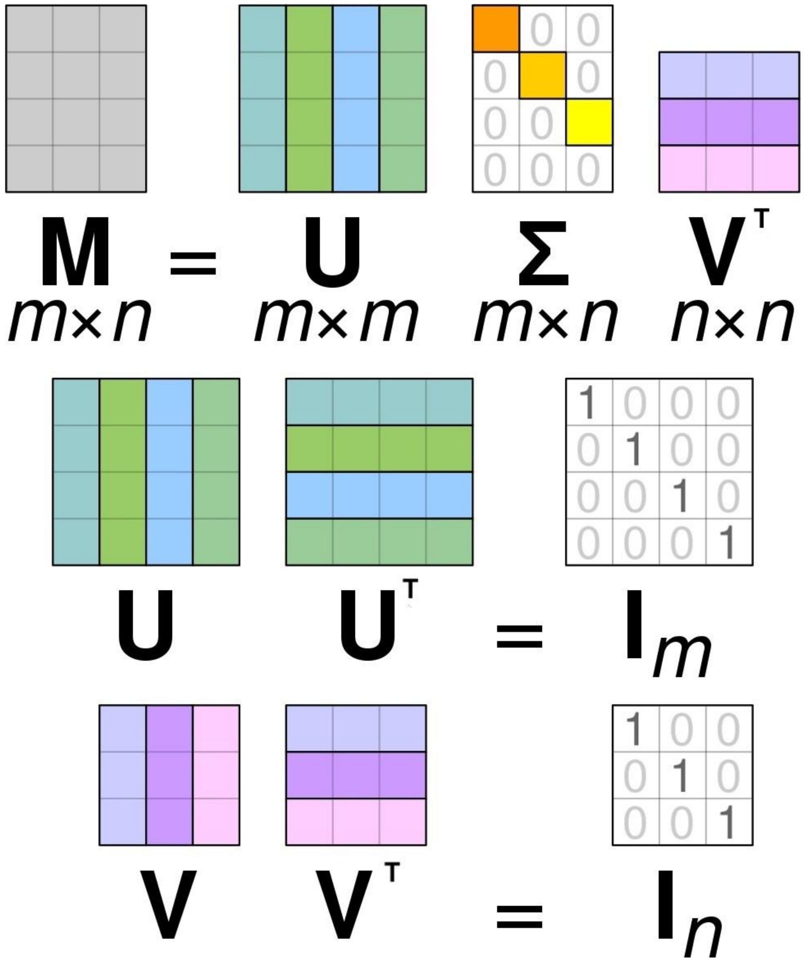
Singular Value Decomposition (SVD)

- It turns out that any matrix M can be decomposed into the product of three matrices U , Σ and V^T such that:

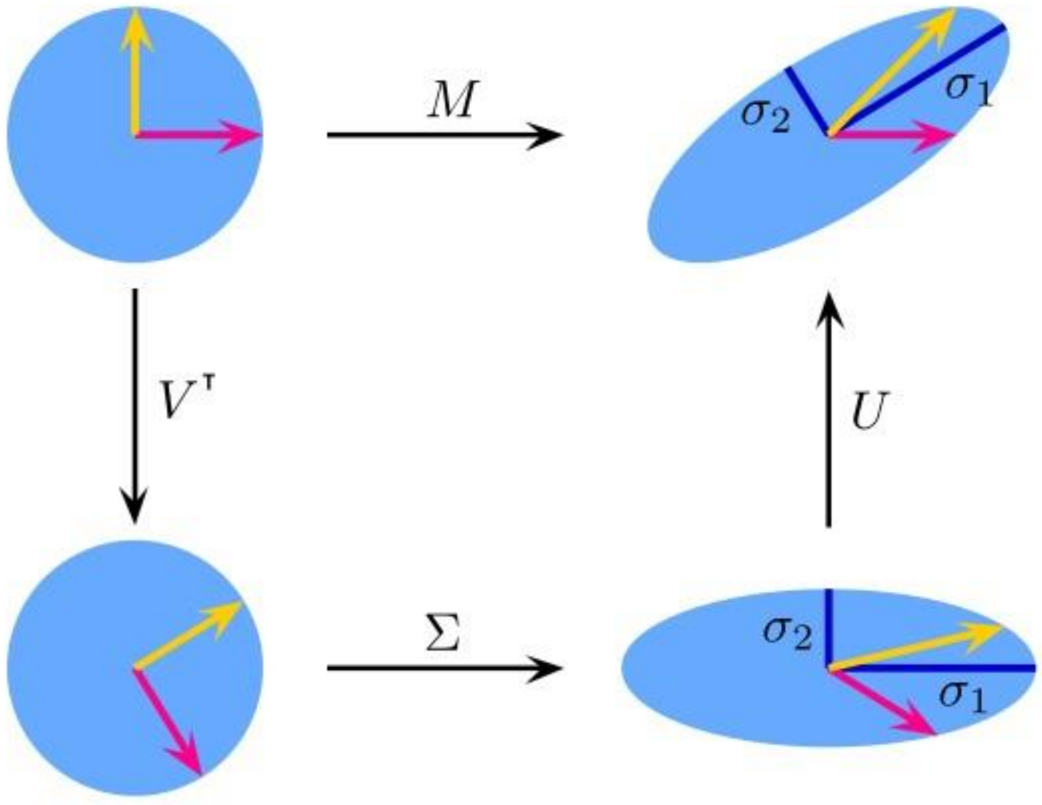


Singular Value Decomposition (SVD)

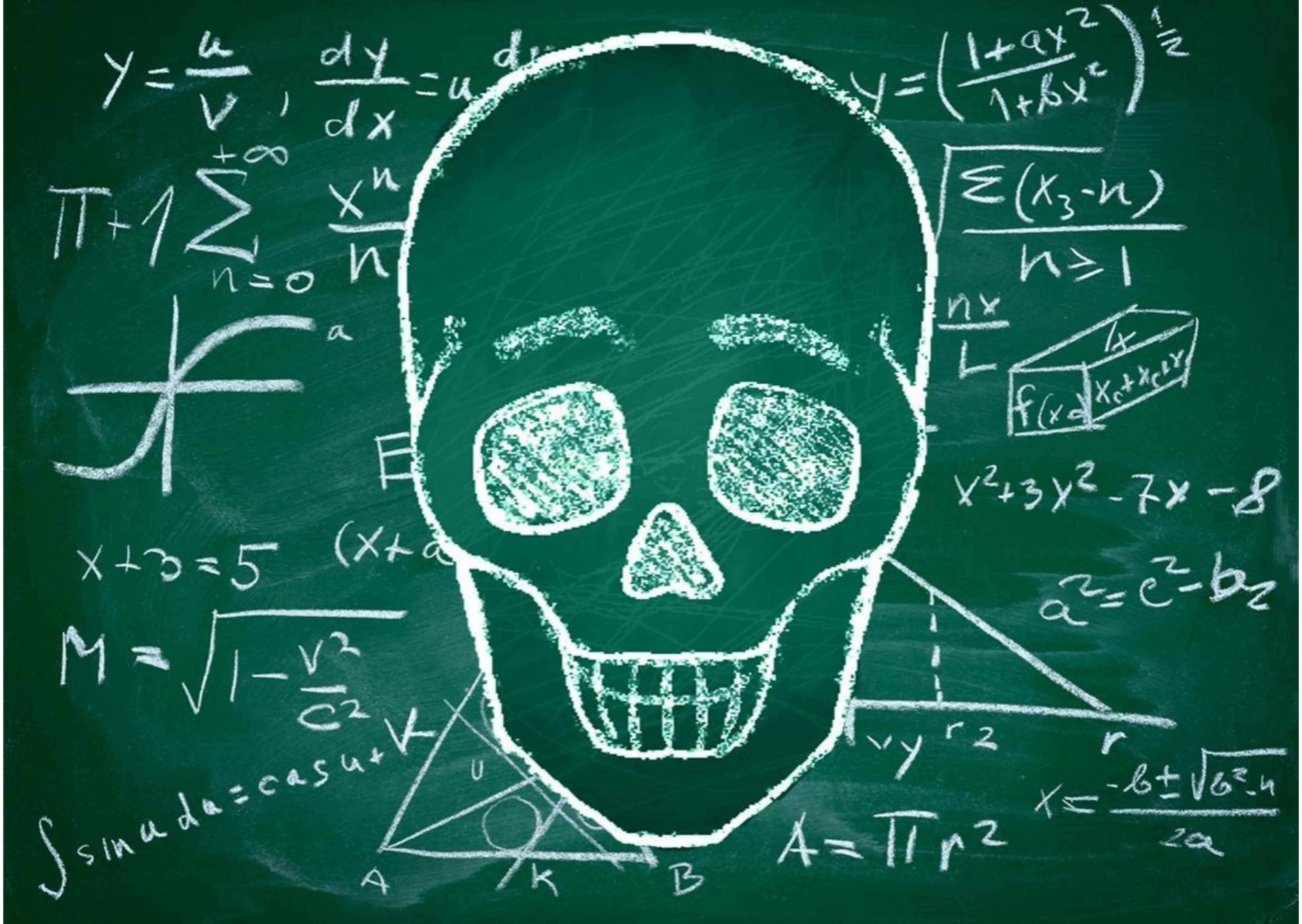
- It turns out that any matrix M can be decomposed into the product of three matrices U , Σ and V^T such that:
  - $M \in \mathbb{R}^{m \times n}$ — The original matrix;
  - $U \in \mathbb{R}^{m \times m}$— The matrix of left singular vectors;
  - $\Sigma \in \mathbb{R}^{m \times n}$— The matrix whose main diagonal contains the so-called singular values (in non-increasing order);
  - $V \in \mathbb{R}^{n \times n}$— The matrix of right singular vectors.



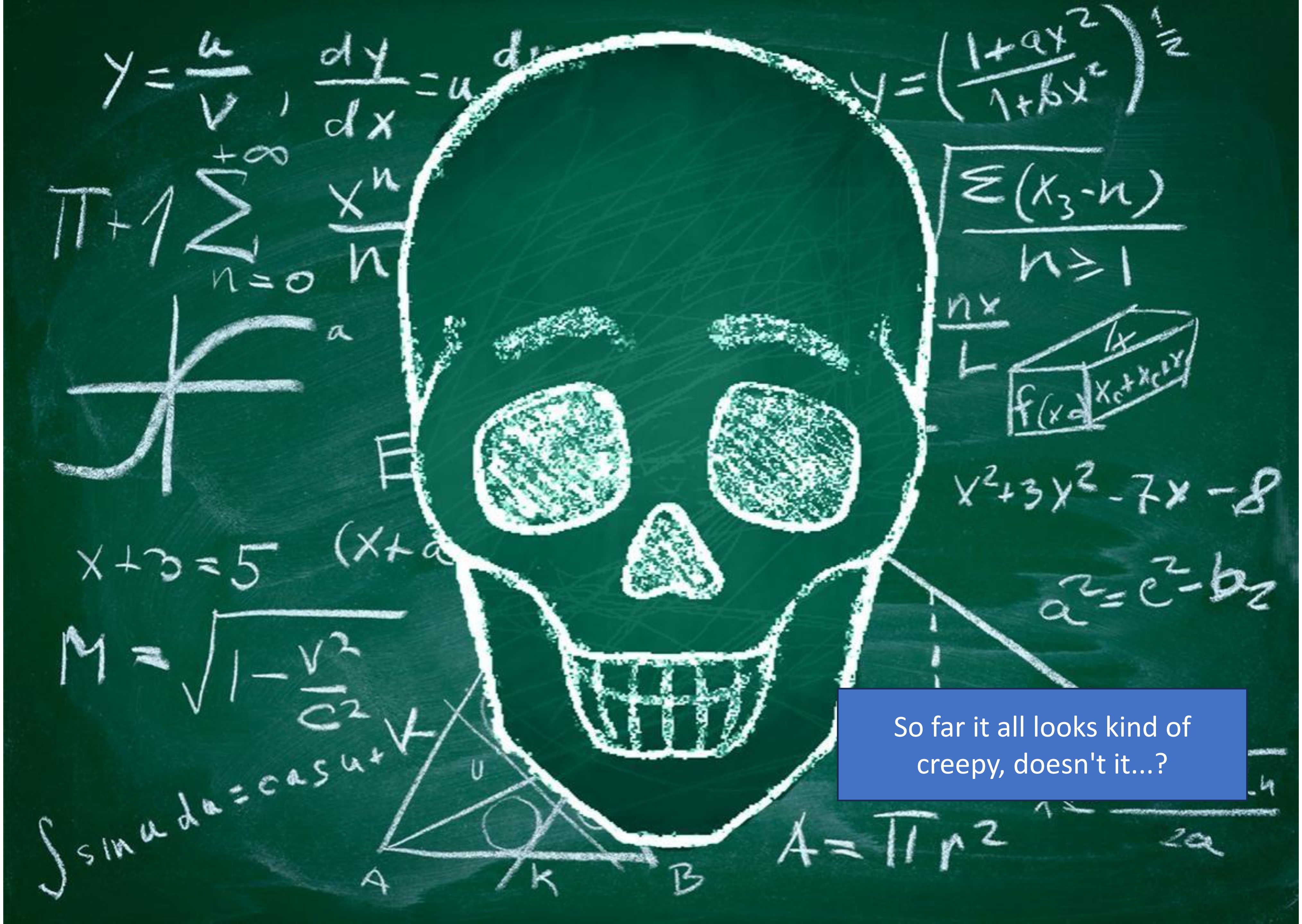
Singular Value Decomposition (SVD)



$$M = U \cdot \Sigma \cdot V^\intercal$$



x+3=5
A=πr²
n≥1
x²+3x²-7x-8

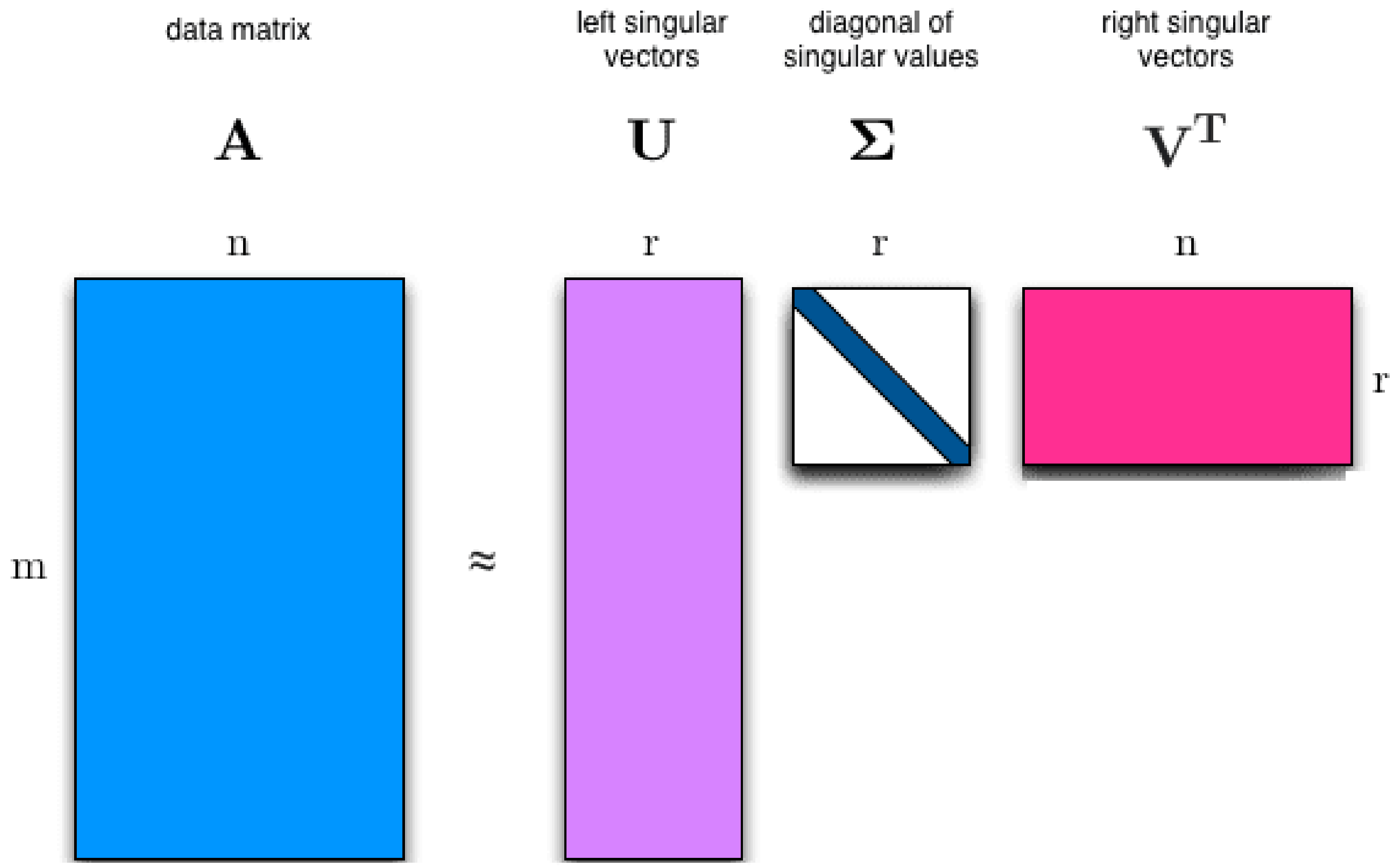


So far it all looks kind of creepy, doesn't it...?

Truncated SVD decomposition of rank r

- Truncated SVD decomposition of rank r for a matrix $M \in \mathbb{R}^{m \times n}$ is a regular SVD decomposition where only the largest singular values and their corresponding singular vectors are retained.

Truncated SVD decomposition of rank r



Truncated SVD decomposition of rank r

- Truncated SVD decomposition of rank r for a matrix $M \in \mathbb{R}^{m \times n}$ is a regular SVD decomposition where only the largest singular values and their corresponding singular vectors are retained.

Truncated SVD decomposition of rank r

- Truncated SVD decomposition of rank r for a matrix $M \in \mathbb{R}^{m \times n}$ is a regular SVD decomposition where only the largest singular values and their corresponding singular vectors are retained.
- It can be proven that the truncated SVD decomposition of rank r is the optimal rank- r approximation in terms of the Frobenius norm.

Truncated SVD decomposition of rank r

- Truncated SVD decomposition of rank r for a matrix $M \in \mathbb{R}^{m \times n}$ is a regular SVD decomposition where only the largest singular values and their corresponding singular vectors are retained.
- It can be proven that the truncated SVD decomposition of rank r is the optimal rank- r approximation in terms of the Frobenius norm.

Truncated SVD decomposition of rank r

- Truncated SVD decomposition of rank r for a matrix $M \in \mathbb{R}^{m\times n}$ is a regular SVD decomposition where only the largest singular values and their corresponding singular vectors are retained.
- It can be proven that the truncated SVD decomposition of rank r is the optimal rank- r approximation in terms of the Frobenius norm.

The Frobenius norm, or Euclidean norm (for Euclidean space), represents a special case of p -norm for p =2:

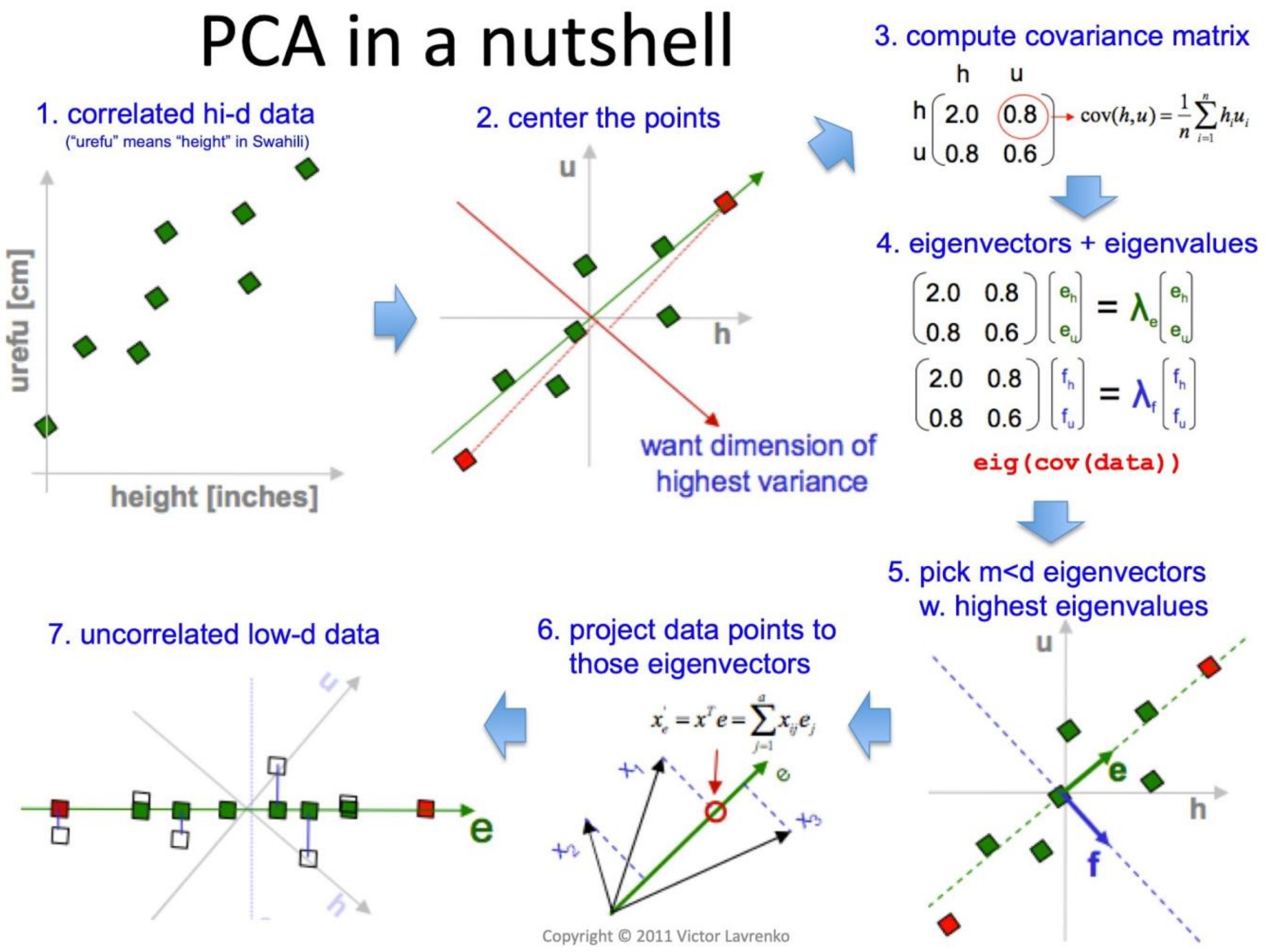
$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}\left|a_{ij}\right|^2}$$

Truncated SVD decomposition of rank r

- Truncated SVD decomposition of rank r for a matrix $M \in \mathbb{R}^{m \times n}$ is a regular SVD decomposition where only the largest singular values and their corresponding singular vectors are retained.
- It can be proven that the truncated SVD decomposition of rank r is the optimal rank-r approximation in terms of the Frobenius norm.
- Congratulations, it is precisely because of this fact that we can finalize the algorithm for Principal Component Analysis! :)

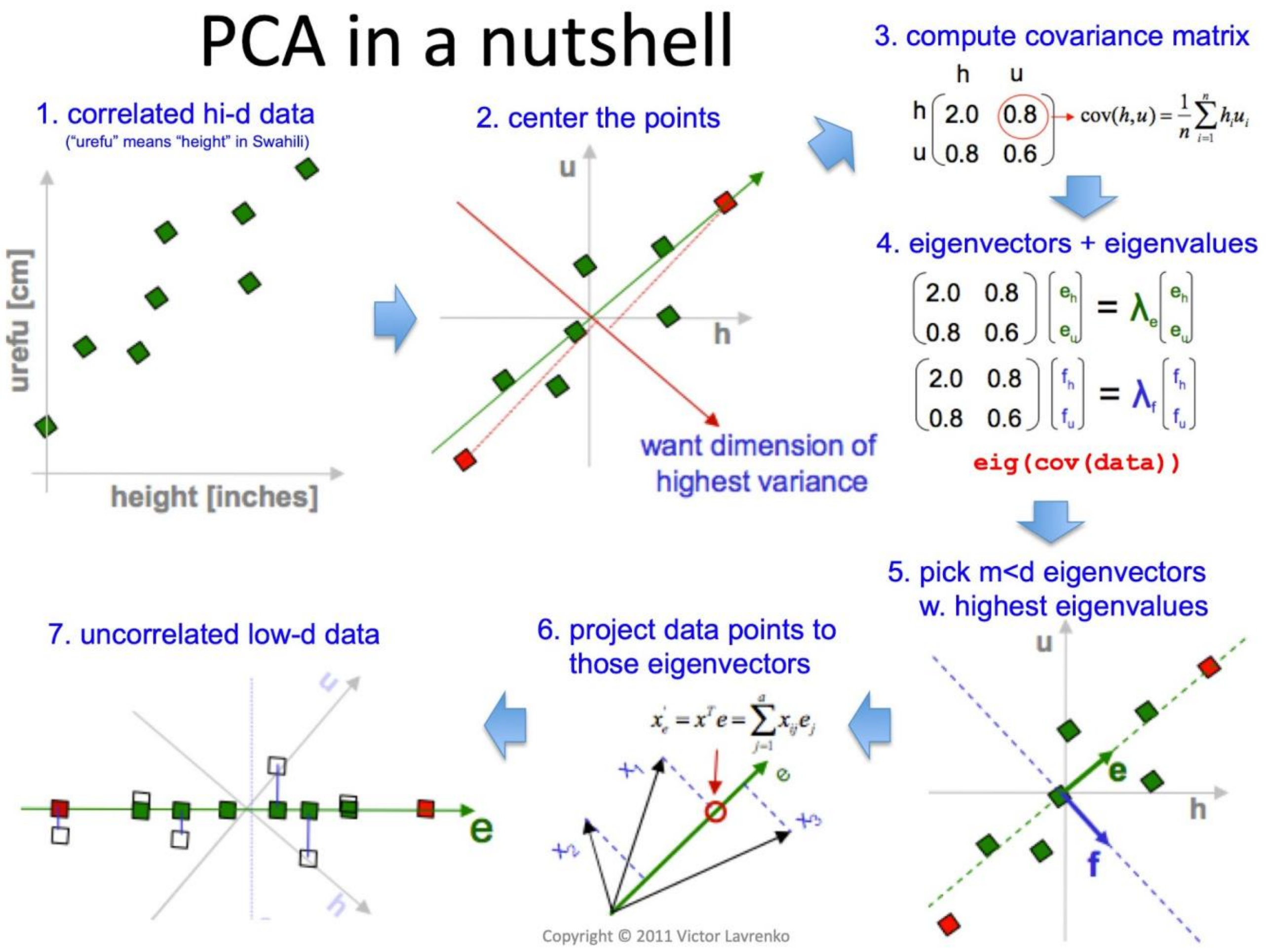
The connection between SVD and PCA

- PCA transformation could be performed by finding eigenvectors (as in the algorithm on the right), but unfortunately, our datasets almost always have different numbers of rows and columns, and therefore we are compelled to use SVD.



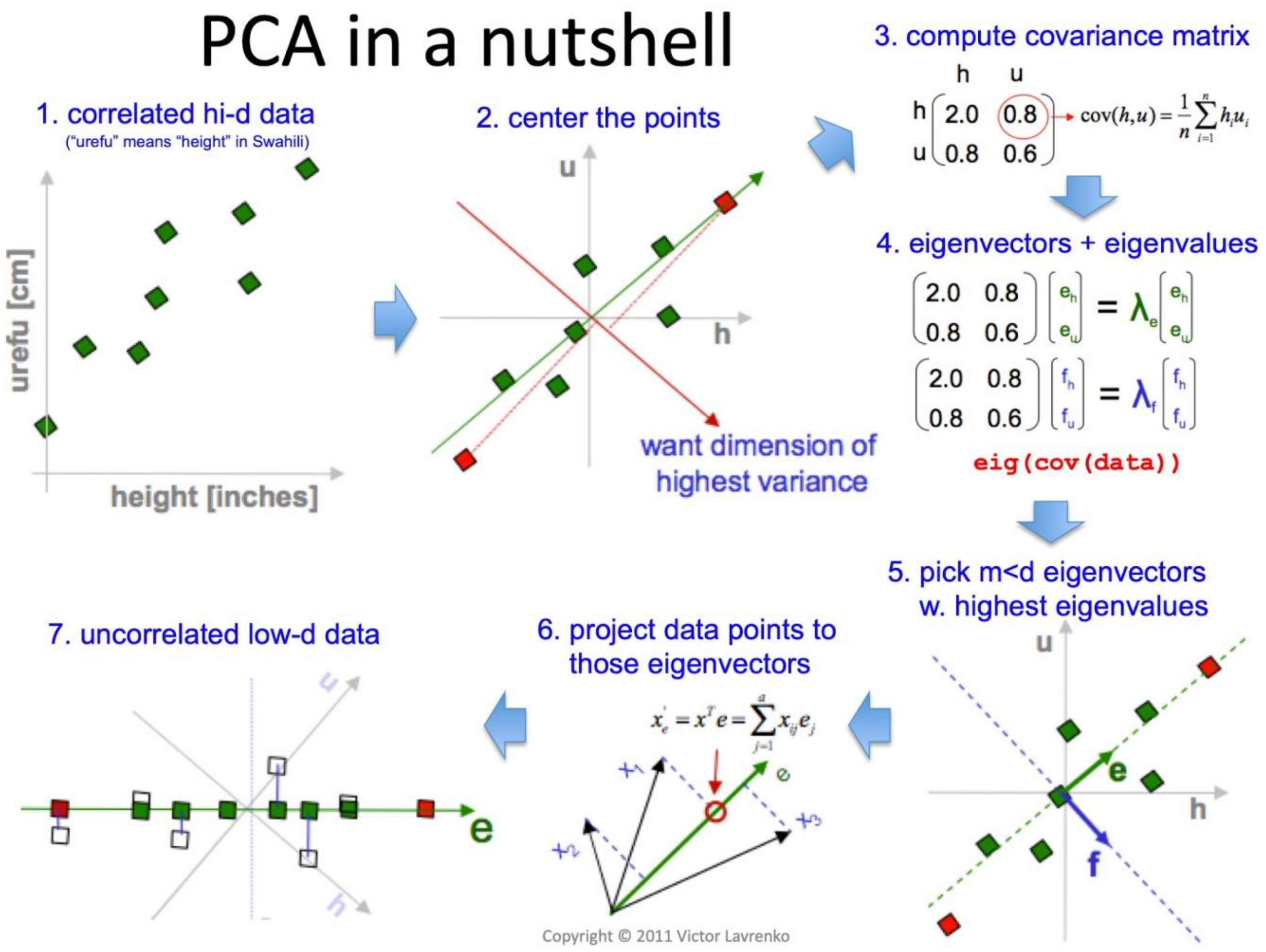
The connection between SVD and PCA

- But there's good news! To obtain the PCA transformation, it is sufficient to simply calculate $U\Sigma$, which is the product of the first two matrices in the SVD decomposition.



The connection between SVD and PCA

- But there's good news! To obtain the PCA transformation, it is sufficient to simply calculate $U\Sigma$, which is the product of the first two matrices in the SVD decomposition.
- In more detail, we will definitely analyze it at the seminars!



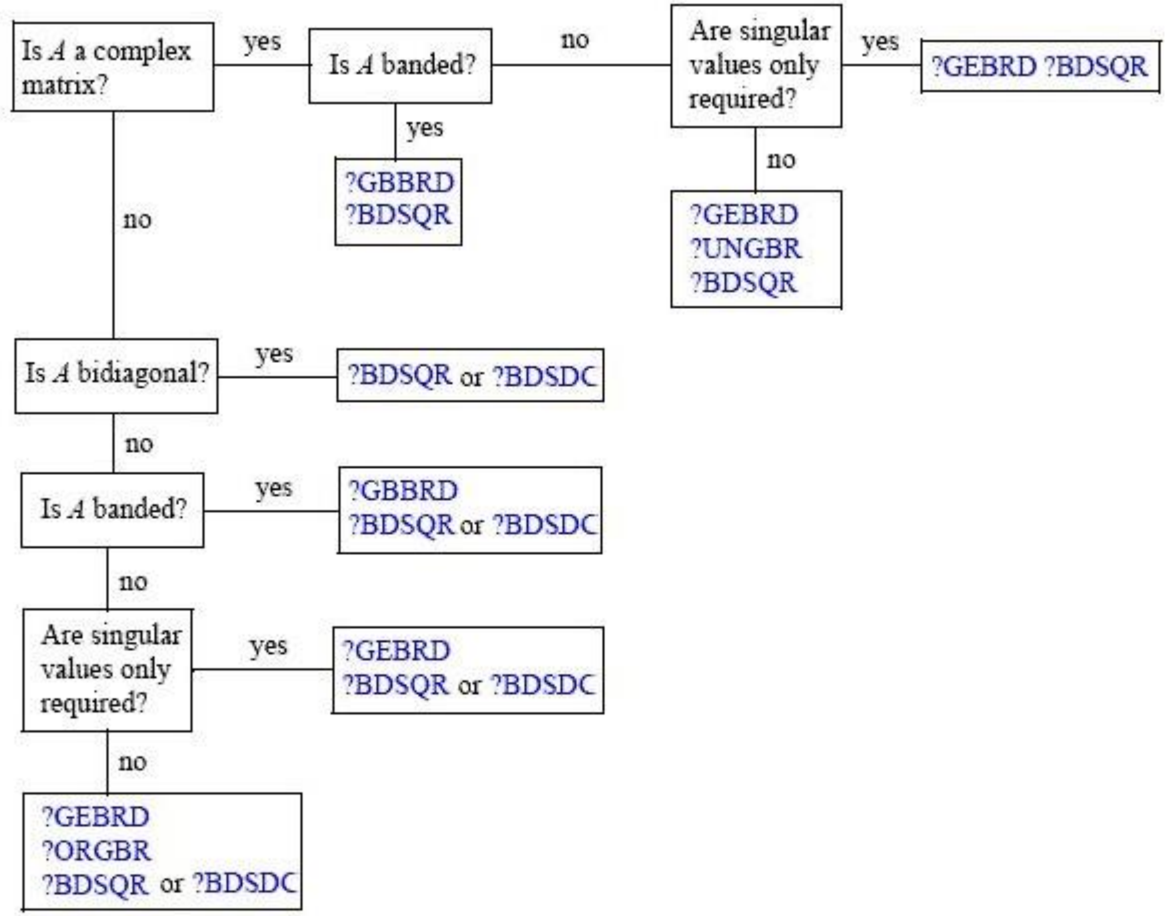
The connection between SVD and PCA

- The competitive advantage of SVD decomposition lies in the availability of efficient algorithms for its computation.
- This includes algorithms for huge sparse matrices, which are often encountered in practice, such as in recommendation systems, bioinformatics, and the like.

The connection between SVD and PCA

- The competitive advantage of SVD decomposition lies in the availability of efficient algorithms for its computation.
- This includes algorithms for huge sparse matrices, which are often encountered in practice, such as in recommendation systems, bioinformatics, and the like.

A simplified scheme for computing SVD decomposition. Each case has its own algorithm! And this is without considering the sparsity of the matrix.



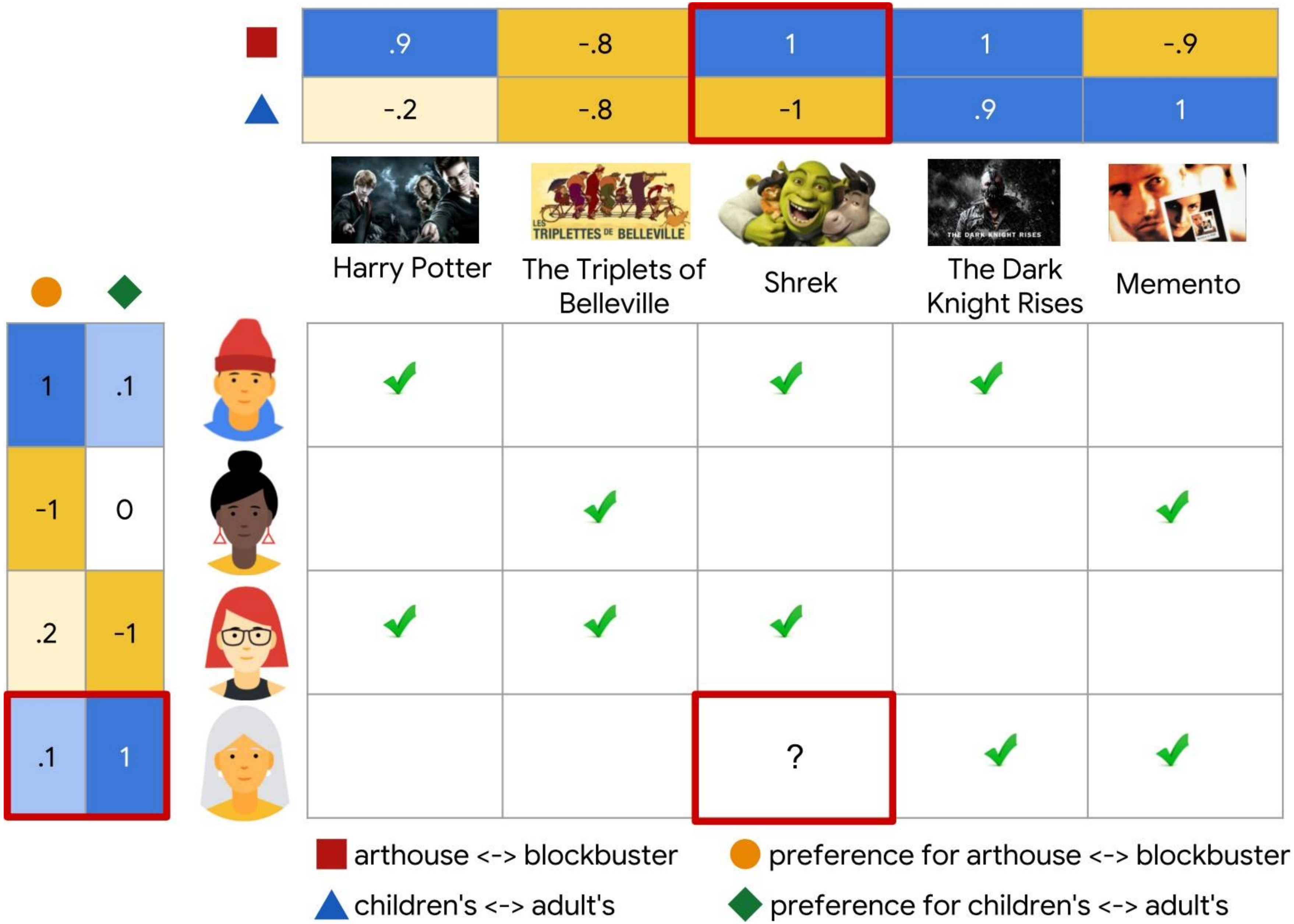
The connection between SVD and PCA

- The competitive advantage of SVD decomposition lies in the availability of efficient algorithms for its computation.
- This includes algorithms for huge sparse matrices, which are often encountered in practice, such as in recommendation systems, bioinformatics, and the like.
- But PCA is not the only good use of SVD!

SVD Applications

SVD Applications

- In the task of building recommendation systems, the matrix of user interactions with content plays an important role.



.9
-.8
1
1
-.9
-.2
-.8
-1
.9
1
Harry Potter
The Triplets of Belleville
LES TRIPLETTES DE BELLEVILLE
Shrek
The Dark Knight Rises
THE DARK KNIGHT RISES
Memento
1
.1
-1
0
.2
-1
.1
1
?
arthouse <-> blockbuster
children's <-> adult's
preference for arthouse <-> blockbuster
preference for children's <-> adult's

SVD Applications

- In the task of building recommendation systems, the matrix of user interactions with content plays an important role.
- Imagine Netflix (familiar to us):

SVD Applications

- In the task of building recommendation systems, the matrix of user interactions with content plays an important role.
- Imagine Netflix (familiar to us):
  - Millions of users;
  - Thousands of movies and shows;
  - Each user, on average, watches < 100 of them.

SVD Applications

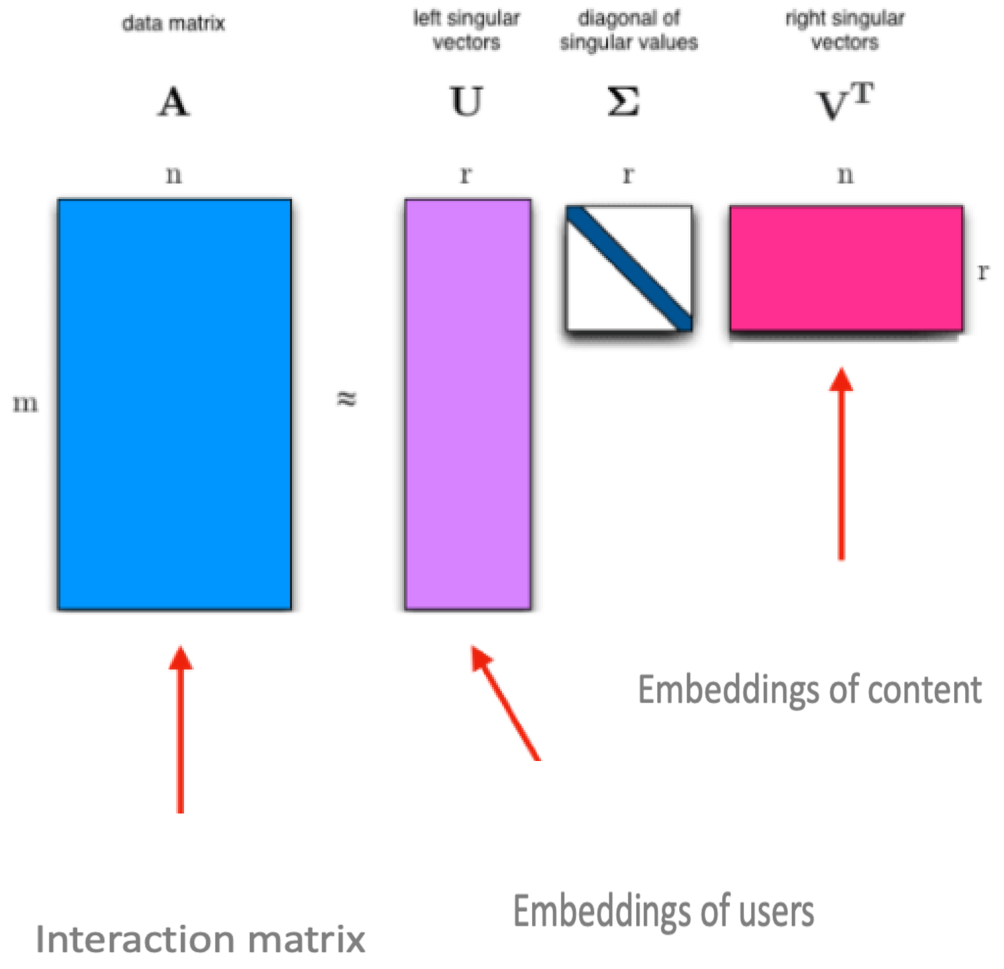
- In the task of building recommendation systems, the matrix of user interactions with content plays an important role.
- Imagine Netflix (familiar to us):
    - Millions of users;
    - Thousands of movies and shows;
    - Each user, on average, watches < 100 of them.
- What can we say about the score matrix?

SVD Applications

- In the task of building recommendation systems, the matrix of user interactions with content plays an important role.
- Imagine Netflix (familiar to us):
  - Millions of users;
  - Thousands of movies and shows;
  - Each user, on average, watches < 100 of them.
- What can we say about the score matrix?
- Giant sparse matrix. At the intersection (user, column) is the score or any other useful information ("whether watched to the end" and other proxy metrics)

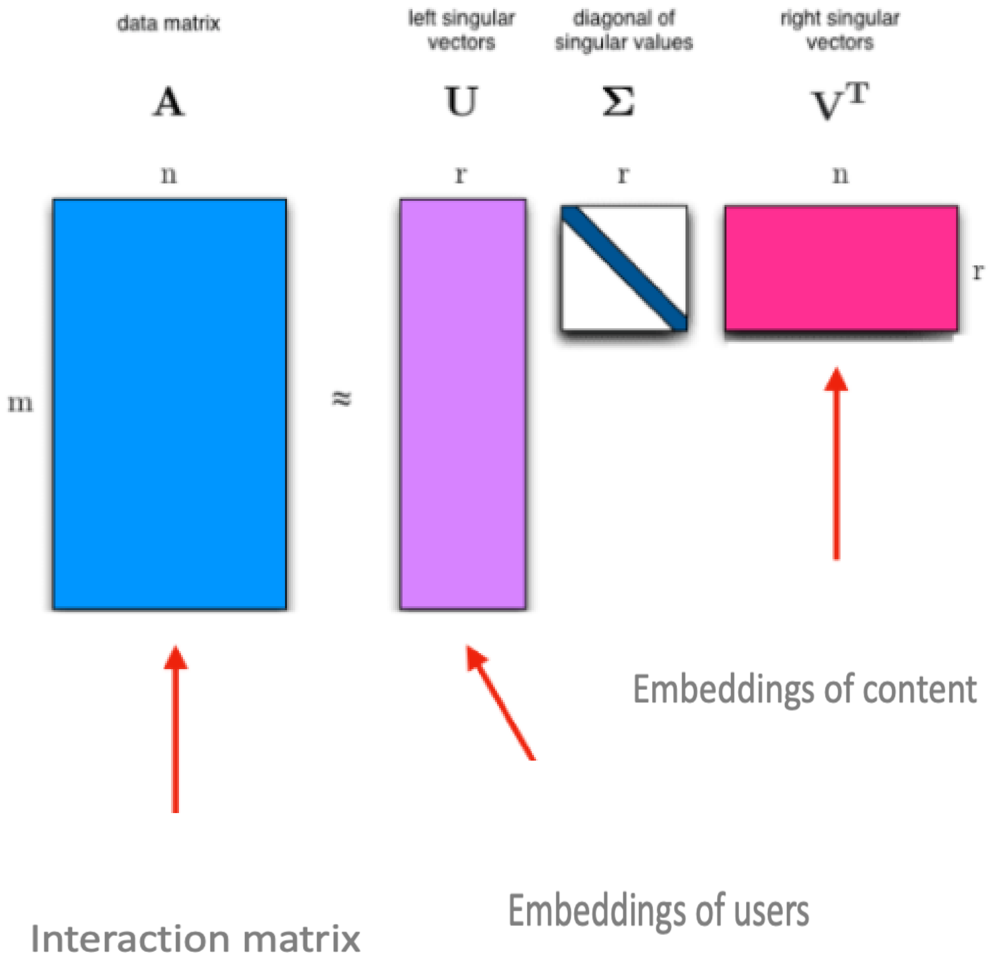
SVD Applications

- Truncated SVD decomposition of rank r for the interaction matrix allows us to obtain embeddings of dimension r for both users and content!



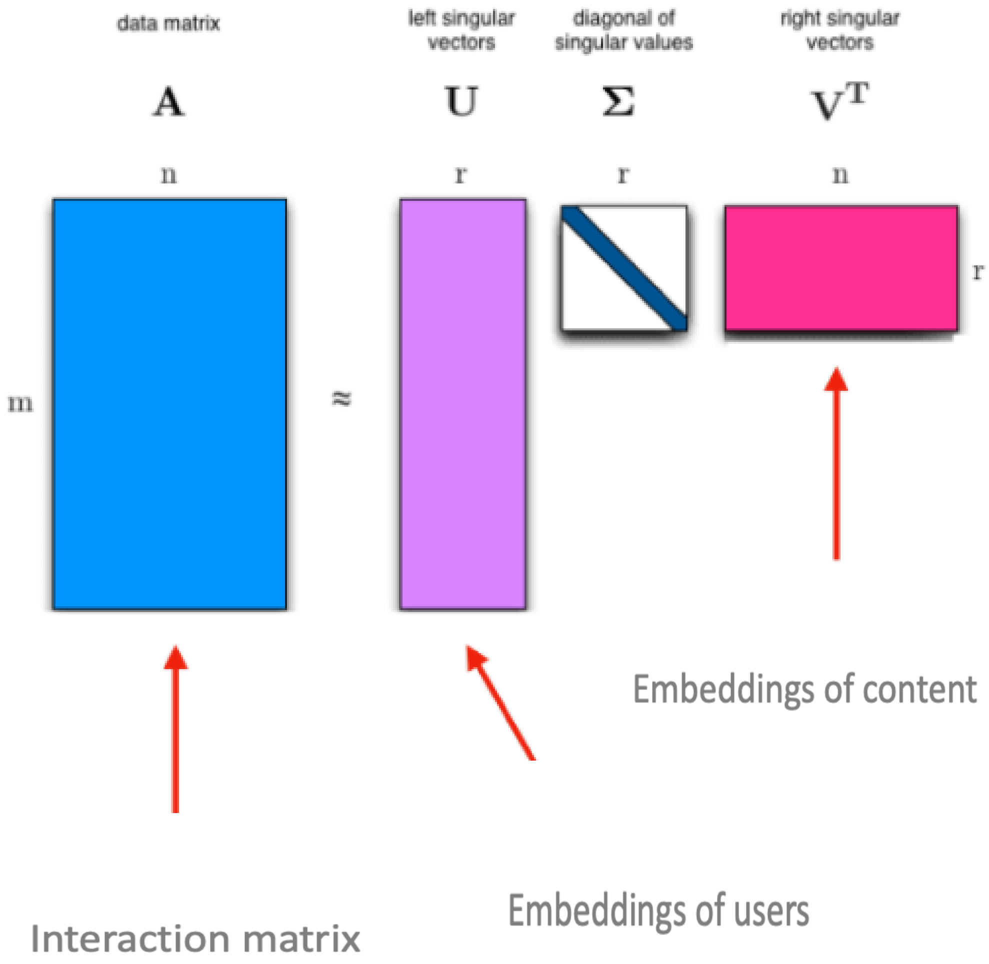
SVD Applications

- Truncated SVD decomposition of rank r for the interaction matrix allows us to obtain embeddings of dimension r for both users and content!
- Moreover, the higher the scalar product of user and content embeddings, the higher the chance that the user will like the content!



SVD Applications

- Truncated SVD decomposition of rank r for the interaction matrix allows us to obtain embeddings of dimension r for both users and content!
- Moreover, the higher the scalar product of user and content embeddings, the higher the chance that the user will like the content!
- This allows you to build the simplest recommendation system!



SVD Applications

- Collaborative filtering-based recommender systems are the foundation of recommendation systems at Yandex—in advertising, Yandex Zen, and so on.

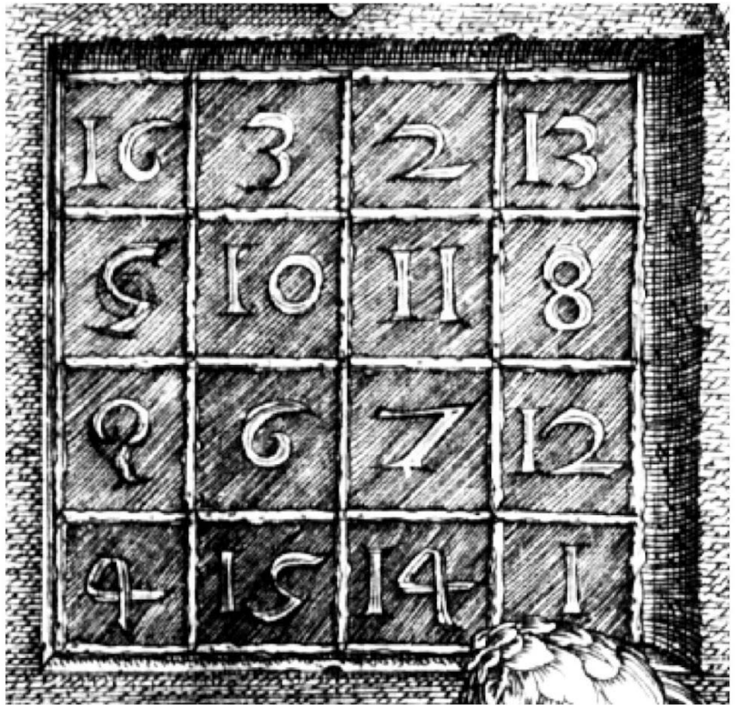
SVD Applications

- Collaborative filtering-based recommender systems are the foundation of recommendation systems at Yandex—in advertising, Yandex Zen, and so on.
- Of course, more advanced matrix factorizations are used there.
- But the idea remains the same!

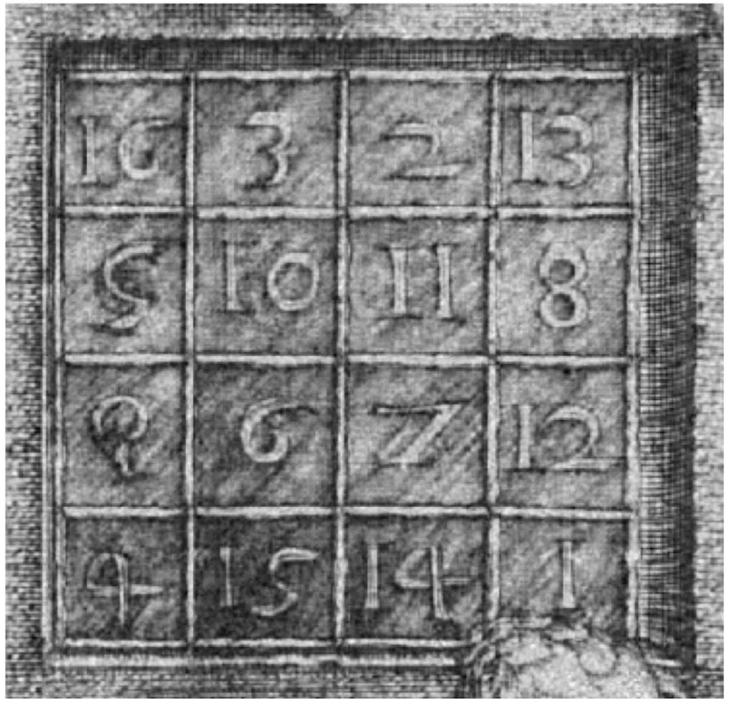
SVD Applications

- Furthermore, SVD decomposition can be used as a mechanism for compressing images and videos...

Detail from Durer's Melancolia, dated 1514., 359x371 image



EOF reconstruction with 50 modes

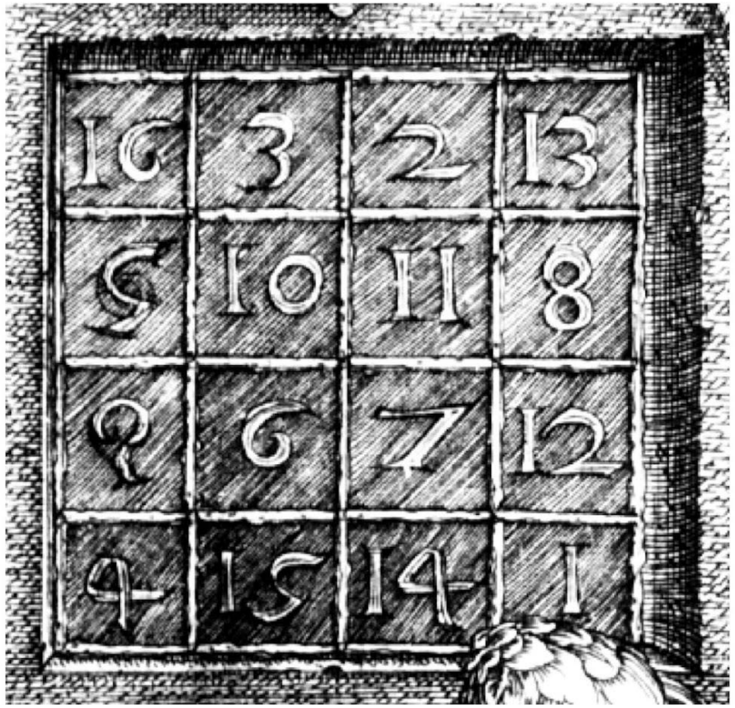


SVD Applications

Of course, we'll definitely come back to this in seminars!

- Furthermore, SVD decomposition can be used as a mechanism for compressing images and videos...

Detail from Durer's Melancolia, dated 1514., 359x371 image



EOF reconstruction with 50 modes