

Vector space

Definition (over reals)

A set \mathcal{X} is called a *vector space* over \mathbb{R} if addition and scalar multiplication are defined and satisfy for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{X}$ and $\lambda, \mu \in \mathbb{R}$:

- Addition:

associative $\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$

commutative $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$

identity element $\exists \mathbf{0} \in \mathcal{X} : \mathbf{x} + \mathbf{0} = \mathbf{x}$

inverse element $\forall \mathbf{x} \in \mathcal{X} \exists \mathbf{x}' \in \mathcal{X} : \mathbf{x} + \mathbf{x}' = \mathbf{0}$

- Scalar multiplication:

distributive over elements $\lambda(\mathbf{x} + \mathbf{y}) = \lambda\mathbf{x} + \lambda\mathbf{y}$

distributive over scalars $(\lambda + \mu)\mathbf{x} = \lambda\mathbf{x} + \mu\mathbf{x}$

associative over scalars $\lambda(\mu\mathbf{x}) = (\lambda\mu)\mathbf{x}$

identity element $\exists 1 \in \mathbb{R} : 1\mathbf{x} = \mathbf{x}$

Properties and operations in vector spaces

subspace Any non-empty subset of \mathcal{X} being itself a vector space (E.g. projection)

linear combination given $\lambda_i \in \mathbb{R}, \mathbf{x}_i \in \mathcal{X}$

$$\sum_{i=1}^n \lambda_i \mathbf{x}_i$$

span The span of vectors $\mathbf{x}_1, \dots, \mathbf{x}_n$ is defined as the set of their linear combinations

$$\left\{ \sum_{i=1}^n \lambda_i \mathbf{x}_i, \lambda_i \in \mathbb{R} \right\}$$

Basis in vector space

Linear independency

A set of vectors \mathbf{x}_i is *linearly independent* if none of them can be written as a linear combination of the others

Basis

- A set of vectors \mathbf{x}_i is a *basis* for \mathcal{X} if any element in \mathcal{X} can be *uniquely* written as a linear combination of vectors \mathbf{x}_i .
- Necessary condition is that vectors \mathbf{x}_i are linearly independent
- All bases of \mathcal{X} have the same number of elements, called the *dimension* of the vector space.

Linear maps

Definition

Given two vector spaces \mathcal{X}, \mathcal{Z} , a function $f : \mathcal{X} \rightarrow \mathcal{Z}$ is a *linear map* if for all $\mathbf{x}, \mathbf{y} \in \mathcal{X}, \lambda \in \mathbb{R}$:

- $f(\mathbf{x} + \mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y})$
- $f(\lambda\mathbf{x}) = \lambda f(\mathbf{x})$

Linear maps as matrices

A linear map between two finite-dimensional spaces \mathcal{X}, \mathcal{Z} of dimensions n, m can always be written as a matrix:

- Let $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ and $\{\mathbf{z}_1, \dots, \mathbf{z}_m\}$ be some bases for \mathcal{X} and \mathcal{Z} respectively.
- For any $\mathbf{x} \in \mathcal{X}$ we have:

$$\begin{aligned} f(\mathbf{x}) &= f\left(\sum_{i=1}^n \lambda_i \mathbf{x}_i\right) = \sum_{i=1}^n \lambda_i f(\mathbf{x}_i) \\ f(\mathbf{x}_i) &= \sum_{j=1}^m a_{ji} \mathbf{z}_j \\ f(\mathbf{x}) &= \sum_{i=1}^n \sum_{j=1}^m \lambda_i a_{ji} \mathbf{z}_j = \sum_{j=1}^m \left(\sum_{i=1}^n \lambda_i a_{ji}\right) \mathbf{z}_j = \sum_{j=1}^m \mu_j \mathbf{z}_j \end{aligned}$$

Linear maps as matrices

- Matrix of basis transformation

$$M \in \mathbb{R}^{m \times n} = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}$$

- Mapping from basis coefficients to basis coefficients

$$M\lambda = \mu$$

Change of Coordinate Matrix

2D example

- let $B = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$ be the standard basis in \mathbb{R}^2
- let $B' = \left\{ \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \end{bmatrix} \right\}$ be an alternative basis
- The change of coordinate matrix from B' to B is:

$$P = \begin{bmatrix} 3 & -2 \\ 1 & 1 \end{bmatrix}$$

- So that:

$$[\mathbf{v}]_B = P \cdot [\mathbf{v}]_{B'} \quad \text{and} \quad [\mathbf{v}]_{B'} = P^{-1} \cdot [\mathbf{v}]_B$$

Note

- For arbitrary B and B' , P 's columns must be the B' vectors written in terms of the B ones (straightforward here)

Matrix properties

transpose Matrix obtained exchanging rows with columns (indicated with M^T). Properties:

$$(MN)^T = N^T M^T$$

trace Sum of diagonal elements of a matrix

$$\text{tr}(M) = \sum_{i=1}^n M_{ii}$$

inverse The matrix which multiplied with the original matrix gives the identity

$$MM^{-1} = I$$

rank The rank of an $n \times m$ matrix is the dimension of the space spanned by its columns

Matrix derivatives

$$\begin{aligned} \frac{\partial M\mathbf{x}}{\partial \mathbf{x}} &= M \\ \frac{\partial \mathbf{y}^T M\mathbf{x}}{\partial \mathbf{x}} &= M^T \mathbf{y} \\ \frac{\partial \mathbf{x}^T M\mathbf{x}}{\partial \mathbf{x}} &= (M^T + M)\mathbf{x} \\ \frac{\partial \mathbf{x}^T M\mathbf{x}}{\partial \mathbf{x}} &= 2M\mathbf{x} \quad \text{if } M \text{ is symmetric} \\ \frac{\partial \mathbf{x}^T \mathbf{x}}{\partial \mathbf{x}} &= 2\mathbf{x} \end{aligned}$$

Note

Results are column vectors. Transpose them if row vectors are needed instead.

Metric structure

Norm

A function $\|\cdot\| : \mathcal{X} \rightarrow \mathbb{R}_0^+$ is a *norm* if for all $\mathbf{x}, \mathbf{y} \in \mathcal{X}, \lambda \in \mathbb{R}$:

- $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$
- $\|\lambda\mathbf{x}\| = |\lambda| \|\mathbf{x}\|$
- $\|\mathbf{x}\| > 0$ if $\mathbf{x} \neq 0$

Metric

A norm defines a metric $d : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}_0^+$:

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$$

Note

The concept of norm is stronger than that of metric: not any metric gives rise to a norm

Dot product

Bilinear form

A function $Q : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ is a *bilinear form* if for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{X}, \lambda, \mu \in \mathbb{R}$:

- $Q(\lambda\mathbf{x} + \mu\mathbf{y}, \mathbf{z}) = \lambda Q(\mathbf{x}, \mathbf{z}) + \mu Q(\mathbf{y}, \mathbf{z})$
- $Q(\mathbf{x}, \lambda\mathbf{y} + \mu\mathbf{z}) = \lambda Q(\mathbf{x}, \mathbf{y}) + \mu Q(\mathbf{x}, \mathbf{z})$

A bilinear form is *symmetric* if for all $\mathbf{x}, \mathbf{y} \in \mathcal{X}$:

- $Q(\mathbf{x}, \mathbf{y}) = Q(\mathbf{y}, \mathbf{x})$

Dot product

Dot product

A dot product $\langle \cdot, \cdot \rangle : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ is a symmetric bilinear form which is *positive semi-definite*:

$$\langle \mathbf{x}, \mathbf{x} \rangle \geq 0 \quad \forall \mathbf{x} \in \mathcal{X}$$

A *positive definite* dot product satisfies

$$\langle \mathbf{x}, \mathbf{x} \rangle = 0 \quad \text{iff} \quad \mathbf{x} = \mathbf{0}$$

Norm

Any dot product defines a corresponding norm via:

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$$

Properties of dot product

angle The angle θ between two vectors is defined as:

$$\cos\theta = \frac{\langle \mathbf{x}, \mathbf{z} \rangle}{\|\mathbf{x}\| \|\mathbf{z}\|}$$

orthogonal Two vectors are *orthogonal* if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$

orthonormal A set of vectors $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ is *orthonormal* if

$$\langle \mathbf{x}_i, \mathbf{x}_j \rangle = \delta_{ij}$$

where $\delta_{ij} = 1$ if $i = j$, 0 otherwise.

Note

If \mathbf{x} and \mathbf{y} are n -dimensional column vectors, their dot product is computed as:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = \sum_{i=1}^n x_i y_i$$

Eigenvalues and eigenvectors

Definition

Given an $n \times n$ matrix M , the real value λ and (non-zero) vector \mathbf{x} are an *eigenvalue* and corresponding *eigenvector* of M if

$$M\mathbf{x} = \lambda\mathbf{x}$$

Cardinality

- An $n \times n$ matrix has n eigenvalues (roots of characteristic polynomial)
- An $n \times n$ matrix can have **less than n distinct** eigenvalues
- An $n \times n$ matrix can have **less than n linear independent** eigenvectors (also fewer than the number of distinct eigenvalues)

Eigenvalues and eigenvectors

Singular matrices

- A matrix is *singular* if it has a zero eigenvalue

$$M\mathbf{x} = 0\mathbf{x} = \mathbf{0}$$

- A singular matrix has linearly dependent columns:

Eigenvalues and eigenvectors

Symmetric matrices

Eigenvectors corresponding to distinct eigenvalues are orthogonal:

$$\begin{aligned}\lambda\langle\mathbf{x}, \mathbf{z}\rangle &= \langle A\mathbf{x}, \mathbf{z}\rangle \\ &= (A\mathbf{x})^T \mathbf{z} \\ &= \mathbf{x}^T A^T \mathbf{z} \\ &= \mathbf{x}^T A\mathbf{z} \\ &= \langle\mathbf{x}, A\mathbf{z}\rangle \\ &= \mu\langle\mathbf{x}, \mathbf{z}\rangle\end{aligned}$$

Eigen-decomposition

Raleigh quotient

$$\begin{aligned}A\mathbf{x} &= \lambda\mathbf{x} \\ \frac{\mathbf{x}^T A\mathbf{x}}{\mathbf{x}^T \mathbf{x}} &= \lambda \frac{\mathbf{x}^T \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \lambda\end{aligned}$$

Finding eigenvector

1. Maximize eigenvalue:

$$\mathbf{x} = \max_{\mathbf{v}} \frac{\mathbf{v}^T A \mathbf{v}}{\mathbf{v}^T \mathbf{v}}$$

2. Normalize eigenvector (solution is invariant to rescaling)

$$\mathbf{x} \leftarrow \frac{\mathbf{x}}{\|\mathbf{x}\|}$$

Eigen-decomposition

Deflating matrix

$$\tilde{A} = A - \lambda \mathbf{x} \mathbf{x}^T$$

- Deflation turns \mathbf{x} into a zero-eigenvalue eigenvector:

$$\begin{aligned} \tilde{A} \mathbf{x} &= A \mathbf{x} - \lambda \mathbf{x} \mathbf{x}^T \mathbf{x} \quad (\mathbf{x} \text{ is normalized}) \\ &= A \mathbf{x} - \lambda \mathbf{x} = 0 \end{aligned}$$

- Other eigenvalues are unchanged as eigenvectors with distinct eigenvalues are orthogonal (symmetric matrix):

$$\begin{aligned} \tilde{A} \mathbf{z} &= A \mathbf{z} - \lambda \mathbf{x} \mathbf{x}^T \mathbf{z} \quad (\mathbf{x} \text{ and } \mathbf{z} \text{ orthonormal}) \\ \tilde{A} \mathbf{z} &= A \mathbf{z} \end{aligned}$$

Eigen-decomposition

Iterating

- The maximization procedure is repeated on the deflated matrix (until solution is zero)
- Minimization is iterated to get eigenvectors with negative eigenvalues
- Eigenvectors with zero eigenvalues are obtained extending the obtained set to an orthonormal basis

Eigen-decomposition

Eigen-decomposition

- Let $V = [\mathbf{v}_1 \dots \mathbf{v}_n]$ be a matrix with orthonormal eigenvectors as columns
- Let Λ be the diagonal matrix of corresponding eigenvalues
- A square symmetric matrix can be *diagonalized* as:

$$V^T A V = \Lambda$$

proof follows..

Note

- A diagonalized matrix is much simpler to manage and has the same properties as the original one (e.g. same eigen-decomposition)
- E.g. change of coordinate system

Eigen-decomposition

Proof

$$A[\mathbf{v}_1 \dots \mathbf{v}_n] = [\mathbf{v}_1 \dots \mathbf{v}_n] \begin{bmatrix} \lambda_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \lambda_n \end{bmatrix}$$
$$AV = V\Lambda$$
$$V^{-1}AV = V^{-1}V\Lambda$$
$$V^T AV = \Lambda$$

Note

V is a *unitary* matrix (orthonormal columns), for which:

$$V^{-1} = V^T$$

Positive semi-definite matrix

Definition

An $n \times n$ symmetric matrix M is *positive semi-definite* if all its eigenvalues are non-negative.

Alternative sufficient and necessary conditions

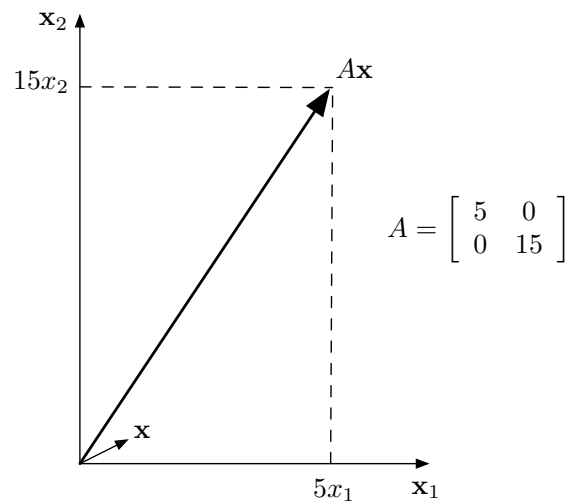
- for all $\mathbf{x} \in \mathbb{R}^n$

$$\mathbf{x}^T M \mathbf{x} \geq 0$$

- there exists a real matrix B s.t.

$$M = B^T B$$

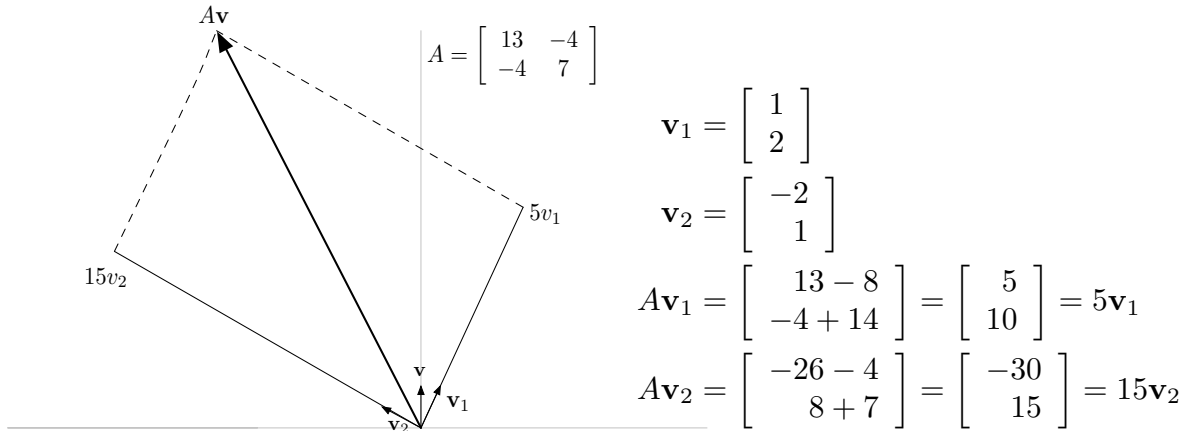
Understanding eigendecomposition



Scaling transformation in standard basis

- let $\mathbf{x}_1 = [1, 0], \mathbf{x}_2 = [0, 1]$ be the standard orthonormal basis in \mathbb{R}^2
- let $\mathbf{x} = [x_1, x_2]$ be an arbitrary vector in \mathbb{R}^2
- A linear transformation is a *scaling* transformation if it only stretches \mathbf{x} along its directions

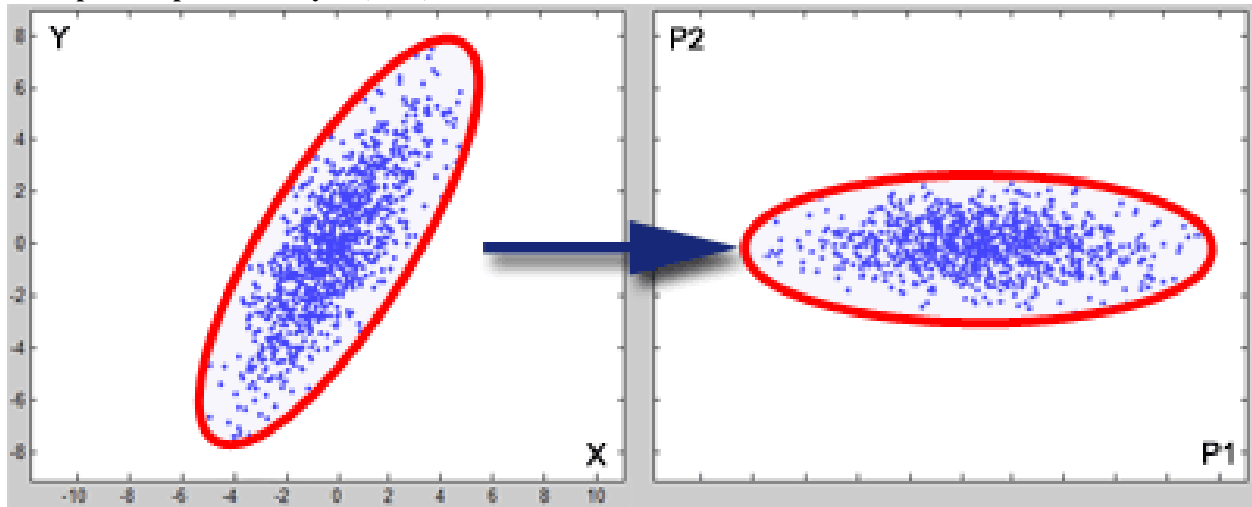
Understanding eigendecomposition



Scaling transformation in eigenbasis

- let A be a non-scaling linear transformation in \mathbb{R}^2 .
- let $\{\mathbf{v}_1, \mathbf{v}_2\}$ be an *eigenbasis* for A .
- By representing vectors in \mathbb{R}^2 in terms of the $\{\mathbf{v}_1, \mathbf{v}_2\}$ basis (instead of the standard $\{\mathbf{x}_1, \mathbf{x}_2\}$), A becomes a *scaling* transformation

Principal Component Analysis (PCA)



Description

- Let X be a data matrix with correlated coordinates.

- PCA is a linear transformation mapping data to a system of *uncorrelated* coordinates.
- It corresponds to fitting an *ellipsoid* to the data, whose axes are the coordinates of the new space.

Principal Component Analysis (PCA)

Procedure (1)

Given a dataset $X \in \mathbb{R}^{n \times d}$ in d dimensions.

- 1 Compute the mean of the data (X_i is i^{th} row vector of X):

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n X_i$$

- 2 Center the data into the origin:

$$X - \begin{bmatrix} \bar{\mathbf{x}} \\ \vdots \\ \bar{\mathbf{x}} \end{bmatrix}$$

- 3 Compute the data covariance: $C = \frac{1}{n} X^T X$

Principal Component Analysis (PCA)

Procedure (2)

- 4 Compute the (orthonormal) eigendecomposition of C :

$$V^T C V = \Lambda$$

- 5 Use it as the new coordinate system:

$$\mathbf{x}' = V^{-1} \mathbf{x} = V^T \mathbf{x}$$

($V^{-1} = V^T$ as V is unitary)

Warning

- It assumes linear correlations (and Gaussian distributions)

Principal Component Analysis (PCA)

Dimensionality reduction

- Each eigenvalue corresponds to the amount of variance in that direction
- Select only the k eigenvectors with largest eigenvalues for dimensionality reduction (e.g. visualization)

Procedure

- 1 $W = [\mathbf{v}_1, \dots, \mathbf{v}_k]$
- 2 $\mathbf{x}' = W^T \mathbf{x}$