Machine Learning with Python

MTH786U/P 2023/24

Lecture 7: Review for mid-term

Nicola Perra, Queen Mary University of London (QMUL)

n.perra@qmul.ac.uk

LINEAR ALGEBRA

Matrices and Vectors

Matrix: $$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1d} \\ x_{21} & x_{22} & \dots & x_{2d} \\ \vdots & & \ddots & \vdots \\ x_{s1} & x_{s2} & \dots & x_{sd} \end{pmatrix} \in \mathbb{R}^{s \times d}$$

Matrices and Vectors

Matrix: $\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1d} \\ x_{21} & x_{22} & \dots & x_{2d} \\ \vdots & & \ddots & \vdots \\ x_{s1} & x_{s2} & \dots & x_{sd} \end{pmatrix} \in \mathbb{R}^{s \times d}$

Vectors: $\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{pmatrix} \in \mathbb{R}^{d \times 1}, \ \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_s \end{pmatrix} \in \mathbb{R}^{s \times 1}$

Matrices and Vectors

Matrix: $\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1d} \\ x_{21} & x_{22} & \dots & x_{2d} \\ \vdots & & \ddots & \vdots \\ x_{s1} & x_{s2} & \dots & x_{sd} \end{pmatrix} \in \mathbb{R}^{s \times d}$

Vectors: $\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{pmatrix} \in \mathbb{R}^{d \times 1}, \ \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_s \end{pmatrix} \in \mathbb{R}^{s \times 1}$

Matrix-vector multiplication: $\mathbf{X}\,\mathbf{w} = \mathbf{y}$

Matrices and Vectors

Matrix: $\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1d} \\ x_{21} & x_{22} & \dots & x_{2d} \\ \vdots & & \ddots & \vdots \\ x_{s1} & x_{s2} & \dots & x_{sd} \end{pmatrix} \in \mathbb{R}^{s \times d}$

Vectors: $\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{pmatrix} \in \mathbb{R}^{d \times 1}, \ \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_s \end{pmatrix} \in \mathbb{R}^{s \times 1}$

Matrix-vector multiplication: $\mathbf{X}\,\mathbf{w} = \mathbf{y}$

You need to pay attention to the dimensions!

Matrices and Vectors

Matrix: $\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1d} \\ x_{21} & x_{22} & \dots & x_{2d} \\ \vdots & & \ddots & \vdots \\ x_{s1} & x_{s2} & \dots & x_{sd} \end{pmatrix} \in \mathbb{R}^{s \times d}$

Vectors: $\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{pmatrix} \in \mathbb{R}^{d \times 1}$, $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_s \end{pmatrix} \in \mathbb{R}^{s \times 1}$

Matrix-vector multiplication: $\mathbf{X}\,\mathbf{w} = \mathbf{y}$

You need to pay attention to the dimensions! $s \times d \;\; d \times 1 \rightarrow s \times 1$

Matrices and Vectors

Matrix: $\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1d} \\ x_{21} & x_{22} & \dots & x_{2d} \\ \vdots & & \ddots & \vdots \\ x_{s1} & x_{s2} & \dots & x_{sd} \end{pmatrix} \in \mathbb{R}^{s\times d}$

Vectors: $\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{pmatrix} \in \mathbb{R}^{d\times 1}$, $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_s \end{pmatrix} \in \mathbb{R}^{s\times 1}$

Matrix-vector multiplication: $\mathbf{X}\,\mathbf{w} = \mathbf{y}$

You need to pay attention to the dimensions! $s \times d \;\; d \times 1 \rightarrow s \times 1$

n.perra@qmul.ac.uk

Matrices and Vectors

The row times column rule can be written as

$$\mathbf{X}\,\mathbf{w} = \mathbf{y} \qquad \Leftrightarrow \qquad \sum_{j=1}^{d} x_{ij} w_j = y_i \quad \forall i \in \{1,\ldots,s\}$$

Matrices and Vectors

The row times column rule can be written as

$$\mathbf{X}\,\mathbf{w} = \mathbf{y} \qquad \Leftrightarrow \qquad \sum_{j=1}^{d} x_{ij} w_j = y_i \quad \forall i \in \{1, \ldots, s\}$$

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

Matrices and Vectors

The row times column rule can be written as

$$\mathbf{X}\,\mathbf{w} = \mathbf{y} \qquad \Leftrightarrow \qquad \sum_{j=1}^{d} x_{ij} w_j = y_i \quad \forall i \in \{1, \ldots, s\}$$

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

Matrices and Vectors

The row times column rule can be written as

$$\mathbf{X}\,\mathbf{w} = \mathbf{y} \qquad \Leftrightarrow \qquad \sum_{j=1}^{d} x_{ij} w_j = y_i \quad \forall i \in \{1, \ldots, s\}$$

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

Matrices and Vectors

The row times column rule can be written as

$$\mathbf{X}\,\mathbf{w} = \mathbf{y} \qquad \Leftrightarrow \qquad \sum_{j=1}^{d} x_{ij} w_j = y_i \quad \forall i \in \{1, \ldots, s\}$$

$$\sum_{j=1}^{3} x_{1j} w_j = 1 \times 1 + 2 \times 4 + 3 \times 0$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 4 \\ 0 \end{pmatrix} \rightarrow$$

Matrices and Vectors

The row times column rule can be written as

$$\mathbf{X}\,\mathbf{w} = \mathbf{y} \qquad \Leftrightarrow \qquad \sum_{j=1}^{d} x_{ij} w_j = y_i \quad \forall i \in \{1,\ldots,s\}$$

$$\sum_{j=1}^{3} x_{1j} w_j = 1 \times 1 + 2 \times 4 + 3 \times 0$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 4 \\ 0 \end{pmatrix} \rightarrow$$

Matrices and Vectors

The row times column rule can be written as

$$\mathbf{X}\,\mathbf{w} = \mathbf{y} \qquad \Leftrightarrow \qquad \sum_{j=1}^{d} x_{ij} w_j = y_i \quad \forall i \in \{1,\ldots,s\}$$

$$\sum_{j=1}^{3} x_{1j} w_j = 1 \times 1 + 2 \times 4 + 3 \times 0$$

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

Matrices and Vectors

The row times column rule can be written as

$$\mathbf{X}\,\mathbf{w} = \mathbf{y} \qquad \Leftrightarrow \qquad \sum_{j=1}^{d} x_{ij} w_j = y_i \quad \forall i \in \{1, \ldots, s\}$$

$$\sum_{j=1}^{3} x_{1j} w_j = 1 \times 1 + 2 \times 4 + 3 \times 0$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 4 \\ 0 \end{pmatrix} \rightarrow \sum_{j=1}^{3} x_{2j} w_j = 0 \times 1 + 1 \times 4 + 2 \times 0$$

Matrices and Vectors

The row times column rule can be written as

$$\mathbf{X}\,\mathbf{w} = \mathbf{y} \qquad \Leftrightarrow \qquad \sum_{j=1}^{d} x_{ij} w_j = y_i \quad \forall i \in \{1,\ldots,s\}$$

$$\sum_{j=1}^{3} x_{1j} w_j = 1 \times 1 + 2 \times 4 + 3 \times 0$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 4 \\ 0 \end{pmatrix} \rightarrow \sum_{j=1}^{3} x_{2j} w_j = 0 \times 1 + 1 \times 4 + 2 \times 0$$

Matrices and Vectors

The row times column rule can be written as

$$\mathbf{X}\,\mathbf{w} = \mathbf{y} \qquad \Leftrightarrow \qquad \sum_{j=1}^{d} x_{ij} w_j = y_i \quad \forall i \in \{1,\ldots,s\}$$

$$\sum_{j=1}^{3} x_{1j} w_j = 1 \times 1 + 2 \times 4 + 3 \times 0$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 4 \\ 0 \end{pmatrix} \rightarrow \sum_{j=1}^{3} x_{2j} w_j = 0 \times 1 + 1 \times 4 + 2 \times 0$$

Matrices and Vectors

The row times column rule can be written as

$$\mathbf{X}\,\mathbf{w} = \mathbf{y} \qquad \Leftrightarrow \qquad \sum_{j=1}^{d} x_{ij} w_j = y_i \quad \forall i \in \{1,\ldots,s\}$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 4 \\ 0 \end{pmatrix} \rightarrow \begin{aligned} \sum_{j=1}^{3} x_{1j} w_j &= 1 \times 1 + 2 \times 4 + 3 \times 0 \\ \sum_{j=1}^{3} x_{2j} w_j &= 0 \times 1 + 1 \times 4 + 2 \times 0 \\ \sum_{j=1}^{3} x_{3j} w_j &= 1 \times 1 - 1 \times 4 + 1 \times 0 \end{aligned}$$

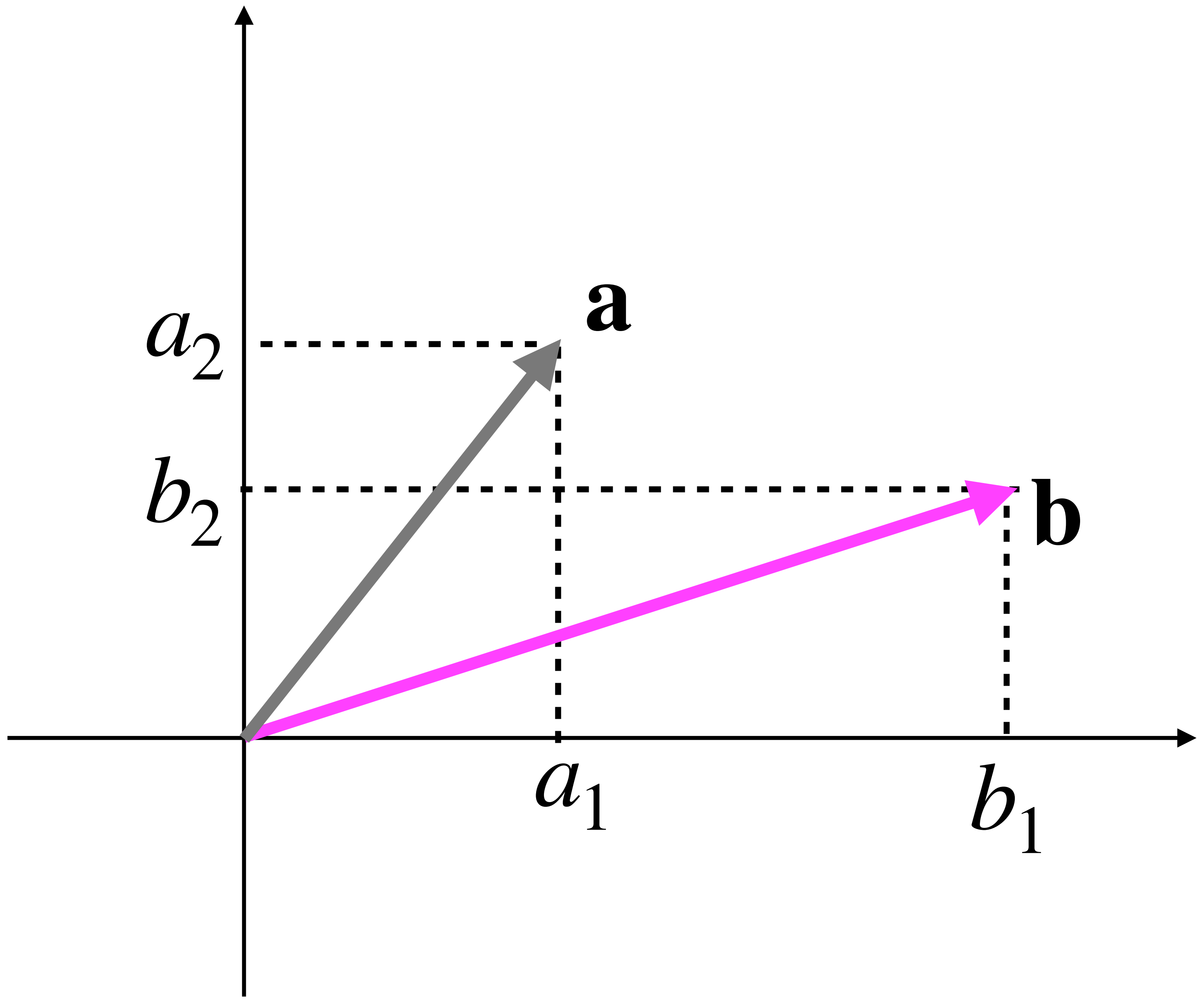
Matrices and Vectors

The row times column rule can be written as

$$\mathbf{X}\,\mathbf{w} = \mathbf{y} \qquad \Leftrightarrow \qquad \sum_{j=1}^{d} x_{ij} w_j = y_i \quad \forall i \in \{1, \ldots, s\}$$

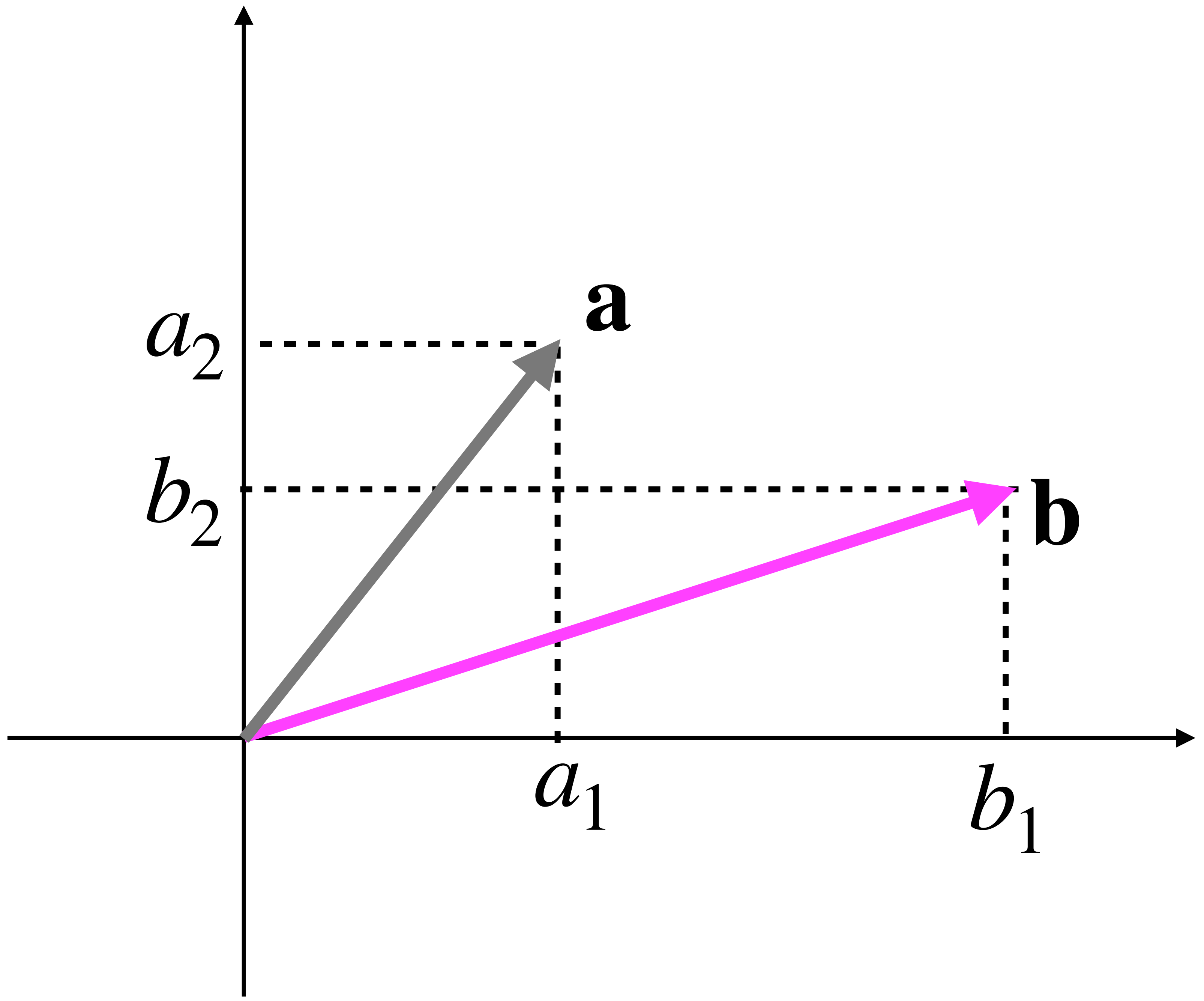
$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 4 \\ 0 \end{pmatrix} \rightarrow \begin{matrix} \sum_{j=1}^{3} x_{1j} w_j = 1 \times 1 + 2 \times 4 + 3 \times 0 \\ \sum_{j=1}^{3} x_{2j} w_j = 0 \times 1 + 1 \times 4 + 2 \times 0 \\ \sum_{j=1}^{3} x_{3j} w_j = 1 \times 1 - 1 \times 4 + 1 \times 0 \end{matrix} \rightarrow \sum_{j=1}^{3} x_{ij} w_j$$

Inner/dot/scalar product



$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad \mathbf{a}, \mathbf{b} \in \mathbb{R}^{2\times 1}$$

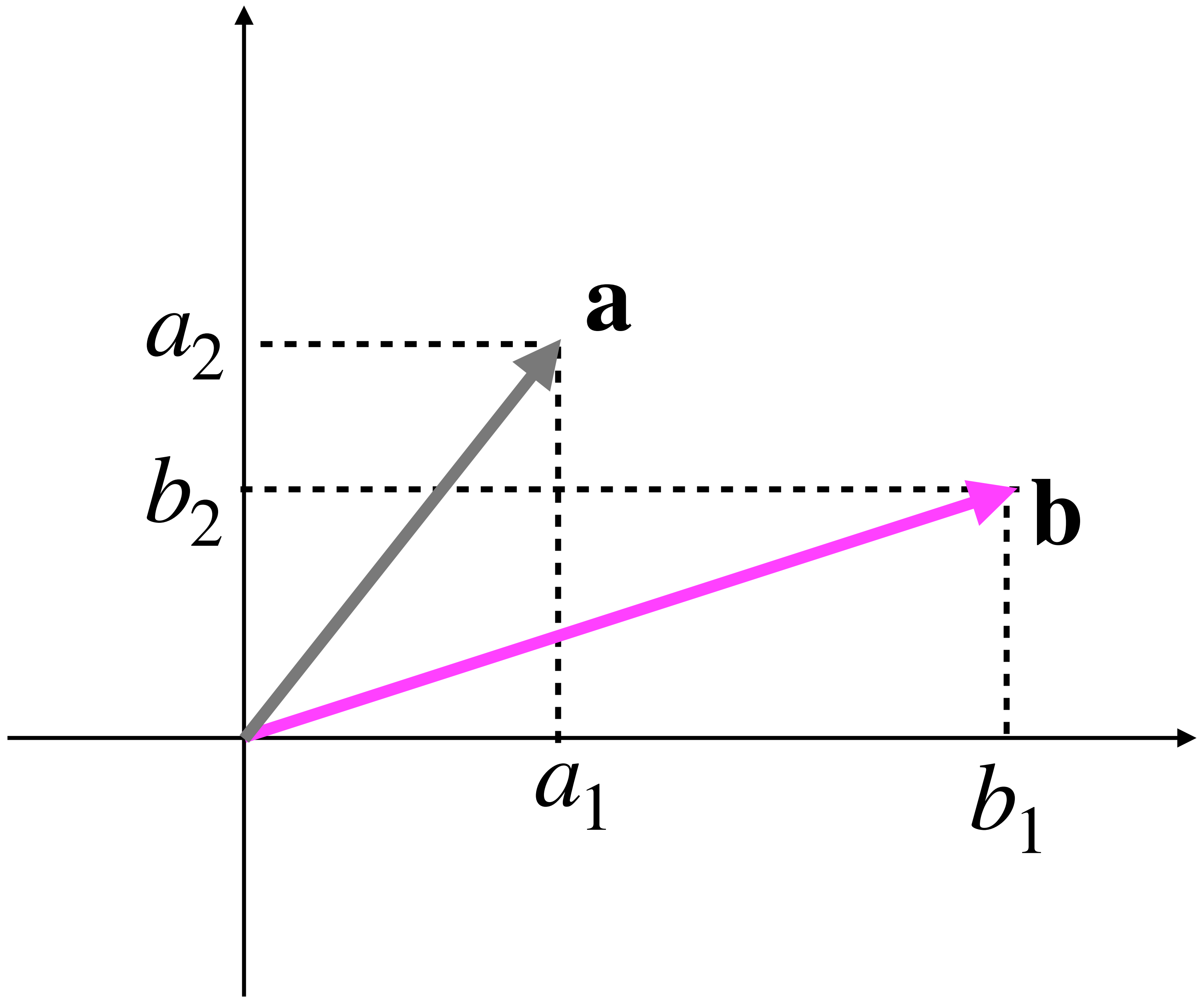
Inner/dot/scalar product



$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad \mathbf{a}, \mathbf{b} \in \mathbb{R}^{2\times 1}$$

$$\mathbf{a} \cdot \mathbf{b} = \sum_{j=1}^{d} a_j b_j = a_1 b_1 + a_2 b_2$$

Inner/dot/scalar product

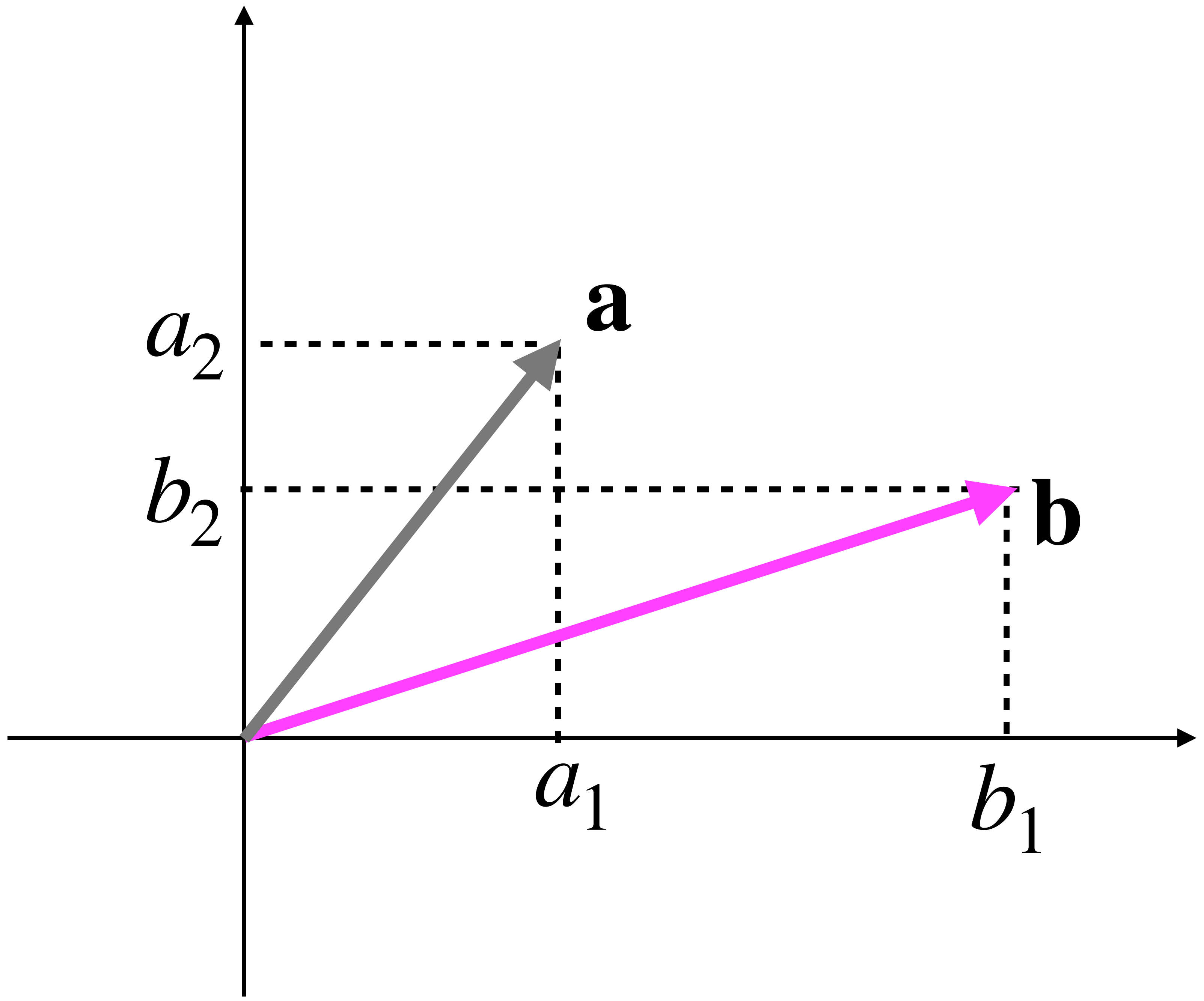


$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad \mathbf{a}, \mathbf{b} \in \mathbb{R}^{2\times1}$$

$$\mathbf{a} \cdot \mathbf{b} = \sum_{j=1}^{d} a_j b_j = a_1 b_1 + a_2 b_2$$

$$\mathbf{b} \cdot \mathbf{a} = \sum_{j=1}^{d} b_j a_j = b_1 a_1 + b_2 a_2$$

Inner/dot/scalar product



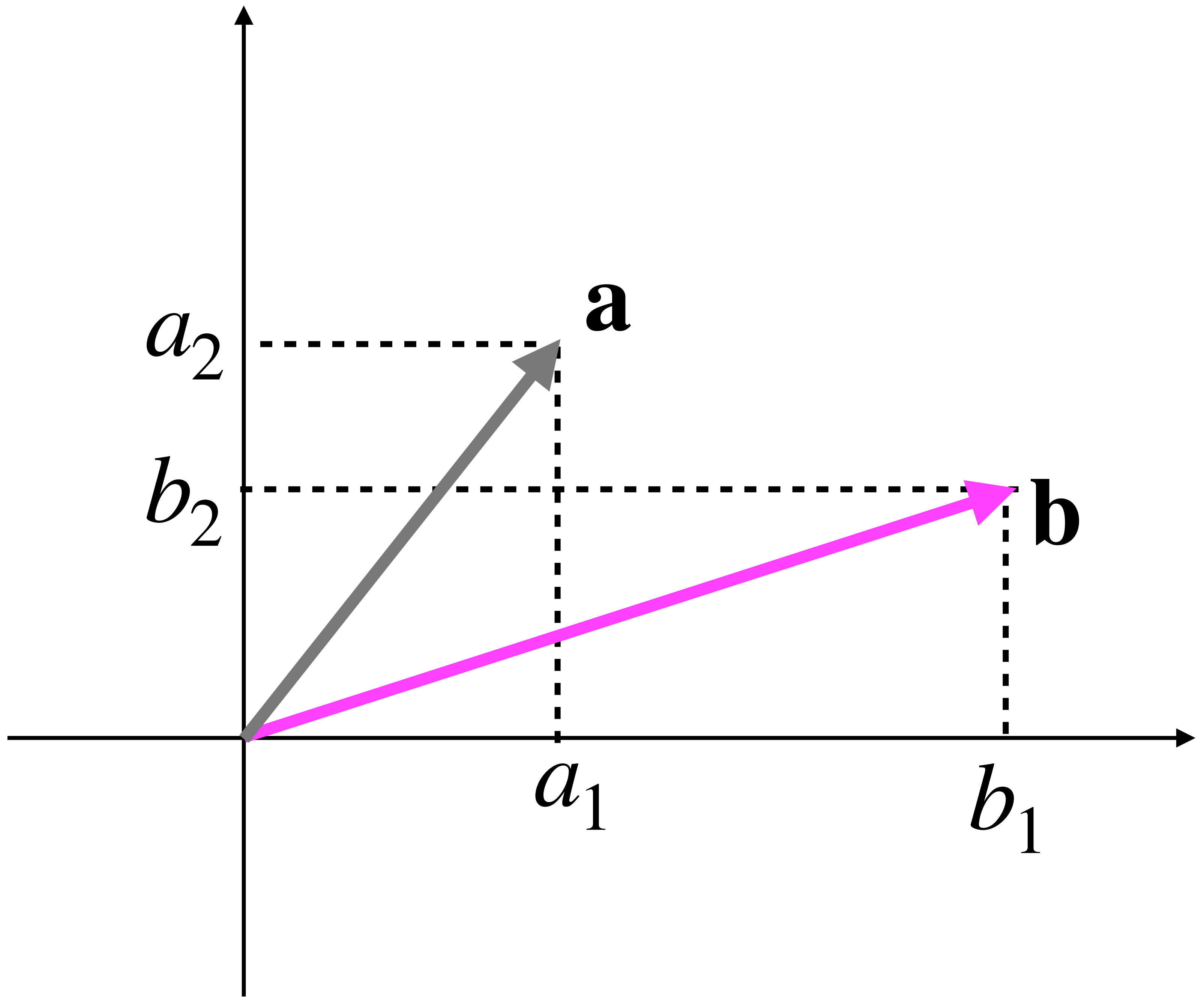
$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad \mathbf{a}, \mathbf{b} \in \mathbb{R}^{2\times 1}$$

$$\mathbf{a} \cdot \mathbf{b} = \sum_{j=1}^{d} a_j b_j = a_1 b_1 + a_2 b_2$$

$$\mathbf{b} \cdot \mathbf{a} = \sum_{j=1}^{d} b_j a_j = b_1 a_1 + b_2 a_2$$

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$$

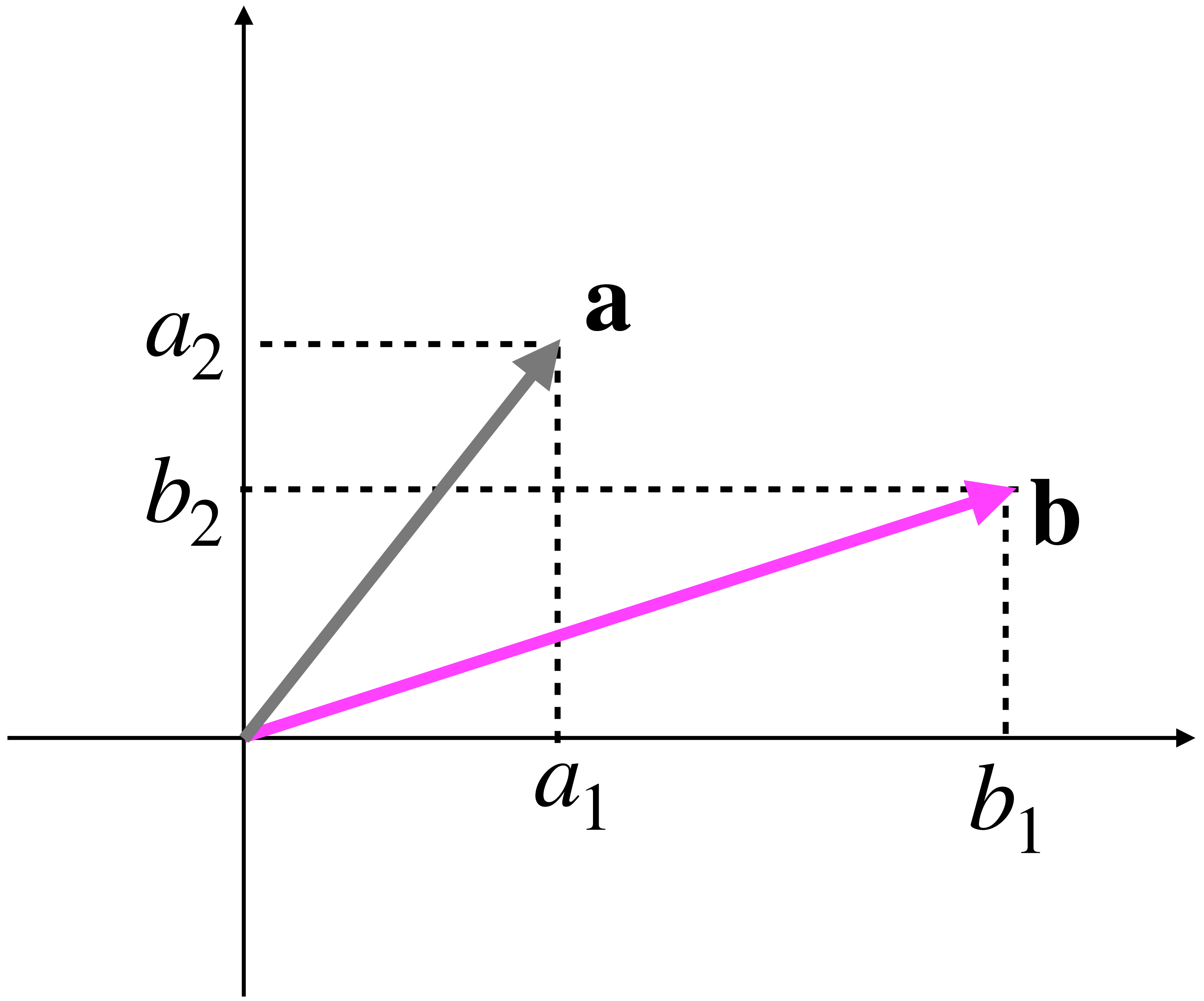
Inner/dot/scalar product



$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad \mathbf{a}, \mathbf{b} \in \mathbb{R}^{2\times 1}$$

$$\mathbf{a} \cdot \mathbf{b} = \sum_{j=1}^{d} a_j b_j = a_1 b_1 + a_2 b_2$$

Inner/dot/scalar product

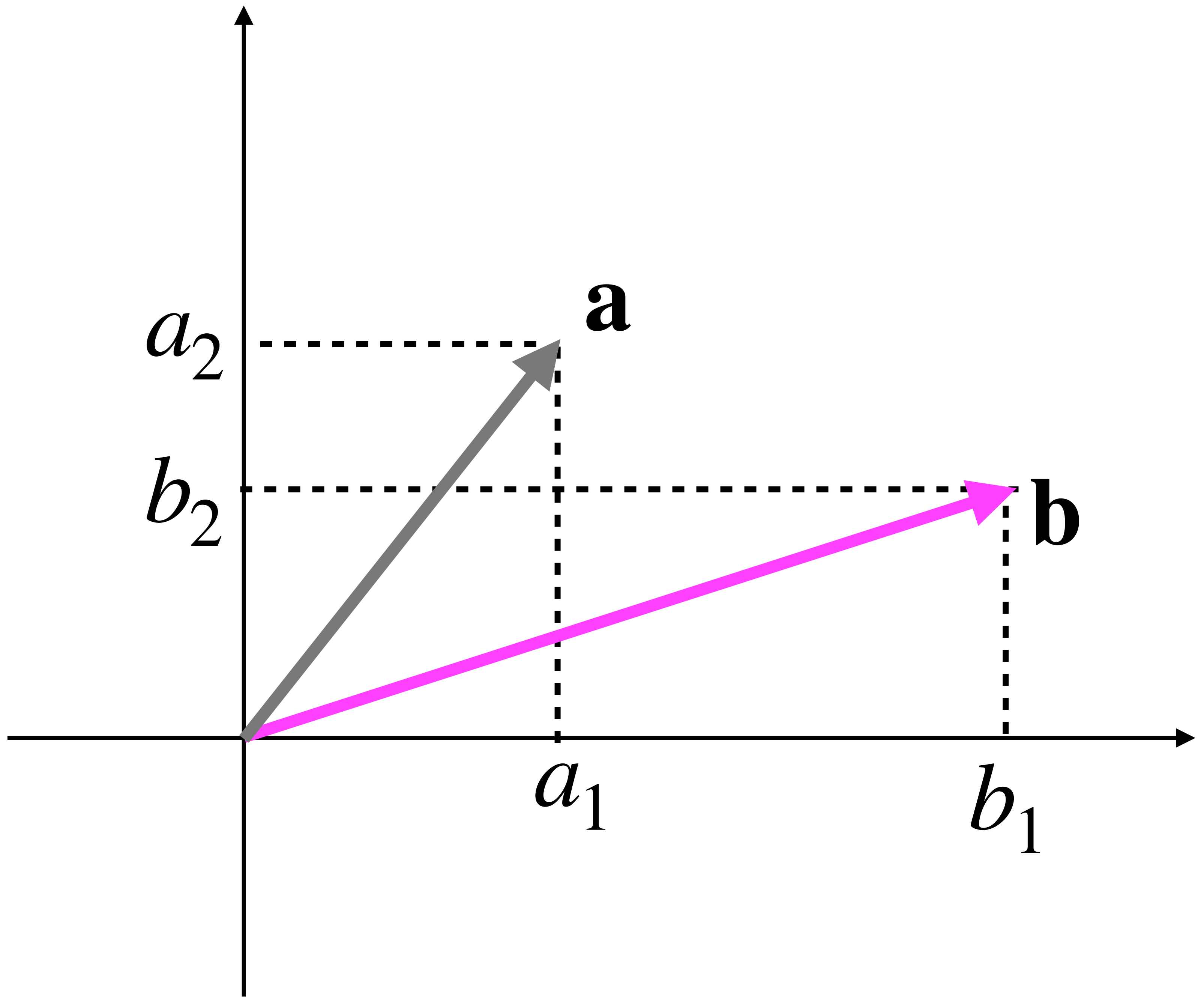


$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad \mathbf{a}, \mathbf{b} \in \mathbb{R}^{2\times 1}$$

$$\mathbf{a} \cdot \mathbf{b} = \sum_{j=1}^{d} a_j b_j = a_1 b_1 + a_2 b_2$$

$$\mathbf{a} \cdot \mathbf{b} = \langle \mathbf{a}, \mathbf{b} \rangle$$

Inner/dot/scalar product



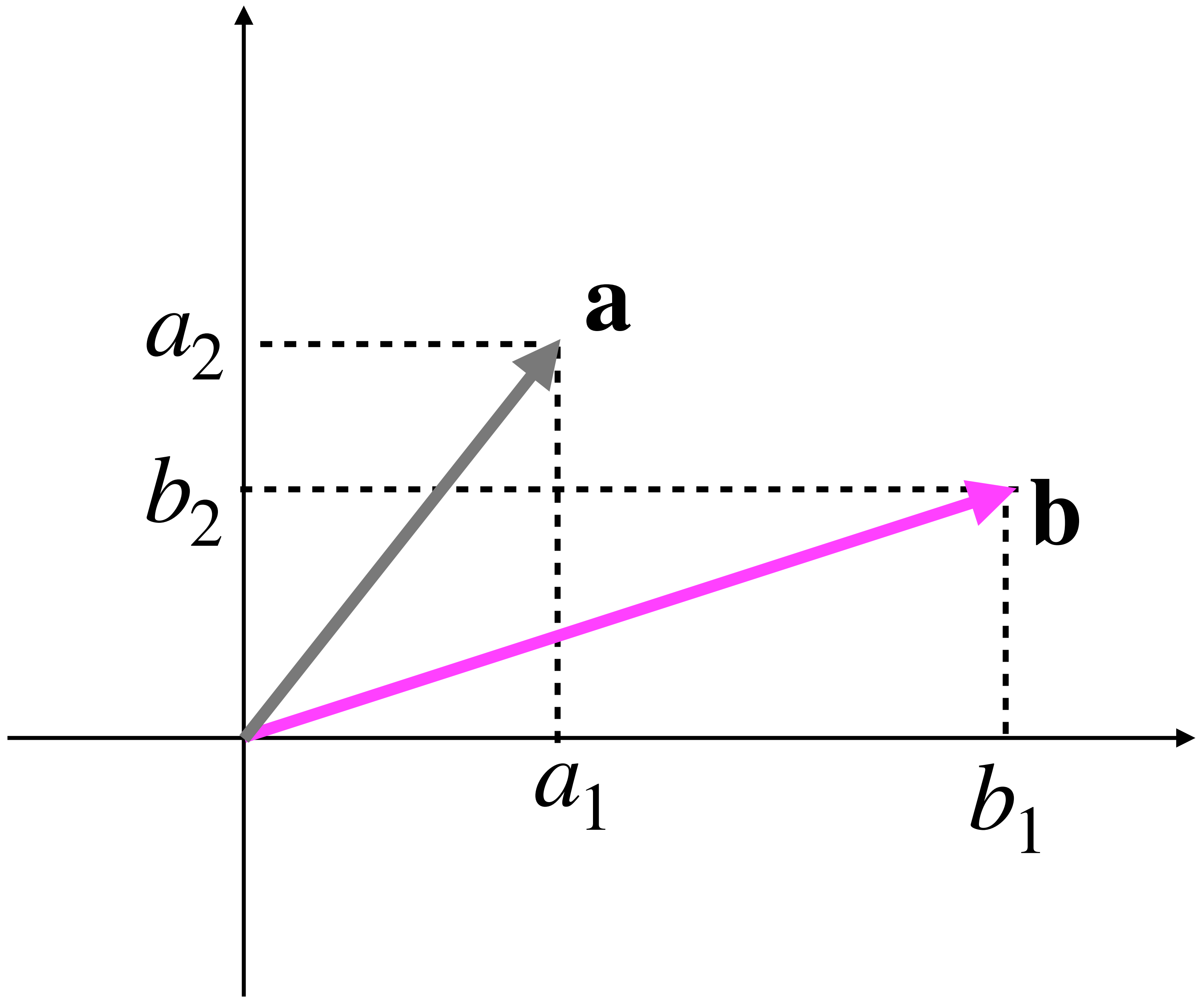
$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad \mathbf{a}, \mathbf{b} \in \mathbb{R}^{2\times 1}$$

$$\mathbf{a} \cdot \mathbf{b} = \sum_{j=1}^{d} a_j b_j = a_1 b_1 + a_2 b_2$$

$$\mathbf{a} \cdot \mathbf{b} = \langle \mathbf{a}, \mathbf{b} \rangle$$

$$\mathbf{a} \cdot \mathbf{b} = \langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^\top \mathbf{b}$$

Inner/dot/scalar product



$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad \mathbf{a}, \mathbf{b} \in \mathbb{R}^{2\times 1}$$

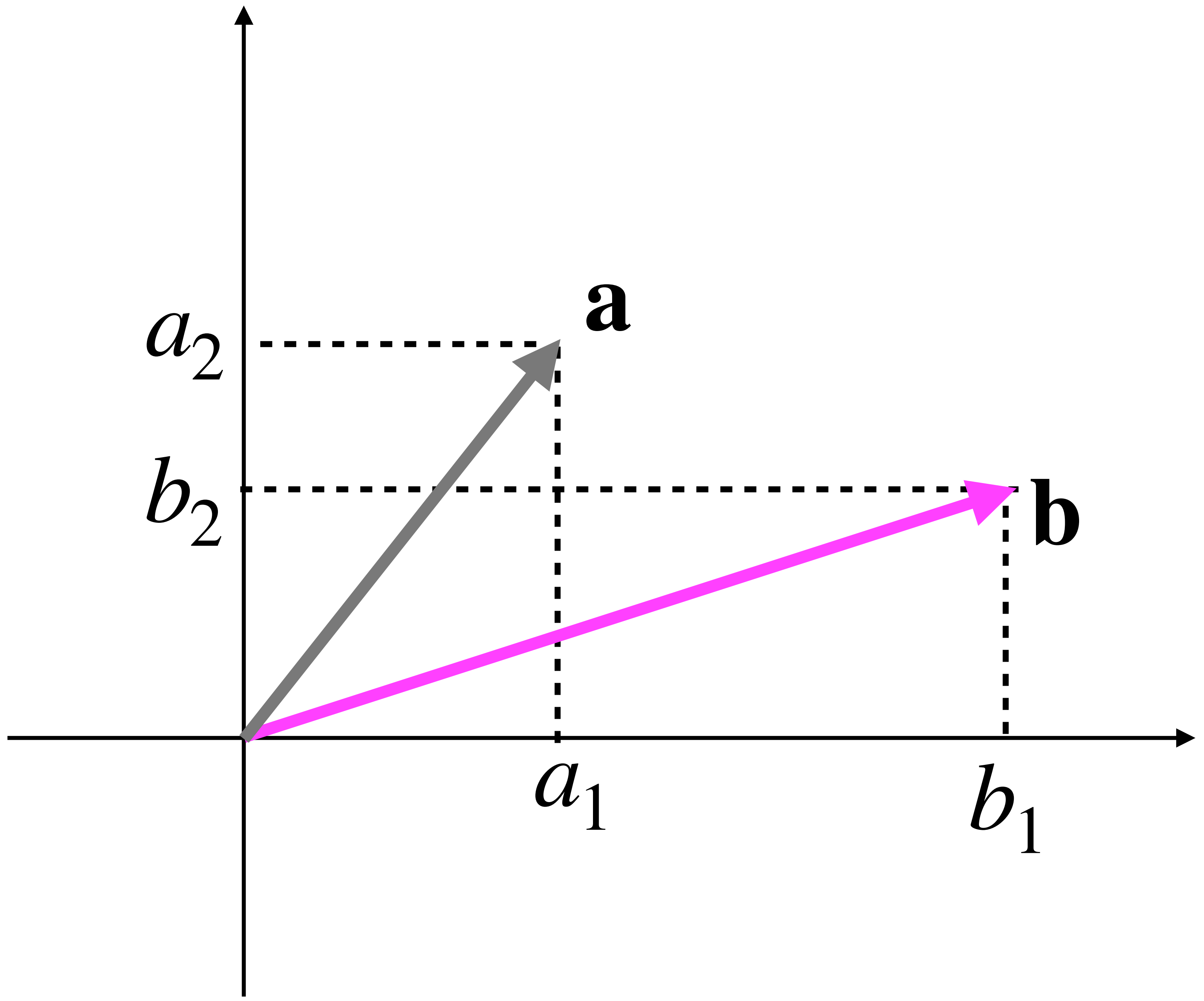
$$\mathbf{a} \cdot \mathbf{b} = \sum_{j=1}^{d} a_j b_j = a_1 b_1 + a_2 b_2$$

$$\mathbf{a} \cdot \mathbf{b} = \langle \mathbf{a}, \mathbf{b} \rangle$$

$$\mathbf{a} \cdot \mathbf{b} = \langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^\top \mathbf{b}$$

$$\mathbf{a}^\top \mathbf{b} = (a_1 \quad a_2) \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

Inner/dot/scalar product



$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad \mathbf{a}, \mathbf{b} \in \mathbb{R}^{2\times 1}$$

$$\mathbf{a} \cdot \mathbf{b} = \sum_{j=1}^{d} a_j b_j = a_1 b_1 + a_2 b_2$$

$$\mathbf{a} \cdot \mathbf{b} = \langle \mathbf{a}, \mathbf{b} \rangle$$

$$\mathbf{a} \cdot \mathbf{b} = \langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^\top \mathbf{b}$$

$$\mathbf{a}^\top \mathbf{b} = \begin{pmatrix} a_1 & a_2 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \mathbf{b}^\top \mathbf{a} = \begin{pmatrix} b_1 & b_2 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

Inner/dot/scalar product

$$\mathbf{a} \cdot \mathbf{a} = \sum_{j=1}^{d} a_j a_j = \sum_{j=1}^{d} a_j^2$$

Inner/dot/scalar product

$$\mathbf{a} \cdot \mathbf{a} = \sum_{j=1}^{d} a_j a_j = \sum_{j=1}^{d} a_j^2$$

$$\mathbf{a} \cdot \mathbf{a} = \|\mathbf{a}\|^2$$

Inner/dot/scalar product

$$\mathbf{a} \cdot \mathbf{a} = \sum_{j=1}^{d} a_j a_j = \sum_{j=1}^{d} a_j^2$$

$$\mathbf{a} \cdot \mathbf{a} = \|\mathbf{a}\|^2$$

Norm:

$$\|\mathbf{a}\| := \sqrt{\langle \mathbf{a}, \mathbf{a} \rangle} := \sqrt{\sum_{j=1}^{d} a_j^2}$$

Eigenvalues and Eigenvectors

Given a squared matrix $\mathbf{X} \in \mathbb{R}^{d \times d}$ a vector $\mathbf{w_i} \in \mathbb{R}^{d \times 1}$ that satisfies this equation

$$\mathbf{X}\mathbf{w_i} = \lambda_\mathbf{i}\mathbf{w_i}$$

Eigenvalues and Eigenvectors

Given a squared matrix $\mathbf{X} \in \mathbb{R}^{d \times d}$ a vector $\mathbf{w_i} \in \mathbb{R}^{d \times 1}$ that satisfies this equation

$$\mathbf{X}\mathbf{w_i} = \lambda_{\mathbf{i}}\mathbf{w_i}$$

It is called eigenvector of the matrix and lambda is the correspondent eigenvalue

Eigenvalues and Eigenvectors

The equation

$$\mathbf{X}\mathbf{w_i} = \lambda_\mathbf{i}\mathbf{w_i}$$

Eigenvalues and Eigenvectors

The equation

$$\mathbf{X}\mathbf{w}_\mathbf{i} = \lambda_\mathbf{i}\mathbf{w}_\mathbf{i}$$

Can be written as

$$(\mathbf{X} - \lambda_\mathbf{i}\mathbf{I})\mathbf{w}_\mathbf{i} = 0$$

Eigenvalues and Eigenvectors

The equation

$$\mathbf{X}\mathbf{w_i} = \lambda_\mathbf{i}\mathbf{w_i}$$

Can be written as

$$(\mathbf{X} - \lambda_\mathbf{i}\mathbf{I})\mathbf{w_i} = 0$$

This has non zero solutions if and only if

Eigenvalues and Eigenvectors

The equation

$$\mathbf{X}\mathbf{w_i} = \lambda_{\mathbf{i}}\mathbf{w_i}$$

Can be written as

$$(\mathbf{X} - \lambda_{\mathbf{i}}\mathbf{I})\mathbf{w_i} = 0$$

This has non zero solutions if and only if

$$\det(\mathbf{X} - \lambda_{\mathbf{i}}\mathbf{I}) = 0$$

Eigenvalues and Eigenvectors

The equation

$$\mathbf{X}\mathbf{w_i} = \lambda_\mathbf{i}\mathbf{w_i}$$

Can be written as

$$(\mathbf{X} - \lambda_\mathbf{i}\mathbf{I})\mathbf{w_i} = 0$$

This has non zero solutions if and only if

$$\det(\mathbf{X} - \lambda_\mathbf{i}\mathbf{I}) = 0$$

Hence the goal is find solutions of this so-called characteristic equation

Eigenvalues and Eigenvectors

$$\det(\mathbf{X} - \lambda_{\mathbf{i}}\mathbf{I}) = 0$$

Eigenvalues and Eigenvectors

$$\det(\mathbf{X} - \lambda_{\mathbf{i}}\mathbf{I}) = 0$$

If the matrix is $d \times d$ we have d solutions (some might be complex, some might have an algebraic multiplicity >1)

Eigenvalues and Eigenvectors

$$\det(\mathbf{X} - \lambda_{\mathbf{i}}\mathbf{I}) = 0$$

If the matrix is $d \times d$ we have d solutions (some might be complex, some might have an algebraic multiplicity >1)

Symmetric matrices lead to d real solutions and orthogonal eigenvectors!

(i) Find eigenvalues, eigenvectors and eigenvalue decomposition of matrix A.

$$A = \begin{pmatrix} 3 & 4 \\ 0 & 5 \end{pmatrix}$$

(i) Find eigenvalues, eigenvectors and eigenvalue decomposition of matrix A.

$$A = \begin{pmatrix} 3 & 4 \\ 0 & 5 \end{pmatrix}$$

The eigenvalues of matrix A can be found by solving $\det(A - \lambda I) = 0$. In our case one has

$$\det(A - \lambda I) = \det \begin{pmatrix} 3-\lambda & 4 \\ 0 & 5-\lambda \end{pmatrix} = (3-\lambda)(5-\lambda) = 0 \Rightarrow \lambda_{1,2} = 3, 5.$$

(i) Find eigenvalues, eigenvectors and eigenvalue decomposition of matrix A.

Corresponding eigenvectors can be found by solving $Au = \lambda u$ for the values of λ found above.

$$Au^{(1)} = 3u^{(1)} \Rightarrow \begin{cases} 3u_1^{(1)} + & 4u_2^{(1)} = 3u_1^{(1)} \\ & 5u_2^{(1)} = 3u_2^{(1)} \end{cases} \Rightarrow u^{(1)} = (1, 0)^T .$$

$$Au^{(2)} = 5u^{(2)} \Rightarrow \begin{cases} 3u_1^{(2)} + & 4u_2^{(2)} = 5u_1^{(2)} \\ & 5u_2^{(2)} = 5u_2^{(2)} \end{cases} \Rightarrow u^{(2)} = (2, 1)^T .$$

$$\begin{pmatrix} 3 & 4 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} u_1^{(1)} \\ u_1^{(2)} \end{pmatrix} = \lambda_1 \begin{pmatrix} u_1^{(1)} \\ u_1^{(2)} \end{pmatrix} = \begin{pmatrix} 3u_1^{(1)} \\ 3u_1^{(2)} \end{pmatrix}$$

Singular value decomposition (SVD)

As we will see soon, in most cases we will work with data matrices that are rarely squared. So how can we decompose such matrices?

Singular value decomposition (SVD)

As we will see soon, in most cases we will work with data matrices that are rarely squared. So how can we decompose such matrices?

Given a $\mathbf{X} \in \mathbb{R}^{s \times d}$ the transpose is $\mathbf{X}^\top \in \mathbb{R}^{d \times s}$ hence $\mathbf{X}^\top \mathbf{X} \in \mathbb{R}^{d \times d}$

Singular value decomposition (SVD)

It can be shown that for any matrix $\mathbf{X} \in \mathbb{R}^{s \times d}$ there exist $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\min(s,d)}$ vectors $\mathbf{U}_i \ \ (i \in [0,s])$ and vectors $\mathbf{V}_j \ \ (j \in [1,d])$ such that

Singular value decomposition (SVD)

It can be shown that for any matrix $\mathbf{X} \in \mathbb{R}^{s \times d}$ there exist $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\min(s,d)}$ vectors $\mathbf{U}_i \ \ (i \in [0,s])$ and vectors $\mathbf{V}_j \ \ (j \in [1,d])$ such that

$$\mathbf{X}\mathbf{V}_i = \sigma_i \mathbf{U}$$

Singular value decomposition (SVD)

It can be shown that for any matrix $\mathbf{X} \in \mathbb{R}^{s \times d}$ there exist $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\min(s,d)}$ vectors $\mathbf{U}_i \ \ (i \in [0,s])$ and vectors $\mathbf{V}_j \ \ (j \in [1,d])$ such that

$$\mathbf{X}\mathbf{V}_i = \sigma_i \mathbf{U}$$

$$\mathbf{X}^\top \mathbf{U}_i = \sigma_i \mathbf{V}$$

Singular value decomposition (SVD)

It can be shown that for any matrix $\mathbf{X} \in \mathbb{R}^{s \times d}$ there exist $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\min(s,d)}$ vectors $\mathbf{U}_i \ \ (i \in [0,s])$ and vectors $\mathbf{V}_j \ \ (j \in [1,d])$ such that

$$\mathbf{X}\mathbf{V}_i = \sigma_i \mathbf{U}$$

$$\mathbf{U}_i = \sigma_i^{-1} \mathbf{X}\mathbf{V}_i$$

$$\mathbf{X}^\top \mathbf{U}_i = \sigma_i \mathbf{V}$$

Singular value decomposition (SVD)

It can be shown that for any matrix $\mathbf{X} \in \mathbb{R}^{s \times d}$ there exist $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\min(s,d)}$ vectors $\mathbf{U}_i \quad (i \in [0,s])$ and vectors $\mathbf{V}_j \quad (j \in [1,d])$ such that

$$\mathbf{X}\mathbf{V}_i = \sigma_i \mathbf{U} \qquad \mathbf{U}_i = \sigma_i^{-1}\mathbf{X}\mathbf{V}_i \qquad \mathbf{V}_i = \sigma_i^{-1}\mathbf{X}^\top\mathbf{U}_i$$

$$\mathbf{X}^\top\mathbf{U}_i = \sigma_i \mathbf{V}$$

Singular value decomposition (SVD)

It can be shown that for any matrix $\mathbf{X} \in \mathbb{R}^{s \times d}$ there exist $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\min(s,d)}$ vectors $\mathbf{U}_i \ \ (i \in [0,s])$ and vectors $\mathbf{V}_j \ \ (j \in [1,d])$ such that

$$\mathbf{X}\mathbf{V}_i = \sigma_i \mathbf{U} \qquad \mathbf{U}_i = \sigma_i^{-1}\mathbf{X}\mathbf{V}_i \qquad \mathbf{V}_i = \sigma_i^{-1}\mathbf{X}^\top\mathbf{U}_i$$

$$\mathbf{X}^\top\mathbf{U}_i = \sigma_i \mathbf{V} \qquad \mathbf{X}^\top\mathbf{X}\mathbf{V}_i = \sigma_i^2\mathbf{V}_i$$

Singular value decomposition (SVD)

It can be shown that for any matrix $\mathbf{X} \in \mathbb{R}^{s \times d}$ there exist $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\min(s,d)}$ vectors $\mathbf{U}_i \ \ (i \in [0,s])$ and vectors $\mathbf{V}_j \ \ (j \in [1,d])$ such that

$$\mathbf{X}\mathbf{V}_i = \sigma_i \mathbf{U} \qquad \mathbf{U}_i = \sigma_i^{-1}\mathbf{X}\mathbf{V}_i \qquad \mathbf{V}_i = \sigma_i^{-1}\mathbf{X}^\top \mathbf{U}_i$$

$$\mathbf{X}^\top \mathbf{U}_i = \sigma_i \mathbf{V} \qquad \mathbf{X}^\top \mathbf{X}\mathbf{V}_i = \sigma_i^2 \mathbf{V}_i \qquad \mathbf{X}\mathbf{X}^\top \mathbf{U}_i = \sigma_i^2 \mathbf{U}_i$$

Singular value decomposition (SVD)

Important for rectangular matrices

It can be shown that for any matrix $\mathbf{X} \in \mathbb{R}^{s \times d}$ there exist $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\min(s,d)}$ vectors $\mathbf{U}_i \;\; (i \in [0,s])$ and vectors $\mathbf{V}_j \;\; (j \in [1,d])$ such that

$$\mathbf{X}\mathbf{V}_i = \sigma_i \mathbf{U} \qquad \mathbf{U}_i = \sigma_i^{-1}\mathbf{X}\mathbf{V}_i \qquad \mathbf{V}_i = \sigma_i^{-1}\mathbf{X}^\top \mathbf{U}_i$$

$$\mathbf{X}^\top \mathbf{U}_i = \sigma_i \mathbf{V} \qquad \mathbf{X}^\top \mathbf{X}\mathbf{V}_i = \sigma_i^2 \mathbf{V}_i \qquad \mathbf{X}\mathbf{X}^\top \mathbf{U}_i = \sigma_i^2 \mathbf{U}_i$$

Singular value decomposition (SVD)

Important for rectangular matrices

It can be shown that for any matrix $\mathbf{X} \in \mathbb{R}^{s \times d}$ there exist $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\min(s,d)}$ vectors $\mathbf{U}_i \quad (i \in [0,s])$ and vectors $\mathbf{V}_j \quad (j \in [1,d])$ such that

$$\mathbf{X}\mathbf{V}_i = \sigma_i \mathbf{U} \qquad \mathbf{U}_i = \sigma_i^{-1}\mathbf{X}\mathbf{V}_i \qquad \mathbf{V}_i = \sigma_i^{-1}\mathbf{X}^\top\mathbf{U}_i$$

$$\mathbf{X}^\top\mathbf{U}_i = \sigma_i \mathbf{V} \qquad \mathbf{X}^\top\mathbf{X}\mathbf{V}_i = \sigma_i^2\mathbf{V}_i \qquad \mathbf{X}\mathbf{X}^\top\mathbf{U}_i = \sigma_i^2\mathbf{U}_i$$

σ_i are called singular values

Singular value decomposition (SVD)

We then define $\mathbf{V} \in \mathbb{R}^{d \times d}$ such that of $\mathbf{X}^\top \mathbf{X} \mathbf{V_i} = \sigma_{\mathbf{i}}^{\mathbf{2}} \mathbf{V_i}$

Singular value decomposition (SVD)

We then define $\mathbf{V} \in \mathbb{R}^{d \times d}$ such that of $\mathbf{X}^\top \mathbf{X} \mathbf{V_i} = \sigma_{\mathbf{i}}^{\mathbf{2}} \mathbf{V_i}$

Each $\mathbf{V_i}$ is an eigenvector of $\mathbf{X}^\top \mathbf{X}$ correspondent to the eigenvalue σ_i^2

Singular value decomposition (SVD)

We then define $\mathbf{U} \in \mathbb{R}^{s \times s}$ such that of $\mathbf{X}\mathbf{X}^\top \mathbf{U_i} = \sigma_{\mathbf{i}}^{\mathbf{2}} \mathbf{U_i}$

Singular value decomposition (SVD)

We then define $\mathbf{U} \in \mathbb{R}^{s \times s}$ such that of $\mathbf{X}\mathbf{X}^{\top}\mathbf{U_i} = \sigma_\mathbf{i}^\mathbf{2}\mathbf{U_i}$

Each $\mathbf{U_i}$ is an eigenvector of $\mathbf{X}\mathbf{X}^{\top}$ correspondent to the eigenvalue σ_i^2

n.perra@qmul.ac.uk

Singular value decomposition (SVD)

Summarizing

$$\mathbf{X}^\top \mathbf{X}\mathbf{V} = \sigma^2 \mathbf{V}$$

$$\mathbf{X}\mathbf{X}^\top \mathbf{U} = \sigma^2 \mathbf{U}$$

Also

$$\mathbf{X} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$$

(c) Consider the following matrix $\mathbf{B} = \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}$ compute its singular values.

(c) Consider the following matrix $\mathbf{B} = \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}$ compute its singular values.

We have two options:

(c) Consider the following matrix $\mathbf{B} = \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}$ compute its singular values.

We have two options:

$$\mathbf{B}^{\top}\mathbf{B}\mathbf{V}_{\mathbf{i}} = \sigma_{\mathbf{i}}^{2}\mathbf{V}_{\mathbf{i}}$$

(c) Consider the following matrix $\mathbf{B} = \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}$ compute its singular values.

We have two options:

$$\mathbf{B}^{\top}\mathbf{B}\mathbf{V}_\mathbf{i} = \sigma_\mathbf{i}^2\mathbf{V}_\mathbf{i}$$

$$\mathbf{B}\mathbf{B}^{\top}\mathbf{U}_\mathbf{i} = \sigma_\mathbf{i}^2\mathbf{U}_\mathbf{i}$$

(c) Consider the following matrix $\mathbf{B} = \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}$ compute its singular values.

We have two options:

$$\mathbf{B}^\top \mathbf{B} \mathbf{V_i} = \sigma_\mathbf{i}^2 \mathbf{V_i}$$

$$\mathbf{B}\mathbf{B}^\top \mathbf{U_i} = \sigma_\mathbf{i}^2 \mathbf{U_i}$$

The results are the same, but one is more convenient than the other

(c) Consider the following matrix $\mathbf{B} = \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}$ compute its singular values.

We have two options:

$$\mathbf{B}^\top \mathbf{B} \mathbf{V_i} = \sigma_{\mathbf{i}}^2 \mathbf{V_i}$$

$$\mathbf{B}\mathbf{B}^\top \mathbf{U_i} = \sigma_{\mathbf{i}}^2 \mathbf{U_i}$$

The results are the same, but one is more convenient than the other

$$\mathbf{B}^\top \mathbf{B} \rightarrow 2x3 \times 3x2 \rightarrow 2 \times 2$$

(c) Consider the following matrix $\mathbf{B} = \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}$ compute its singular values.

We have two options:

$$\mathbf{B}^\top \mathbf{B} \mathbf{V_i} = \sigma_{\mathbf{i}}^2 \mathbf{V_i}$$

$$\mathbf{B}\mathbf{B}^\top \mathbf{U_i} = \sigma_{\mathbf{i}}^2 \mathbf{U_i}$$

The results are the same, but one is more convenient than the other

$$\mathbf{B}^\top \mathbf{B} \rightarrow 2x3 \times 3x2 \rightarrow 2 \times 2$$
$$\mathbf{B}\mathbf{B}^\top \rightarrow 3x2 \times 2x3 \rightarrow 3 \times 3$$

(c) Consider the following matrix $\mathbf{B} = \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}$ compute its singular values.

We have two options:

$$\mathbf{B}^\top \mathbf{B} \mathbf{V_i} = \sigma_\mathbf{i}^2 \mathbf{V_i}$$

$$\mathbf{B}\mathbf{B}^\top \mathbf{U_i} = \sigma_\mathbf{i}^2 \mathbf{U_i}$$

The results are the same, but one is more convenient than the other

$$\mathbf{B}^\top \mathbf{B} \rightarrow 2x3 \times 3x2 \rightarrow 2 \times 2$$
$$\mathbf{B}\mathbf{B}^\top \rightarrow 3x2 \times 2x3 \rightarrow 3 \times 3$$

Hence, it is a good idea picking the first!

(c) Consider the following matrix $\mathbf{B} = \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}$ compute its singular values.

(c) Consider the following matrix $\mathbf{B} = \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}$ compute its singular values.

$$\mathbf{B}^\top \mathbf{B} = \begin{pmatrix} 1 & 1 & 0 \\ -2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}$$

(c) Consider the following matrix $\mathbf{B} = \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}$ compute its singular values.

$$\mathbf{B}^\top \mathbf{B} = \begin{pmatrix} 1 & 1 & 0 \\ -2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix}$$

(c) Consider the following matrix $\mathbf{B} = \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}$ compute its singular values.

$$\mathbf{B}^{\top}\mathbf{B} = \begin{pmatrix} 1 & 1 & 0 \\ -2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix}$$

To find the singular values, we need to compute the eingevalues of this matrix!

(c) Consider the following matrix $\mathbf{B} = \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}$ compute its singular values.

$$\mathbf{B}^\top \mathbf{B} = \begin{pmatrix} 1 & 1 & 0 \\ -2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix}$$

To find the singular values, we need to compute the eingevalues of this matrix!

$$\mathbf{B}^\top \mathbf{B} \mathbf{V} = \sigma^2 \mathbf{V}$$

(c) Consider the following matrix $\mathbf{B} = \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}$ compute its singular values.

$$\mathbf{B}^\top \mathbf{B} = \begin{pmatrix} 1 & 1 & 0 \\ -2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix}$$

To find the singular values, we need to compute the eingevalues of this matrix!

$$\mathbf{B}^\top \mathbf{B} \mathbf{V} = \sigma^2 \mathbf{V}$$

$$(\mathbf{B}^\top \mathbf{B} - \sigma^2 \mathbf{I})\mathbf{V} = 0$$

(c) Consider the following matrix $\mathbf{B} = \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}$ compute its singular values.

$$\mathbf{B}^\top\mathbf{B} = \begin{pmatrix} 1 & 1 & 0 \\ -2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix}$$

To find the singular values, we need to compute the eingevalues of this matrix!

$$\mathbf{B}^\top\mathbf{B}\mathbf{V} = \sigma^2\mathbf{V}$$

$$(\mathbf{B}^\top\mathbf{B} - \sigma^2\mathbf{I})\mathbf{V} = 0$$

Non-zero solutions exist if and only if

(c) Consider the following matrix $\mathbf{B} = \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}$ compute its singular values.

$$\mathbf{B}^\top \mathbf{B} = \begin{pmatrix} 1 & 1 & 0 \\ -2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix}$$

To find the singular values, we need to compute the eingevalues of this matrix!

$$\mathbf{B}^\top \mathbf{B} \mathbf{V} = \sigma^2 \mathbf{V}$$

$$(\mathbf{B}^\top \mathbf{B} - \sigma^2 \mathbf{I})\mathbf{V} = 0$$

Non-zero solutions exist if and only if

$$\det[\mathbf{B}^\top \mathbf{B} - \sigma^2 I] = 0$$

(c) Consider the following matrix $\mathbf{B} = \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}$ compute its singular values.

$$\det \begin{vmatrix} 2-\sigma^2 & -1 \\ -1 & 5-\sigma^2 \end{vmatrix} = 0$$

(c) Consider the following matrix $\mathbf{B} = \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}$ compute its singular values.

$$\det \begin{vmatrix} 2-\sigma^2 & -1 \\ -1 & 5-\sigma^2 \end{vmatrix} = 0$$

$$(2-\sigma^2)(5-\sigma^2) - 1 = 0$$

(c) Consider the following matrix $\mathbf{B} = \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}$ compute its singular values.

$$\det \begin{vmatrix} 2-\sigma^2 & -1 \\ -1 & 5-\sigma^2 \end{vmatrix} = 0$$

$$(2-\sigma^2)(5-\sigma^2) - 1 = 0$$

$$\sigma^4 - 7\sigma^2 + 9 = 0$$

(c) Consider the following matrix $\mathbf{B} = \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}$ compute its singular values.

$$\det \begin{vmatrix} 2-\sigma^2 & -1 \\ -1 & 5-\sigma^2 \end{vmatrix} = 0$$

$$(2-\sigma^2)(5-\sigma^2) - 1 = 0$$

$$\sigma^4 - 7\sigma^2 + 9 = 0$$

$$\sigma_{1,2}^2 = \frac{7 \pm \sqrt{13}}{2}$$

Compute the left and right singular vectors of the matrix **B**.

$$\mathbf{B}^\top \mathbf{B} \mathbf{V}_1 = \sigma_1^2 \mathbf{V}_1$$

$$\begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix} \begin{pmatrix} V_1^{(1)} \\ V_1^{(2)} \end{pmatrix} = \sigma_1^2 \begin{pmatrix} V_1^{(1)} \\ V_1^{(2)} \end{pmatrix}$$

Compute the left and right singular vectors of the matrix **B**.

$$\mathbf{B}^\top \mathbf{B} \mathbf{V}_1 = \sigma_1^2 \mathbf{V}_1$$

$$\begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix} \begin{pmatrix} V_1^{(1)} \\ V_1^{(2)} \end{pmatrix} = \sigma_1^2 \begin{pmatrix} V_1^{(1)} \\ V_1^{(2)} \end{pmatrix}$$

$\{$

Compute the left and right singular vectors of the matrix **B**.

$$\mathbf{B}^\top \mathbf{B} \mathbf{V}_1 = \sigma_1^2 \mathbf{V}_1$$

$$\begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix} \begin{pmatrix} V_1^{(1)} \\ V_1^{(2)} \end{pmatrix} = \sigma_1^2 \begin{pmatrix} V_1^{(1)} \\ V_1^{(2)} \end{pmatrix}$$

$$\begin{cases} 2V_1^{(1)} - V_1^{(2)} = \sigma_1^2 V_1^{(1)} \\ \end{cases}$$

Compute the left and right singular vectors of the matrix **B**.

$$\mathbf{B}^\top \mathbf{B} \mathbf{V}_1 = \sigma_1^2 \mathbf{V}_1$$

$$\begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix} \begin{pmatrix} V_1^{(1)} \\ V_1^{(2)} \end{pmatrix} = \sigma_1^2 \begin{pmatrix} V_1^{(1)} \\ V_1^{(2)} \end{pmatrix}$$

$$\begin{cases} 2V_1^{(1)} - V_1^{(2)} = \sigma_1^2 V_1^{(1)} \\ -V_1^{(1)} + 5V_1^{(2)} = \sigma_1^2 V_1^{(2)} \end{cases}$$

Compute the left and right singular vectors of the matrix **B**.

$$\mathbf{B}^\top \mathbf{B} \mathbf{V}_1 = \sigma_1^2 \mathbf{V}_1$$

$$\begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix} \begin{pmatrix} V_1^{(1)} \\ V_1^{(2)} \end{pmatrix} = \sigma_1^2 \begin{pmatrix} V_1^{(1)} \\ V_1^{(2)} \end{pmatrix}$$

$$\begin{cases} 2V_1^{(1)} - V_1^{(2)} = \sigma_1^2 V_1^{(1)} \\ -V_1^{(1)} + 5V_1^{(2)} = \sigma_1^2 V_1^{(2)} \end{cases}$$

$$V_1^{(2)} = V_1^{(1)}(2 - \sigma_1^2)$$

Compute the left and right singular vectors of the matrix **B**.

$$\mathbf{B}^\top \mathbf{B} \mathbf{V}_1 = \sigma_1^2 \mathbf{V}_1$$

$$\begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix} \begin{pmatrix} V_1^{(1)} \\ V_1^{(2)} \end{pmatrix} = \sigma_1^2 \begin{pmatrix} V_1^{(1)} \\ V_1^{(2)} \end{pmatrix}$$

$$\begin{cases} 2V_1^{(1)} - V_1^{(2)} = \sigma_1^2 V_1^{(1)} \\ -V_1^{(1)} + 5V_1^{(2)} = \sigma_1^2 V_1^{(2)} \end{cases}$$

$$V_1^{(2)} = V_1^{(1)}(2 - \sigma_1^2) = -V_1^{(1)} \frac{3 + \sqrt{13}}{2}$$

Compute the left and right singular vectors of the matrix **B**.

$$\mathbf{B}^\top \mathbf{B} \mathbf{V}_1 = \sigma_1^2 \mathbf{V}_1$$

$$\begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix} \begin{pmatrix} V_1^{(1)} \\ V_1^{(2)} \end{pmatrix} = \sigma_1^2 \begin{pmatrix} V_1^{(1)} \\ V_1^{(2)} \end{pmatrix}$$

$$\begin{cases} 2V_1^{(1)} - V_1^{(2)} = \sigma_1^2 V_1^{(1)} \\ -V_1^{(1)} + 5V_1^{(2)} = \sigma_1^2 V_1^{(2)} \end{cases}$$

$$V_1^{(2)} = V_1^{(1)}(2 - \sigma_1^2) = -V_1^{(1)} \frac{3 + \sqrt{13}}{2}$$

$$\mathbf{V}_1 = \left(c, -c\frac{3 + \sqrt{13}}{2} \right)^\top$$

Compute the left and right singular vectors of the matrix **B**.

$$\mathbf{B}^\top \mathbf{B} \mathbf{V}_1 = \sigma_1^2 \mathbf{V}_1$$

$$\begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix} \begin{pmatrix} V_1^{(1)} \\ V_1^{(2)} \end{pmatrix} = \sigma_1^2 \begin{pmatrix} V_1^{(1)} \\ V_1^{(2)} \end{pmatrix}$$

$$\begin{cases} 2V_1^{(1)} - V_1^{(2)} = \sigma_1^2 V_1^{(1)} \\ -V_1^{(1)} + 5V_1^{(2)} = \sigma_1^2 V_1^{(2)} \end{cases}$$

$$V_1^{(2)} = V_1^{(1)}(2 - \sigma_1^2) = -V_1^{(1)} \frac{3+\sqrt{13}}{2}$$

$$\mathbf{V}_1 = \left(c, -c\frac{3+\sqrt{13}}{2}\right)^\top$$

Another solution can be obtained from the second equation, the two
Are just proportional!

Compute the left and right singular vectors of the matrix **B**.

$$\mathbf{B}^\top \mathbf{B} \mathbf{V}_2 = \sigma_2^2 \mathbf{v}_2$$

$$\begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix} \begin{pmatrix} V_2^{(1)} \\ V_2^{(2)} \end{pmatrix} = \sigma_2^2 \begin{pmatrix} V_2^{(1)} \\ V_2^{(2)} \end{pmatrix}$$

$$\begin{cases} 2V_2^{(1)} - V_2^{(2)} = \sigma_2^2 V_2^{(1)} \\ -V_2^{(1)} + 5V_2^{(2)} = \sigma_2^2 V_2^{(2)} \end{cases}$$

Compute the left and right singular vectors of the matrix **B**.

$$\mathbf{B}^{\top}\mathbf{B}\mathbf{V}_2 = \sigma_2^2 \mathbf{v}_2$$

$$\begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix} \begin{pmatrix} V_2^{(1)} \\ V_2^{(2)} \end{pmatrix} = \sigma_2^2 \begin{pmatrix} V_2^{(1)} \\ V_2^{(2)} \end{pmatrix}$$

$$\begin{cases} 2V_2^{(1)} - V_2^{(2)} = \sigma_2^2 V_2^{(1)} \\ -V_2^{(1)} + 5V_2^{(2)} = \sigma_2^2 V_2^{(2)} \end{cases}$$

$$V_2^{(2)} = V_2^{(1)}(2 - \sigma_2^2)$$

Compute the left and right singular vectors of the matrix **B**.

$$\mathbf{B}^\top \mathbf{B} \mathbf{V}_2 = \sigma_2^2 \mathbf{v}_2$$

$$\begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix} \begin{pmatrix} V_2^{(1)} \\ V_2^{(2)} \end{pmatrix} = \sigma_2^2 \begin{pmatrix} V_2^{(1)} \\ V_2^{(2)} \end{pmatrix}$$

$$\begin{cases} 2V_2^{(1)} - V_2^{(2)} = \sigma_2^2 V_2^{(1)} \\ -V_2^{(1)} + 5V_2^{(2)} = \sigma_2^2 V_2^{(2)} \end{cases}$$

$$V_2^{(2)} = V_2^{(1)}(2 - \sigma_2^2) = V_2^{(1)} \frac{\sqrt{13} - 3}{2}$$

Compute the left and right singular vectors of the matrix **B**.

$$\mathbf{B}^\top \mathbf{B} \mathbf{V}_2 = \sigma_2^2 \mathbf{v}_2$$

$$\begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix} \begin{pmatrix} V_2^{(1)} \\ V_2^{(2)} \end{pmatrix} = \sigma_2^2 \begin{pmatrix} V_2^{(1)} \\ V_2^{(2)} \end{pmatrix}$$

$$\begin{cases} 2V_2^{(1)} - V_2^{(2)} = \sigma_2^2 V_2^{(1)} \\ -V_2^{(1)} + 5V_2^{(2)} = \sigma_2^2 V_2^{(2)} \end{cases}$$

$$V_2^{(2)} = V_2^{(1)}(2 - \sigma_2^2) = V_2^{(1)} \frac{\sqrt{13} - 3}{2}$$

$$\mathbf{V}_2 = \left(b, b\frac{\sqrt{13} - 3}{2} \right)^\top$$

Compute the left and right singular vectors of the matrix **B**.

$$\mathbf{B}^\top\mathbf{B}\mathbf{V}_2 = \sigma_2^2\mathbf{v}_2$$

$$\begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix}\begin{pmatrix} V_2^{(1)} \\ V_2^{(2)} \end{pmatrix} = \sigma_2^2\begin{pmatrix} V_2^{(1)} \\ V_2^{(2)} \end{pmatrix}$$

$$\begin{cases} 2V_2^{(1)} - V_2^{(2)} = \sigma_2^2 V_2^{(1)} \\ -V_2^{(1)} + 5V_2^{(2)} = \sigma_2^2 V_2^{(2)} \end{cases}$$

$$V_2^{(2)} = V_2^{(1)}(2-\sigma_2^2) = V_2^{(1)}\frac{\sqrt{13}-3}{2}$$

$$\mathbf{V}_2 = \left(b, b\frac{\sqrt{13}-3}{2}\right)^\top$$

Another solution can be obtained from the second equation, the two
Are just proportional!

To compute the left singular vectors we can use

$$\mathbf{U}_1 = \sigma_1^{-1}\mathbf{B}\mathbf{V}_1$$

To compute the left singular vectors we can use

$$\mathbf{U}_1 = \sigma_1^{-1}\mathbf{B}\mathbf{V}_1$$

Or….we could compute solve the eingevalues/eingevectors problem for $\mathbf{B}\mathbf{B}^\top$

To compute the left singular vectors we can use

$$\mathbf{U}_1 = \sigma_1^{-1}\mathbf{B}\mathbf{V}_1$$

Or….we could compute solve the eingevalues/eingevectors problem for $\mathbf{B}\mathbf{B}^\top$

$$\mathbf{U}_1 = \sqrt{\frac{2}{7+\sqrt{13}}}\begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} c \\ -c\frac{3+\sqrt{13}}{2} \end{pmatrix}$$

To compute the left singular vectors we can use

$$\mathbf{U}_1 = \sigma_1^{-1}\mathbf{B}\mathbf{V}_1$$

Or….we could compute solve the eingevalues/eingevectors problem for $\mathbf{B}\mathbf{B}^\top$

$$\mathbf{U}_1 = \sqrt{\frac{2}{7+\sqrt{13}}}\begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} c \\ -c\frac{3+\sqrt{13}}{2} \end{pmatrix}$$

$$= c \times \sqrt{\frac{2}{7+\sqrt{13}}}\begin{pmatrix} 4+\sqrt{13} \\ -\frac{1+\sqrt{13}}{2} \\ 0 \end{pmatrix}$$

To compute the left singular vectors we can use

$$\mathbf{U}_1 = \sigma_1^{-1}\mathbf{B}\mathbf{V}_1$$

Or....we could compute solve the eingevalues/eingevectors problem for $\mathbf{B}\mathbf{B}^\top$

$$\mathbf{U}_1 = \sqrt{\frac{2}{7+\sqrt{13}}}\begin{pmatrix}1 & -2\\ 1 & 1\\ 0 & 0\end{pmatrix}\begin{pmatrix}c\\ -c\frac{3+\sqrt{13}}{2}\end{pmatrix}$$

$$= c \times \sqrt{\frac{2}{7+\sqrt{13}}}\begin{pmatrix}4+\sqrt{13}\\ -\frac{1+\sqrt{13}}{2}\\ 0\end{pmatrix}$$

Notice the dimension of this vector!

To compute the left singular vectors we can use

$$\mathbf{U}_2 = \sigma_2^{-1}\mathbf{B}\mathbf{V}_2$$

$$\mathbf{U}_2 = \sqrt{\frac{2}{7-\sqrt{13}}}\begin{pmatrix}1 & -2\\ 1 & 1\\ 0 & 0\end{pmatrix}\begin{pmatrix} b \\ b\frac{\sqrt{13}-3}{2}\end{pmatrix}$$

$$= b\sqrt{\frac{2}{7-\sqrt{13}}}\begin{pmatrix}4-\sqrt{13}\\ \frac{\sqrt{13}-1}{2}\\ 0\end{pmatrix}$$

Notice the dimension of this vector!

Important note:

1) The matrix $\mathbf{B}\mathbf{B}^\top$ is a 3x3 matrix with another eigenvalue equal to zero
2) the associate vector needs to be orthogonal to the others, hence

$$\mathbf{U}_3 = (0,0,d)^\top$$

for data points $(x^{(1)}, y^{(1)})$ with $x^{(1)} = -c$ and $y^{(1)} = 2$, $(x^{(2)}, y^{(2)})$ with $x^{(2)} = 0$ and $y^{(2)} = 2$, and $(x^{(3)}, y^{(3)})$ with $x^{(3)} = c$ and $y^{(3)} = 2$, for some constant $c > 0$.

1. Derive the normal equation for this problem.

for data points $(x^{(1)}, y^{(1)})$ with $x^{(1)} = -c$ and $y^{(1)} = 2$, $(x^{(2)}, y^{(2)})$ with $x^{(2)} = 0$ and $y^{(2)} = 2$, and $(x^{(3)}, y^{(3)})$ with $x^{(3)} = c$ and $y^{(3)} = 2$, for some constant $c > 0$.

1. Derive the normal equation for this problem.

$$X^\top X\hat{w} = X^\top y$$

for data points $(x^{(1)}, y^{(1)})$ with $x^{(1)} = -c$ and $y^{(1)} = 2$, $(x^{(2)}, y^{(2)})$ with $x^{(2)} = 0$ and $y^{(2)} = 2$, and $(x^{(3)}, y^{(3)})$ with $x^{(3)} = c$ and $y^{(3)} = 2$, for some constant $c > 0$.

1. Derive the normal equation for this problem.

$$X^\top X\hat{w} = X^\top y$$

$$X = \begin{pmatrix} 1 & -c \\ 1 & 0 \\ 1 & c \end{pmatrix}$$

for data points $(x^{(1)}, y^{(1)})$ with $x^{(1)} = -c$ and $y^{(1)} = 2$, $(x^{(2)}, y^{(2)})$ with $x^{(2)} = 0$ and $y^{(2)} = 2$, and $(x^{(3)}, y^{(3)})$ with $x^{(3)} = c$ and $y^{(3)} = 2$, for some constant $c > 0$.

1. Derive the normal equation for this problem.

$$X^\top X\hat{w} = X^\top y$$

$$X = \begin{pmatrix} 1 & -c \\ 1 & 0 \\ 1 & c \end{pmatrix} \qquad X^\top = \begin{pmatrix} 1 & 1 & 1 \\ -c & 0 & c \end{pmatrix}$$

for data points $(x^{(1)}, y^{(1)})$ with $x^{(1)} = -c$ and $y^{(1)} = 2$, $(x^{(2)}, y^{(2)})$ with $x^{(2)} = 0$ and $y^{(2)} = 2$, and $(x^{(3)}, y^{(3)})$ with $x^{(3)} = c$ and $y^{(3)} = 2$, for some constant $c > 0$.

1. Derive the normal equation for this problem.

$$X^\top X\hat{w} = X^\top y$$

$$X = \begin{pmatrix} 1 & -c \\ 1 & 0 \\ 1 & c \end{pmatrix} \qquad X^\top = \begin{pmatrix} 1 & 1 & 1 \\ -c & 0 & c \end{pmatrix} \qquad y = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}$$

for data points $(x^{(1)}, y^{(1)})$ with $x^{(1)} = -c$ and $y^{(1)} = 2$, $(x^{(2)}, y^{(2)})$ with $x^{(2)} = 0$ and $y^{(2)} = 2$, and $(x^{(3)}, y^{(3)})$ with $x^{(3)} = c$ and $y^{(3)} = 2$, for some constant $c > 0$.

1. Derive the normal equation for this problem.

$$X^\top X\hat{w} = X^\top y$$

$$X = \begin{pmatrix} 1 & -c \\ 1 & 0 \\ 1 & c \end{pmatrix} \qquad X^\top = \begin{pmatrix} 1 & 1 & 1 \\ -c & 0 & c \end{pmatrix} \qquad y = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}$$

$$X^\top X\hat{w} = \begin{pmatrix} 1 & 1 & 1 \\ -c & 0 & c \end{pmatrix} \begin{pmatrix} 1 & -c \\ 1 & 0 \\ 1 & c \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix}$$

for data points $(x^{(1)}, y^{(1)})$ with $x^{(1)} = -c$ and $y^{(1)} = 2$, $(x^{(2)}, y^{(2)})$ with $x^{(2)} = 0$ and $y^{(2)} = 2$, and $(x^{(3)}, y^{(3)})$ with $x^{(3)} = c$ and $y^{(3)} = 2$, for some constant $c > 0$.

1. Derive the normal equation for this problem.

$$X^\top X\hat{w} = X^\top y$$

$$X = \begin{pmatrix} 1 & -c \\ 1 & 0 \\ 1 & c \end{pmatrix} \qquad X^\top = \begin{pmatrix} 1 & 1 & 1 \\ -c & 0 & c \end{pmatrix} \qquad y = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}$$

$$X^\top X\hat{w} = \begin{pmatrix} 1 & 1 & 1 \\ -c & 0 & c \end{pmatrix} \begin{pmatrix} 1 & -c \\ 1 & 0 \\ 1 & c \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 0 & 2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix}$$

for data points $(x^{(1)}, y^{(1)})$ with $x^{(1)} = -c$ and $y^{(1)} = 2$, $(x^{(2)}, y^{(2)})$ with $x^{(2)} = 0$ and $y^{(2)} = 2$, and $(x^{(3)}, y^{(3)})$ with $x^{(3)} = c$ and $y^{(3)} = 2$, for some constant $c > 0$.

1. Derive the normal equation for this problem.

$$X^\top = \begin{pmatrix} 1 & 1 & 1 \\ -c & 0 & c \end{pmatrix} \qquad y = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}$$

for data points $(x^{(1)}, y^{(1)})$ with $x^{(1)} = -c$ and $y^{(1)} = 2$, $(x^{(2)}, y^{(2)})$ with $x^{(2)} = 0$ and $y^{(2)} = 2$, and $(x^{(3)}, y^{(3)})$ with $x^{(3)} = c$ and $y^{(3)} = 2$, for some constant $c > 0$.

1. Derive the normal equation for this problem.

$$X^\top = \begin{pmatrix} 1 & 1 & 1 \\ -c & 0 & c \end{pmatrix} \qquad y = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}$$

$$\begin{pmatrix} 3 & 0 \\ 0 & 2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = X^\top y = \begin{pmatrix} 1 & 1 & 1 \\ -c & 0 & c \end{pmatrix} \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}$$

for data points $(x^{(1)}, y^{(1)})$ with $x^{(1)} = -c$ and $y^{(1)} = 2$, $(x^{(2)}, y^{(2)})$ with $x^{(2)} = 0$ and $y^{(2)} = 2$, and $(x^{(3)}, y^{(3)})$ with $x^{(3)} = c$ and $y^{(3)} = 2$, for some constant $c > 0$.

1. Derive the normal equation for this problem.

$$X^\top = \begin{pmatrix} 1 & 1 & 1 \\ -c & 0 & c \end{pmatrix} \qquad y = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}$$

$$\begin{pmatrix} 3 & 0 \\ 0 & 2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = X^\top y = \begin{pmatrix} 1 & 1 & 1 \\ -c & 0 & c \end{pmatrix} \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}$$

$$\begin{pmatrix} 3 & 0 \\ 0 & 2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \end{pmatrix}$$

2. Solve the normal equations for your weights $\hat{\mathbf{w}} = (\hat{w}^{(0)}, \hat{w}^{(1)})^\top$.

$$\begin{pmatrix} 3 & 0 \\ 0 & 2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \end{pmatrix}$$

2. Solve the normal equations for your weights $\hat{\mathbf{w}} = (\hat{w}^{(0)}, \hat{w}^{(1)})^\top$.

$$\begin{pmatrix} 3 & 0 \\ 0 & 2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \end{pmatrix}$$

$$3w_0 = 6 \rightarrow w_0 = 2$$

2. Solve the normal equations for your weights $\hat{\mathbf{w}} = (\hat{w}^{(0)}, \hat{w}^{(1)})^\top$.

$$\begin{pmatrix} 3 & 0 \\ 0 & 2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \end{pmatrix}$$

$$3w_0 = 6 \rightarrow w_0 = 2$$

$$2c^2 w_1 = 0 \rightarrow w_1 = 0$$

2. Solve the normal equations for your weights $\hat{\mathbf{w}} = (\hat{w}^{(0)}, \hat{w}^{(1)})^\top$.

$$\begin{pmatrix} 3 & 0 \\ 0 & 2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \end{pmatrix}$$

$$3w_0 = 6 \rightarrow w_0 = 2$$

$$2c^2 w_1 = 0 \rightarrow w_1 = 0$$

$$\begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

3. Repeat the previous exercise, but this time assume you make an error in your measurement. The new, perturbed measurements $\mathbf{y}_\delta$ read $y_\delta^{(1)} = 2 + \varepsilon$, $y_\delta^{(2)} = 2 + \varepsilon$ and $y_\delta^{(3)} = 2 - \varepsilon$.

3. Repeat the previous exercise, but this time assume you make an error in your measurement. The new, perturbed measurements $\mathbf{y}_\delta$ read $y_\delta^{(1)} = 2 + \varepsilon$, $y_\delta^{(2)} = 2 + \varepsilon$ and $y_\delta^{(3)} = 2 - \varepsilon$.

$$\begin{pmatrix} 3 & 0 \\ 0 & 2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = X^\top y = \begin{pmatrix} 1 & 1 & 1 \\ -c & 0 & c \end{pmatrix} \begin{pmatrix} 2 + \epsilon \\ 2 + \epsilon \\ 2 - \epsilon \end{pmatrix}$$

3. Repeat the previous exercise, but this time assume you make an error in your measurement. The new, perturbed measurements $\mathbf{y}_\delta$ read $y_\delta^{(1)} = 2 + \varepsilon$, $y_\delta^{(2)} = 2 + \varepsilon$ and $y_\delta^{(3)} = 2 - \varepsilon$.

$$\begin{pmatrix} 3 & 0 \\ 0 & 2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = X^\top y = \begin{pmatrix} 1 & 1 & 1 \\ -c & 0 & c \end{pmatrix} \begin{pmatrix} 2+\epsilon \\ 2+\epsilon \\ 2-\epsilon \end{pmatrix}$$

$$\begin{pmatrix} 3 & 0 \\ 0 & 2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 6+\epsilon \\ -2c\epsilon \end{pmatrix}$$

3. Repeat the previous exercise, but this time assume you make an error in your measurement. The new, perturbed measurements $\mathbf{y}_\delta$ read $y_\delta^{(1)} = 2 + \varepsilon$, $y_\delta^{(2)} = 2 + \varepsilon$ and $y_\delta^{(3)} = 2 - \varepsilon$.

$$\begin{pmatrix} 3 & 0 \\ 0 & 2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = X^\top y = \begin{pmatrix} 1 & 1 & 1 \\ -c & 0 & c \end{pmatrix} \begin{pmatrix} 2+\epsilon \\ 2+\epsilon \\ 2-\epsilon \end{pmatrix}$$

$$\begin{pmatrix} 3 & 0 \\ 0 & 2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 6+\epsilon \\ -2c\epsilon \end{pmatrix}$$

$$\begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 + \frac{\epsilon}{3} \\ -\frac{\epsilon}{c} \end{pmatrix}$$

4. Compute the error between $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}_\delta$ in the Euclidean norm.

4. Compute the error between $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}_\delta$ in the Euclidean norm.

$$\|\hat{w} - \hat{w}_\delta\| = \sqrt{\sum_i (\hat{w}_i - \hat{w}_{\delta_i})^2}$$

4. Compute the error between $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}_\delta$ in the Euclidean norm.

$$\|\hat{w} - \hat{w}_\delta\| = \sqrt{\sum_i (\hat{w}_i - \hat{w}_{\delta_i})^2}$$

$$\hat{w} = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

4. Compute the error between $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}_\delta$ in the Euclidean norm.

$$\|\hat{w} - \hat{w}_\delta\| = \sqrt{\sum_i (\hat{w}_i - \hat{w}_{\delta_i})^2}$$

$$\hat{w} = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix} \qquad \hat{w}_\delta = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 + \frac{\epsilon}{3} \\ -\frac{\epsilon}{c} \end{pmatrix}$$

4. Compute the error between $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}_\delta$ in the Euclidean norm.

$$\|\hat{w} - \hat{w}_\delta\| = \sqrt{\sum_i (\hat{w}_i - \hat{w}_{\delta_i})^2}$$

$$\hat{w} = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix} \qquad \hat{w}_\delta = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 + \frac{\epsilon}{3} \\ -\frac{\epsilon}{c} \end{pmatrix}$$

$$\|\hat{w} - \hat{w}_\delta\| = \sqrt{\left(2 - \left(2 + \frac{\varepsilon}{3}\right)\right)^2 + \left(0 - \frac{\varepsilon}{c}\right)^2}$$

4. Compute the error between $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}_\delta$ in the Euclidean norm.

$$\|\hat{w} - \hat{w}_\delta\| = \sqrt{\sum_i (\hat{w}_i - \hat{w}_{\delta_i})^2}$$

$$\hat{w} = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix} \qquad \hat{w}_\delta = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 + \frac{\epsilon}{3} \\ -\frac{\epsilon}{c} \end{pmatrix}$$

$$\|\hat{w} - \hat{w}_\delta\| = \sqrt{\left(2 - \left(2 + \frac{\varepsilon}{3}\right)\right)^2 + \left(0 - \frac{\varepsilon}{c}\right)^2} = \sqrt{\frac{\varepsilon^2}{9} + \frac{\varepsilon^2}{c^2}} = \frac{\varepsilon\sqrt{9 + c^2}}{3c}$$

4. Compute the error between $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}_\delta$ in the Euclidean norm.

$$\|\hat{w} - \hat{w}_\delta\| = \sqrt{\sum_i (\hat{w}_i - \hat{w}_{\delta_i})^2}$$

$$\hat{w} = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix} \qquad \hat{w}_\delta = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 + \frac{\epsilon}{3} \\ -\frac{\epsilon}{c} \end{pmatrix}$$

$$\|\hat{w} - \hat{w}_\delta\| = \sqrt{\left(2 - \left(2 + \frac{\varepsilon}{3}\right)\right)^2 + \left(0 - \frac{\varepsilon}{c}\right)^2} = \sqrt{\frac{\varepsilon^2}{9} + \frac{\varepsilon^2}{c^2}} = \frac{\varepsilon\sqrt{9 + c^2}}{3c}$$
$$= \frac{\varepsilon}{c}\sqrt{1 + \left(\frac{c}{3}\right)^2} > \frac{\varepsilon}{c}.$$

5. How does the error compare with the data error $\delta := \|\mathbf{y} - \mathbf{y}_\delta\|$?

5. How does the error compare with the data error $\delta := \|\mathbf{y} - \mathbf{y}_\delta\|$?

$$\|\hat{w} - \hat{w}_\delta\| = \frac{\epsilon}{c}\sqrt{1 + \frac{c^2}{9}}$$

5. How does the error compare with the data error $\delta := \|\mathbf{y} - \mathbf{y}_\delta\|$?

$$\|\hat{w} - \hat{w}_\delta\| = \frac{\epsilon}{c}\sqrt{1 + \frac{c^2}{9}}$$

$$\|y - y_\delta\| = \epsilon\sqrt{3}$$

5. How does the error compare with the data error $\delta := \|\mathbf{y} - \mathbf{y}_\delta\|$?

$$\|\hat{w} - \hat{w}_\delta\| = \frac{\epsilon}{c}\sqrt{1 + \frac{c^2}{9}}$$

$$\|y - y_\delta\| = \epsilon\sqrt{3}$$

$$\|\hat{w} - \hat{w}_\delta\| \gg \| y - y_\delta\| \quad for \quad c \to 0$$

5. How does the error compare with the data error $\delta := \|\mathbf{y} - \mathbf{y}_\delta\|$?

$$\|\hat{w} - \hat{w}_\delta\| = \frac{\epsilon}{c}\sqrt{1 + \frac{c^2}{9}}$$

$$\|y - y_\delta\| = \epsilon\sqrt{3}$$

$$\|\hat{w} - \hat{w}_\delta\| \gg \| y - y_\delta\| \quad for \quad c \to 0$$

The error in reconstruction is dominated by the ratio ε/c. If $c \ll \varepsilon$ the error can get potentially very large compared to the data error $\delta = \|y - y^\delta\| = \varepsilon\sqrt{3}$, which does not depend on c. Suppose $\varepsilon = 1/100$ and $c = 1/1000$, then $\delta \approx 0.01732$ but $\varepsilon/c = 10$. Hence, the data error δ is amplified by a factor larger than 577 in the reconstruction.

for data points $(x^{(1)}, y^{(1)})$ with $x^{(1)} = 1 - c$ and $y^{(1)} = 1$, $(x^{(2)}, y^{(2)})$ with $x^{(2)} = 1 + c$ and $y^{(2)} = 1$ for some constant $c > 0$.

1. Derive the normal equation for this problem.

$$X = \begin{pmatrix} 1 & 1-c \\ 1 & 1+c \end{pmatrix} \qquad X^\top = \begin{pmatrix} 1 & 1 \\ 1-c & 1+c \end{pmatrix} \qquad y = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

for data points $(x^{(1)}, y^{(1)})$ with $x^{(1)} = 1 - c$ and $y^{(1)} = 1$, $(x^{(2)}, y^{(2)})$ with $x^{(2)} = 1 + c$ and $y^{(2)} = 1$ for some constant $c > 0$.

1. Derive the normal equation for this problem.

$$X = \begin{pmatrix} 1 & 1-c \\ 1 & 1+c \end{pmatrix} \qquad X^\top = \begin{pmatrix} 1 & 1 \\ 1-c & 1+c \end{pmatrix} \qquad y = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$X^\top X\hat{w} = \begin{pmatrix} 2 & 2 \\ 2 & 2+2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = X^\top y = \begin{pmatrix} 1 & 1 \\ 1-c & 1+c \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

for data points $(x^{(1)}, y^{(1)})$ with $x^{(1)} = 1 - c$ and $y^{(1)} = 1$, $(x^{(2)}, y^{(2)})$ with $x^{(2)} = 1 + c$ and $y^{(2)} = 1$ for some constant $c > 0$.

1. Derive the normal equation for this problem.

$$X = \begin{pmatrix} 1 & 1-c \\ 1 & 1+c \end{pmatrix} \qquad X^\top = \begin{pmatrix} 1 & 1 \\ 1-c & 1+c \end{pmatrix} \qquad y = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$X^\top X \hat{w} = \begin{pmatrix} 2 & 2 \\ 2 & 2+2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = X^\top y = \begin{pmatrix} 1 & 1 \\ 1-c & 1+c \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 2 \\ 2 & 2+2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

$$X^\top X v_i = \sigma_i^2 v_i$$

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

$$X^\top X v_i = \sigma_i^2 v_i$$

$$\det[X^\top X - \sigma_i^2 I] = 0$$

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

$$X^\top X v_i = \sigma_i^2 v_i$$

$$\det[X^\top X - \sigma_i^2 I] = 0$$

$$\det \begin{vmatrix} 2 - \sigma_i^2 & 2 \\ 2 & 2 + 2c^2 - \sigma_i^2 \end{vmatrix} = 0$$

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

$$X^\top X v_i = \sigma_i^2 v_i$$

$$\det[X^\top X - \sigma_i^2 I] = 0$$

$$\det \begin{vmatrix} 2 - \sigma_i^2 & 2 \\ 2 & 2 + 2c^2 - \sigma_i^2 \end{vmatrix} = 0$$

$$\sigma_i^4 - 2(2 + c^2)\sigma_i^2 + 4c^2 = 0$$

2. For the matrix **X** you have set up find its singular values and left-/right- singular vectors.

$$X^{\top} X v_i = \sigma_i^2 v_i$$

$$\det[X^{\top} X - \sigma_i^2 I] = 0$$

$$\det \begin{vmatrix} 2 - \sigma_i^2 & 2 \\ 2 & 2 + 2c^2 - \sigma_i^2 \end{vmatrix} = 0$$

$$\sigma_i^4 - 2(2 + c^2)\sigma_i^2 + 4c^2 = 0$$

$$\begin{cases} \sigma_1 = \sqrt{c^2 + 2 + \sqrt{c^4 + 4}}, \\ \sigma_2 = \sqrt{c^2 + 2 - \sqrt{c^4 + 4}} \end{cases}$$

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

2. For the matrix **X** you have set up find its singular values and left-/right- singular vectors.

$$X^\top X v_i = \sigma_i^2 v_i$$

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

$$X^\top X v_i = \sigma_i^2 v_i$$

$$\begin{pmatrix} 2 & 2 \\ 2 & 2+2c^2 \end{pmatrix} \begin{pmatrix} v_1^{(j)} \\ v_2^{(j)} \end{pmatrix} = \sigma_j^2 \begin{pmatrix} v_1^{(j)} \\ v_2^{(j)} \end{pmatrix}$$

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

$$X^\top X v_i = \sigma_i^2 v_i$$

$$\begin{pmatrix} 2 & 2 \\ 2 & 2+2c^2 \end{pmatrix} \begin{pmatrix} v_1^{(j)} \\ v_2^{(j)} \end{pmatrix} = \sigma_j^2 \begin{pmatrix} v_1^{(j)} \\ v_2^{(j)} \end{pmatrix}$$

$$2v_1^{(j)} + 2v_2^{(j)} = \sigma_j^2 v_1^{(j)}$$

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

$$X^\top X v_i = \sigma_i^2 v_i$$

$$\begin{pmatrix} 2 & 2 \\ 2 & 2+2c^2 \end{pmatrix} \begin{pmatrix} v_1^{(j)} \\ v_2^{(j)} \end{pmatrix} = \sigma_j^2 \begin{pmatrix} v_1^{(j)} \\ v_2^{(j)} \end{pmatrix}$$

$$2v_1^{(j)} + 2v_2^{(j)} = \sigma_j^2 v_1^{(j)}$$

$$v_2^{(j)} = \frac{\sigma_j^2 - 2}{2} v_1^{(j)}$$

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

$$v_2^{(j)} = \frac{\sigma_j^2 - 2}{2} v_1^{(j)}$$

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

$$v_2^{(j)} = \frac{\sigma_j^2 - 2}{2} v_1^{(j)}$$

$$v^{(j)} = \left(\gamma, \frac{\sigma_j^2 - 2}{2} \gamma \right)^{\top}$$

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

$$v_2^{(j)} = \frac{\sigma_j^2 - 2}{2} v_1^{(j)}$$

$$v^{(j)} = \left(\gamma, \frac{\sigma_j^2 - 2}{2}\gamma\right)^\top$$

$$\|v^{(j)}\|^2 = 1$$

2. For the matrix **X** you have set up find its singular values and left-/right- singular vectors.

$$v_2^{(j)} = \frac{\sigma_j^2 - 2}{2} v_1^{(j)}$$

$$v^{(j)} = \left(\gamma, \frac{\sigma_j^2 - 2}{2} \gamma \right)^\top$$

$$\|v^{(j)}\|^2 = 1$$

$$\|v^{(j)}\|^2 = \gamma^2 + \frac{(\sigma_j - 2)^2}{4} \gamma^2 = 1$$

2. For the matrix **X** you have set up find its singular values and left-/right- singular vectors.

$$v_2^{(j)} = \frac{\sigma_j^2 - 2}{2} v_1^{(j)}$$

$$v^{(j)} = \left(\gamma, \frac{\sigma_j^2 - 2}{2}\gamma\right)^\top$$

$$\|v^{(j)}\|^2 = 1$$

$$\|v^{(j)}\|^2 = \gamma^2 + \frac{(\sigma_j - 2)^2}{4}\gamma^2 = 1 \qquad \rightarrow \gamma^2 = \frac{4}{4 + (\sigma_j^2 - 2)^2}$$

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

$$\gamma^2 = \frac{4}{4 + (\sigma_j^2 - 2)^2}$$

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

$$\gamma^2 = \frac{4}{4 + (\sigma_j^2 - 2)^2}$$

$$v^{(j)} = \left(\gamma, \frac{\sigma_j^2 - 2}{2}\gamma\right)^\top$$

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

$$\gamma^2 = \frac{4}{4 + (\sigma_j^2 - 2)^2}$$

$$v^{(j)} = \left(\gamma, \frac{\sigma_j^2 - 2}{2}\gamma\right)^{\top}$$

$$v^{(j)} = \left(\frac{2}{\sqrt{4 + (\sigma_j^2 - 2)^2}}, \frac{\sigma_j^2 - 2}{\sqrt{4 + (\sigma_j^2 - 2)^2}}\right)^{\top}$$

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

2. For the matrix **X** you have set up find its singular values and left-/right- singular vectors.

$$u^{(j)} = \sigma_j^{-1} X v^{(j)}$$

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

$$u^{(j)} = \sigma_j^{-1} X v^{(j)}$$

$$\begin{pmatrix} u_1^{(j)} \\ u_2^{(j)} \end{pmatrix} = \sigma_j^{-1} \begin{pmatrix} 1 & 1-c \\ 1 & 1+c \end{pmatrix} \begin{pmatrix} \dfrac{2}{\sqrt{4+(\sigma_j^2-2)^2}} \\ \dfrac{\sigma_j^2-2}{\sqrt{4+(\sigma_j^2-2)^2}} \end{pmatrix}$$

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

$$u^{(j)} = \sigma_j^{-1} X v^{(j)}$$

$$\begin{pmatrix} u_1^{(j)} \\ u_2^{(j)} \end{pmatrix} = \sigma_j^{-1} \begin{pmatrix} 1 & 1-c \\ 1 & 1+c \end{pmatrix} \begin{pmatrix} \dfrac{2}{\sqrt{4+(\sigma_j^2-2)^2}} \\ \dfrac{\sigma_j^2-2}{\sqrt{4+(\sigma_j^2-2)^2}} \end{pmatrix}$$

$$\begin{pmatrix} u_1^{(j)} \\ u_2^{(j)} \end{pmatrix} = \sigma_j^{-1} \begin{pmatrix} \dfrac{\sigma_j^2(1-c)+2c}{\sqrt{4+(\sigma_j^2-2)^2}} \\ \dfrac{\sigma_j^2(1+c)-2c}{\sqrt{4+(\sigma_j^2-2)^2}} \end{pmatrix}$$

3. Solve the normal equations for your weights $\hat{\mathbf{w}} = (\hat{w}^{(0)}, \hat{w}^{(1)})^\top$.

3. Solve the normal equations for your weights $\hat{\mathbf{w}} = (\hat{w}^{(0)}, \hat{w}^{(1)})^\top$.

$$\begin{pmatrix} 2 & 2 \\ 2 & 2+2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

3. Solve the normal equations for your weights $\hat{\mathbf{w}} = (\hat{w}^{(0)}, \hat{w}^{(1)})^\top$.

$$\begin{pmatrix} 2 & 2 \\ 2 & 2+2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

$$w_0 + w_1 = 1$$

3. Solve the normal equations for your weights $\hat{\mathbf{w}} = (\hat{w}^{(0)}, \hat{w}^{(1)})^\top$.

$$\begin{pmatrix} 2 & 2 \\ 2 & 2+2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

$$w_0 + w_1 = 1$$

$$w_0 + (1 + c^2)w_1 = 1$$

3. Solve the normal equations for your weights $\hat{\mathbf{w}} = (\hat{w}^{(0)}, \hat{w}^{(1)})^\top$.

$$\begin{pmatrix} 2 & 2 \\ 2 & 2+2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

$$w_0 + w_1 = 1 \qquad\qquad w_0 = 1 - w_1$$

$$w_0 + (1 + c^2)w_1 = 1$$

3. Solve the normal equations for your weights $\hat{\mathbf{w}} = (\hat{w}^{(0)}, \hat{w}^{(1)})^\top$.

$$\begin{pmatrix} 2 & 2 \\ 2 & 2 + 2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

$$w_0 + w_1 = 1 \qquad w_0 = 1 - w_1$$

$$w_0 + (1 + c^2)w_1 = 1 \qquad 1 - w_1 + (1 + c^2)w_1 = 1$$

3. Solve the normal equations for your weights $\hat{\mathbf{w}} = (\hat{w}^{(0)}, \hat{w}^{(1)})^\top$.

$$\begin{pmatrix} 2 & 2 \\ 2 & 2+2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

$$w_0 + w_1 = 1 \qquad w_0 = 1 - w_1$$

$$w_0 + (1+c^2)w_1 = 1 \qquad 1 - w_1 + (1+c^2)w_1 = 1$$

$$\hat{w} = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

4. Repeat the previous exercise, but this time assume you make an error in your measurement. Consider two cases of the new, perturbed measurements

- $\mathbf{y}_\delta$ reads $y_\delta^{(1)} = 1 - \varepsilon$, $y_\delta^{(2)} = 1 + \varepsilon$.
- $\mathbf{y}_\delta$ reads $y_\delta^{(1)} = 1 + \varepsilon$, $y_\delta^{(2)} = 1 + \varepsilon$.

4. Repeat the previous exercise, but this time assume you make an error in your measurement. Consider two cases of the new, perturbed measurements

- $\mathbf{y}_\delta$ reads $y_\delta^{(1)} = 1 - \varepsilon$, $y_\delta^{(2)} = 1 + \varepsilon$.
- $\mathbf{y}_\delta$ reads $y_\delta^{(1)} = 1 + \varepsilon$, $y_\delta^{(2)} = 1 + \varepsilon$.

Repeating the previous exercise with the perturbed data $\mathbf{y}_\delta = \begin{pmatrix} 1-\varepsilon & 1+\varepsilon \end{pmatrix}^\top$ yields the normal equation

$$\begin{pmatrix} 2 & 2 \\ 2 & 2+2c^2 \end{pmatrix} \hat{\mathbf{w}}_\delta = \begin{pmatrix} 1 & 1 \\ 1-c & 1+c \end{pmatrix} \mathbf{y}_\delta$$
$$= \begin{pmatrix} 2 \\ 2+2c\varepsilon \end{pmatrix},$$

with the solution

$$\hat{\mathbf{w}}_\delta = \begin{pmatrix} 1 - \frac{\varepsilon}{c} \\ \frac{\varepsilon}{c} \end{pmatrix}.$$

4. Repeat the previous exercise, but this time assume you make an error in your measurement. Consider two cases of the new, perturbed measurements

- $\mathbf{y}_\delta$ reads $y_\delta^{(1)} = 1 - \varepsilon$, $y_\delta^{(2)} = 1 + \varepsilon$.
- $\mathbf{y}_\delta$ reads $y_\delta^{(1)} = 1 + \varepsilon$, $y_\delta^{(2)} = 1 + \varepsilon$.

4. Repeat the previous exercise, but this time assume you make an error in your measurement. Consider two cases of the new, perturbed measurements

- $\mathbf{y}_\delta$ reads $y_\delta^{(1)} = 1 - \varepsilon$, $y_\delta^{(2)} = 1 + \varepsilon$.
- $\mathbf{y}_\delta$ reads $y_\delta^{(1)} = 1 + \varepsilon$, $y_\delta^{(2)} = 1 + \varepsilon$.

For the perturbed data $\mathbf{y}_\delta = \begin{pmatrix} 1+\varepsilon & 1+\varepsilon \end{pmatrix}^\top$ the normal equation takes the form

$$\begin{pmatrix} 2 & 2 \\ 2 & 2+2c^2 \end{pmatrix} \hat{\mathbf{w}}_\delta = \begin{pmatrix} 1 & 1 \\ 1-c & 1+c \end{pmatrix} \mathbf{y}_\delta$$
$$= \begin{pmatrix} 2+2\varepsilon \\ 2+2\varepsilon \end{pmatrix},$$

with the solution

$$\hat{\mathbf{w}}_\delta = \begin{pmatrix} 1+\varepsilon \\ 0 \end{pmatrix}.$$

5. In both cases compute the error between $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}_\delta$ in the Euclidean norm and compare with the data error $\delta := \|\mathbf{y} - \mathbf{y}_\delta\|$?

5. In both cases compute the error between $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}_\delta$ in the Euclidean norm and compare with the data error $\delta := \|\mathbf{y} - \mathbf{y}_\delta\|$?

$$\hat{w} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \hat{w}_\delta = \begin{pmatrix} 1 - \frac{\epsilon}{c} \\ \frac{\epsilon}{c} \end{pmatrix}$$

5. In both cases compute the error between $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}_\delta$ in the Euclidean norm and compare with the data error $\delta := \|\mathbf{y} - \mathbf{y}_\delta\|$?

$$\hat{w} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \hat{w}_\delta = \begin{pmatrix} 1 - \frac{\epsilon}{c} \\ \frac{\epsilon}{c} \end{pmatrix} \qquad \|\hat{w} - \hat{w}_\delta\| = \sqrt{\frac{\epsilon^2}{c^2} + \frac{\epsilon^2}{c^2}} = \frac{\epsilon}{c}\sqrt{2}$$

5. In both cases compute the error between $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}_\delta$ in the Euclidean norm and compare with the data error $\delta := \|\mathbf{y} - \mathbf{y}_\delta\|$?

$$\hat{w} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \hat{w}_\delta = \begin{pmatrix} 1 - \frac{\epsilon}{c} \\ \frac{\epsilon}{c} \end{pmatrix} \qquad \|\hat{w} - \hat{w}_\delta\| = \sqrt{\frac{\epsilon^2}{c^2} + \frac{\epsilon^2}{c^2}} = \frac{\epsilon}{c}\sqrt{2}$$

$$\hat{w} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \hat{w}_\delta = \begin{pmatrix} 1 + \epsilon \\ 0 \end{pmatrix}$$

5. In both cases compute the error between $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}_\delta$ in the Euclidean norm and compare with the data error $\delta := \|\mathbf{y} - \mathbf{y}_\delta\|$?

$$\hat{w} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \hat{w}_\delta = \begin{pmatrix} 1 - \frac{\epsilon}{c} \\ \frac{\epsilon}{c} \end{pmatrix} \qquad \|\hat{w} - \hat{w}_\delta\| = \sqrt{\frac{\epsilon^2}{c^2} + \frac{\epsilon^2}{c^2}} = \frac{\epsilon}{c}\sqrt{2}$$

$$\hat{w} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \hat{w}_\delta = \begin{pmatrix} 1 + \epsilon \\ 0 \end{pmatrix} \qquad \|\hat{w} - \hat{w}_\delta\| = \sqrt{\epsilon^2} = \epsilon$$

CALCULUS

Calculus

Assume we have a function $f(\mathbf{x}) = f(x_1, x_2, \ldots, x_d)$

Where $\mathbf{x} = (x_1, x_2, \ldots, x_d)$

Calculus

Assume we have a function $f(\mathbf{x}) = f(x_1, x_2, \ldots, x_d)$

Where $\mathbf{x} = (x_1, x_2, \ldots, x_d)$

The gradient is defined as

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \frac{\partial}{\partial x_1} f(\mathbf{x}) \\ \frac{\partial}{\partial x_2} f(\mathbf{x}) \\ \vdots \\ \frac{\partial}{\partial x_d} f(\mathbf{x}) \end{pmatrix}$$

Calculus

Assume we have a function $f(\mathbf{x}) = f(x_1, x_2, \ldots, x_d)$

Where $\mathbf{x} = (x_1, x_2, \ldots, x_d)$

The gradient is defined as

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \frac{\partial}{\partial x_1} f(\mathbf{x}) \\ \frac{\partial}{\partial x_2} f(\mathbf{x}) \\ \vdots \\ \frac{\partial}{\partial x_d} f(\mathbf{x}) \end{pmatrix}$$

It is a vector that indicates the direction of fastest increase of the function

Calculus

Calculus

Example: $f(x_1, x_2) = (x_1 x_2 - y)^2$

Calculus

Example: $f(x_1, x_2) = (x_1 x_2 - y)^2$

Then
$$\frac{\partial}{\partial x_1} f(x) = 2(x_1 x_2 - y) x_2$$
$$\frac{\partial}{\partial x_2} f(x) = 2x_1 (x_1 x_2 - y)$$

Calculus

Example: $f(x_1, x_2) = (x_1 x_2 - y)^2$

Then

$$\frac{\partial}{\partial x_1} f(x) = 2(x_1 x_2 - y)x_2$$

$$\frac{\partial}{\partial x_2} f(x) = 2x_1(x_1 x_2 - y)$$

$$\Rightarrow \qquad \nabla f(x) = 2 \begin{pmatrix} x_1 x_2^2 - y x_2 \\ x_1^2 x_2 - x_1 y \end{pmatrix}$$

Calculus

Example: $f(x_1, \ldots, x_d) = \|\mathbf{x}\|^2$

Calculus

Example: $f(x_1, \ldots, x_d) = \|\mathbf{x}\|^2 = \sum_{i=1}^{d} x_i^2$

Calculus

Example: $f(x_1, \ldots, x_d) = \|\mathbf{x}\|^2 = \sum_{i=1}^{d} x_i^2$

Very simple to get the gradient!

Calculus

Example: $f(x_1, \ldots, x_d) = \|\mathbf{x}\|^2 = \sum_{i=1}^{d} x_i^2$

Very simple to get the gradient! $\frac{\partial}{\partial x_1} f(\mathbf{x}) = \frac{\partial}{\partial x_1} \sum_{i=1}^{d} x_i^2$

Calculus

Example: $f(x_1, \ldots, x_d) = \|\mathbf{x}\|^2 = \sum_{i=1}^{d} x_i^2$

Very simple to get the gradient! $\frac{\partial}{\partial x_1} f(\mathbf{x}) = \frac{\partial}{\partial x_1} \sum_{i=1}^{d} x_i^2 = \sum_{i=1}^{d} \frac{\partial}{\partial x_1} x_i^2$

Calculus

Example: $f(x_1, \ldots, x_d) = \|\mathbf{x}\|^2 = \sum_{i=1}^{d} x_i^2$

Very simple to get the gradient! $\frac{\partial}{\partial x_1} f(\mathbf{x}) = \frac{\partial}{\partial x_1} \sum_{i=1}^{d} x_i^2 = \sum_{i=1}^{d} \frac{\partial}{\partial x_1} x_i^2 = 2x_1$

Calculus

Example: $f(x_1, \ldots, x_d) = \|\mathbf{x}\|^2 = \sum_{i=1}^{d} x_i^2$

Very simple to get the gradient! $\frac{\partial}{\partial x_1} f(\mathbf{x}) = \frac{\partial}{\partial x_1} \sum_{i=1}^{d} x_i^2 = \sum_{i=1}^{d} \frac{\partial}{\partial x_1} x_i^2 = 2x_1$

Let's see the details

$$\sum_{i=1}^{d} \frac{\partial}{\partial x_1} x_i^2 = \frac{\partial}{\partial x_1} x_1^2 + \frac{\partial}{\partial x_1} x_2^2 + \ldots + \frac{\partial}{\partial x_1} x_d^2$$

Calculus

Example: $$f(x_1, \ldots, x_d) = \|\mathbf{x}\|^2 = \sum_{i=1}^{d} x_i^2$$

Very simple to get the gradient! $$\frac{\partial}{\partial x_1} f(\mathbf{x}) = \frac{\partial}{\partial x_1} \sum_{i=1}^{d} x_i^2 = \sum_{i=1}^{d} \frac{\partial}{\partial x_1} x_i^2 = 2x_1$$

Let's see the details

$$\sum_{i=1}^{d} \frac{\partial}{\partial x_1} x_i^2 = \frac{\partial}{\partial x_1} x_1^2 + \frac{\partial}{\partial x_1} x_2^2 + \ldots + \frac{\partial}{\partial x_1} x_d^2$$

Calculus

Example: $f(x_1, \ldots, x_d) = \|\mathbf{x}\|^2 = \sum_{i=1}^{d} x_i^2$

Very simple to get the gradient! $\frac{\partial}{\partial x_1} f(\mathbf{x}) = \frac{\partial}{\partial x_1} \sum_{i=1}^{d} x_i^2 = \sum_{i=1}^{d} \frac{\partial}{\partial x_1} x_i^2 = 2x_1$

Let's see the details

$$\sum_{i=1}^{d} \frac{\partial}{\partial x_1} x_i^2 = \frac{\partial}{\partial x_1} x_1^2 + \frac{\partial}{\partial x_1} x_2^2 + \cdots + \frac{\partial}{\partial x_1} x_d^2$$

Calculus

Example: $f(x_1, \ldots, x_d) = \|\mathbf{x}\|^2 = \sum_{i=1}^{d} x_i^2$

Very simple to get the gradient! $\frac{\partial}{\partial x_1} f(\mathbf{x}) = \frac{\partial}{\partial x_1} \sum_{i=1}^{d} x_i^2 = \sum_{i=1}^{d} \frac{\partial}{\partial x_1} x_i^2 = 2x_1$

Let's see the details

$$\sum_{i=1}^{d} \frac{\partial}{\partial x_1} x_i^2 = \frac{\partial}{\partial x_1} x_1^2 + \frac{\partial}{\partial x_1} x_2^2 + \quad + \frac{\partial}{\partial x_1} x_d^2$$

Calculus

Example: $f(x_1, \ldots, x_d) = \|\mathbf{x}\|^2 = \sum_{i=1}^{d} x_i^2$

Calculus

Example: $$f(x_1, \ldots, x_d) = \|\mathbf{x}\|^2 = \sum_{i=1}^{d} x_i^2$$

$$\frac{\partial}{\partial x_p} f(\mathbf{x}) = \partial_{x_p} \sum_{i=1}^{d} x_i^2 = 2x_p$$

Calculus

Example: $$f(x_1, \ldots, x_d) = \|\mathbf{x}\|^2 = \sum_{i=1}^{d} x_i^2$$

$$\frac{\partial}{\partial x_p} f(\mathbf{x}) = \partial_{x_p} \sum_{i=1}^{d} x_i^2 = 2x_p$$

$$\nabla f(\mathbf{x}) = 2\mathbf{x}$$

1. Show that the gradient ∇f is given by $\nabla f(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{u}$.

1. Show that the gradient ∇f is given by $\nabla f(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{u}$.

$$f(x) = \frac{1}{2}\langle x, Ax \rangle + \langle u, x \rangle$$

1. Show that the gradient ∇f is given by $\nabla f(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{u}$.

$$f(x) = \frac{1}{2}\langle x, Ax\rangle + \langle u, x\rangle$$

$$f(x) = \frac{1}{2}\sum_i x_i^\top \sum_j A_{ij}x_j + \sum_i u_i^\top x_i$$

1. Show that the gradient ∇f is given by $\nabla f(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{u}$.

$$f(x) = \frac{1}{2}\langle x, Ax\rangle + \langle u, x\rangle$$

$$f(x) = \frac{1}{2}\sum_i x_i^\top \sum_j A_{ij}x_j + \sum_i u_i^\top x_i$$

$$\partial_{x_p} f(x) = \frac{1}{2}\left(\partial_{x_p}\sum_i x_i^\top\right)\sum_j A_{ij}x_j + \frac{1}{2}\sum_i x_i^\top\left(\partial_{x_p}\sum_j A_{ij}x_j\right) + \partial_{x_p}\sum_i u_i^\top x_i$$

1. Show that the gradient ∇f is given by $\nabla f(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{u}$.

$$\partial_{x_p} f(x) = \frac{1}{2}\left(\partial_{x_p} \sum_i x_i^\top\right) \sum_j A_{ij} x_j + \frac{1}{2} \sum_i x_i^\top \left(\partial_{x_p} \sum_j A_{ij} x_j\right) + \partial_{x_p} \sum_i u_i^\top x_i$$

1. Show that the gradient ∇f is given by $\nabla f(\mathbf{x}) = \mathbf{Ax} + \mathbf{u}$.

$$\partial_{x_p} f(x) = \frac{1}{2}\left(\partial_{x_p} \sum_i x_i^\top\right) \sum_j A_{ij} x_j + \frac{1}{2} \sum_i x_i^\top \left(\partial_{x_p} \sum_j A_{ij} x_j\right) + \partial_{x_p} \sum_i u_i^\top x_i$$

i=p

1. Show that the gradient ∇f is given by $\nabla f(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{u}$.

$$\partial_{x_p} f(x) = \frac{1}{2}\left(\partial_{x_p} \sum_{\substack{i \\ i=p}} x_i^\top\right) \sum_j A_{ij} x_j + \frac{1}{2} \sum_i x_i^\top \left(\partial_{x_p} \sum_{\substack{j \\ j=p}} A_{ij} x_j\right) + \partial_{x_p} \sum_i u_i^\top x_i$$

1. Show that the gradient ∇f is given by $\nabla f(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{u}$.

$$\partial_{x_p} f(x) = \frac{1}{2}\left(\partial_{x_p} \sum_{i} x_i^\top\right) \sum_{j} A_{ij} x_j + \frac{1}{2} \sum_{i} x_i^\top \left(\partial_{x_p} \sum_{j} A_{ij} x_j\right) + \partial_{x_p} \sum_{i} u_i^\top x_i$$

i=p   j=p   i=p

1. Show that the gradient ∇f is given by $\nabla f(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{u}$.

$$\partial_{x_p} f(x) = \frac{1}{2}\left(\partial_{x_p} \sum_i x_i^\top\right) \sum_j A_{ij} x_j + \frac{1}{2} \sum_i x_i^\top \left(\partial_{x_p} \sum_j A_{ij} x_j\right) + \partial_{x_p} \sum_i u_i^\top x_i$$

i=p j=p i=p

$$\partial_{x_p} f(x) = \frac{1}{2} \sum_j A_{pj} x_j + \frac{1}{2} \sum_i x_i^\top A_{ip} + u_p$$

1. Show that the gradient ∇f is given by $\nabla f(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{u}$.

$$\partial_{x_p} f(x) = \frac{1}{2}\left(\partial_{x_p} \sum_i x_i^\top\right) \sum_j A_{ij} x_j + \frac{1}{2} \sum_i x_i^\top \left(\partial_{x_p} \sum_j A_{ij} x_j\right) + \partial_{x_p} \sum_i u_i^\top x_i$$

i=p j=p i=p

$$\partial_{x_p} f(x) = \frac{1}{2} \sum_j A_{pj} x_j + \frac{1}{2} \sum_i x_i^\top A_{ip} + u_p$$

$$\partial_{x_p} f(x) = \frac{1}{2} \sum_j A_{pj} x_j + \frac{1}{2} \sum_i A_{pi}^\top x_i + u_p$$

1. Show that the gradient ∇f is given by $\nabla f(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{u}$.

$$\partial_{x_p} f(x) = \frac{1}{2}\left(\partial_{x_p} \sum_i x_i^\top\right) \sum_j A_{ij} x_j + \frac{1}{2} \sum_i x_i^\top \left(\partial_{x_p} \sum_j A_{ij} x_j\right) + \partial_{x_p} \sum_i u_i^\top x_i$$

i=p j=p i=p

$$\partial_{x_p} f(x) = \frac{1}{2} \sum_j A_{pj} x_j + \frac{1}{2} \sum_i x_i^\top A_{ip} + u_p$$

$$\partial_{x_p} f(x) = \frac{1}{2} \sum_j A_{pj} x_j + \frac{1}{2} \sum_i A_{pi}^\top x_i + u_p$$

$$\partial_{x_p} f(x) = \left(\frac{1}{2}(A + A^\top)x + u\right)_p$$

Reminder

A function $f : C \to \mathbb{R}$ over a convex set C is called *convex* if

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$$

is satisfied for all $x, y \in C$ and $\lambda \in [0,1]$.

Reminder

A function $f : C \to \mathbb{R}$ over a convex set C is called *convex* if

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$$

is satisfied for all $x, y \in C$ and $\lambda \in [0,1]$.

This definition assumes any property of the function f

Reminder

A twice differentiable function $f: C \to \mathbb{R}$ over a convex set C is called *convex* if

$$\frac{d^2 f}{dx^2} \geq 0$$

is satisfied for all $x \in C$.

Reminder

A twice differentiable function $f : C \to \mathbb{R}$ over a convex set C is called *convex* if

$$\frac{d^2 f}{dx^2} \geq 0$$

is satisfied for all $x \in C$.

For a function of n variables the condition is on the Hessian which should be positive semi-definite

5. Verify that the function $g : \mathbb{R} \to \mathbb{R}_{\geq 0}$ with $g(x) := \frac{1}{2}x^2$ is convex.

5. Verify that the function $g : \mathbb{R} \to \mathbb{R}_{\geq 0}$ with $g(x) := \frac{1}{2}x^2$ is convex.

Alternatively: we know that the function is twice differentiable

5. Verify that the function $g : \mathbb{R} \to \mathbb{R}_{\geq 0}$ with $g(x) := \frac{1}{2}x^2$ is convex.

Alternatively: we know that the function is twice differentiable

$$\frac{d}{dx}\left(\frac{1}{2}x^2\right) = x$$

5. Verify that the function $g : \mathbb{R} \to \mathbb{R}_{\geq 0}$ with $g(x) := \frac{1}{2}x^2$ is convex.

Alternatively: we know that the function is twice differentiable

$$\frac{d}{dx}\left(\frac{1}{2}x^2\right) = x$$

$$\frac{d^2}{dx^2}\left(\frac{1}{2}x^2\right) = 1 \geq 0$$

(c) Discuss the convex properties of the function $f(x) = ax^2$ where $a \in \mathbb{R}$.

(c) Discuss the convex properties of the function $f(x) = ax^2$ where $a \in \mathbb{R}$.

Alternatively: we know that the function is twice differentiable

(c) Discuss the convex properties of the function $f(x) = ax^2$ where $a \in \mathbb{R}$.

Alternatively: we know that the function is twice differentiable

$$\frac{d}{dx}\left(ax^2\right) = 2ax$$

(c) Discuss the convex properties of the function $f(x) = ax^2$ where $a \in \mathbb{R}$.

Alternatively: we know that the function is twice differentiable

$$\frac{d}{dx}\left(ax^2\right) = 2ax$$

$$\frac{d^2}{dx^2}\left(ax^2\right) = 2a \geq 0 \quad for \quad a \geq 0$$

(d) Discuss the convex properties the function $f(x) = -a\log(x)$ where $a \in \mathbb{R}$.

(d) Discuss the convex properties the function $f(x) = -a\log(x)$ where $a \in \mathbb{R}$.

Alternatively: we know that the function is twice differentiable

(d) Discuss the convex properties the function $f(x) = -a\log(x)$ where $a \in \mathbb{R}$.

Alternatively: we know that the function is twice differentiable

$$\frac{d}{dx}\left(-a\log x\right) = -\frac{a}{x}$$

(d) Discuss the convex properties the function $f(x) = -a\log(x)$ where $a \in \mathbb{R}$.

Alternatively: we know that the function is twice differentiable

$$\frac{d}{dx}\left(-a\log x\right) = -\frac{a}{x}$$

$$\frac{d^2}{dx^2}\left(-a\log x\right) = \frac{a}{x^2} \geq 0 \quad for \quad a \geq 0$$

PROBABILITY & STATISTICS

Probability & statistics

Assume we have a random variable X with a finite no. of outcomes $x_1, x_2, \ldots, x_s$ and probabilities $\rho_1 = P(X = x_1), \rho_2 = P(X = x_2), \ldots, \rho_s = P(X = x_s)$. The expectation of X is defined as

$$\mathbb{E}_x[x_i] := \sum_{i=1}^{s} x_i \, \rho_i$$

Probability & statistics

Assume we have a random variable X with a finite no. of outcomes $x_1, x_2, \ldots, x_s$ and probabilities $\rho_1 = P(X = x_1), \rho_2 = P(X = x_2), \ldots, \rho_s = P(X = x_s)$. The expectation of X is defined as

$$\mathbb{E}_x[x_i] := \sum_{i=1}^{s} x_i \rho_i$$

It is simply a weighted average!

Probability & statistics

$$\mathbb{E}_x[x_i] := \sum_{i=1}^{s} x_i \, \rho_i$$

Probability & statistics

$$\mathbb{E}_x[x_i] := \sum_{i=1}^{s} x_i \rho_i$$

Example: $s = 3$, $x_1 = 1$ $(\rho_1 = 1/2)$, $x_2 = 11/10$ $(\rho_2 = 1/3)$, $x_3 = 1/2$ $(\rho_3 = 1/6)$

Probability & statistics

$$\mathbb{E}_x[x_i] := \sum_{i=1}^{s} x_i \, \rho_i$$

Example: $s = 3$, $x_1 = 1$ ($\rho_1 = 1/2$), $x_2 = 11/10$ ($\rho_2 = 1/3$), $x_3 = 1/2$ ($\rho_3 = 1/6$)

$$\implies \qquad \mathbb{E}_x[x_i] = \sum_{i=1}^{3} x_i \, \rho_i = \frac{1}{2} + \frac{11}{30} + \frac{1}{12} = \frac{19}{20} = 0.95$$

Probability & statistics

Assume we have an absolutely continuous random variable X with probability density function ρ. The expectation of X is defined as

$$\mathbb{E}_x[x] := \int_{\mathbb{R}} x\,\rho(x)\,dx$$

Probability & statistics

Probability & statistics

Example: uniform random variable X in $[a, b]$ with $\rho(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$

Probability & statistics

Example: uniform random variable X in $[a, b]$ with $\rho(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$

$$\implies \quad \mathbb{E}_x[x] = \int_{\mathbb{R}} x\,\rho(x)\,dx = \frac{1}{b-a}\int_a^b x\,dx$$

Probability & statistics

Example: uniform random variable X in $[a, b]$ with $\rho(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$

$$\implies \quad \mathbb{E}_x[x] = \int_{\mathbb{R}} x\,\rho(x)\,dx = \frac{1}{b-a}\int_a^b x\,dx = \frac{b^2 - a^2}{2(b-a)}$$

Probability & statistics

Example: uniform random variable X in $[a, b]$ with $\rho(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$

$$\implies \quad \mathbb{E}_x[x] = \int_{\mathbb{R}} x\,\rho(x)\,dx = \frac{1}{b-a}\int_a^b x\,dx = \frac{b^2 - a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)}$$

Probability & statistics

Example: uniform random variable X in $[a, b]$ with $\rho(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$

$$\implies \quad \mathbb{E}_x[x] = \int_{\mathbb{R}} x\,\rho(x)\,dx = \frac{1}{b-a}\int_a^b x\,dx = \frac{b^2-a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2}$$

Probability & statistics

The variance of a random variable X is defined as

$$\mathrm{Var}_x[x] := \mathbb{E}_x\left[\left(x - \mathbb{E}_x[x]\right)^2\right]$$

Probability & statistics

The variance of a random variable X is defined as

$$\text{Var}_x[x] := \mathbb{E}_x\left[\left(x - \mathbb{E}_x[x]\right)^2\right] = \mathbb{E}_x[x^2] - \mathbb{E}_x[x]^2$$

Probability & statistics

The variance of a random variable X is defined as

$$\text{Var}_x[x] := \mathbb{E}_x\left[\left(x - \mathbb{E}_x[x]\right)^2\right] = \mathbb{E}_x[x^2] - \mathbb{E}_x[x]^2$$

$$\text{Var}_x[x] = \mathbb{E}_x\left[x^2 - 2x\mathbb{E}_x[x] + \mathbb{E}_x[x]^2\right]$$

Probability & statistics

The variance of a random variable X is defined as

$$\text{Var}_x[x] := \mathbb{E}_x\left[\left(x - \mathbb{E}_x[x]\right)^2\right] = \mathbb{E}_x[x^2] - \mathbb{E}_x[x]^2$$

$$\text{Var}_x[x] = \mathbb{E}_x\left[x^2 - 2x\mathbb{E}_x[x] + \mathbb{E}_x[x]^2\right]$$

$$= \mathbb{E}_x[x^2] - 2\mathbb{E}_x[x]\mathbb{E}_x[x] + \mathbb{E}_x\left[\mathbb{E}_x[x]^2\right]$$

Probability & statistics

The variance of a random variable X is defined as

$$\text{Var}_x[x] := \mathbb{E}_x\left[\left(x - \mathbb{E}_x[x]\right)^2\right] = \mathbb{E}_x[x^2] - \mathbb{E}_x[x]^2$$

$$\text{Var}_x[x] = \mathbb{E}_x\left[x^2 - 2x\mathbb{E}_x[x] + \mathbb{E}_x[x]^2\right]$$

$$= \mathbb{E}_x[x^2] - 2\mathbb{E}_x[x]\mathbb{E}_x[x] + \mathbb{E}_x\left[\mathbb{E}_x[x]^2\right]$$

$$\mathbb{E}_x[x_i] := \sum_{i=1}^{s} x_i\,\rho_i$$

Probability & statistics

The variance of a random variable X is defined as

$$\text{Var}_x[x] := \mathbb{E}_x\left[\left(x - \mathbb{E}_x[x]\right)^2\right] = \mathbb{E}_x[x^2] - \mathbb{E}_x[x]^2$$

$$\text{Var}_x[x] = \mathbb{E}_x\left[x^2 - 2x\mathbb{E}_x[x] + \mathbb{E}_x[x]^2\right]$$

$$= \mathbb{E}_x[x^2] - 2\mathbb{E}_x[x]\mathbb{E}_x[x] + \mathbb{E}_x\left[\mathbb{E}_x[x]^2\right]$$

$$\mathbb{E}_x[x_i] := \sum_{i=1}^{s} x_i \rho_i$$

If x is a constant?

Probability & statistics

The variance of a random variable X is defined as

$$\text{Var}_x[x] := \mathbb{E}_x\left[\left(x - \mathbb{E}_x[x]\right)^2\right] = \mathbb{E}_x[x^2] - \mathbb{E}_x[x]^2$$

$$\text{Var}_x[x] = \mathbb{E}_x\left[x^2 - 2x\mathbb{E}_x[x] + \mathbb{E}_x[x]^2\right]$$

$$= \mathbb{E}_x[x^2] - 2\mathbb{E}_x[x]\mathbb{E}_x[x] + \mathbb{E}_x\left[\mathbb{E}_x[x]^2\right]$$

$$\mathbb{E}_x[x_i] := \sum_{i=1}^{s} x_i \rho_i$$

If x is a constant?

$$\mathbb{E}_x[c] := \sum_{i=1}^{s} c \rho_i$$

Probability & statistics

The variance of a random variable X is defined as

$$\text{Var}_x[x] := \mathbb{E}_x\left[\left(x - \mathbb{E}_x[x]\right)^2\right] = \mathbb{E}_x[x^2] - \mathbb{E}_x[x]^2$$

$$\text{Var}_x[x] = \mathbb{E}_x\left[x^2 - 2x\mathbb{E}_x[x] + \mathbb{E}_x[x]^2\right]$$

$$= \mathbb{E}_x[x^2] - 2\mathbb{E}_x[x]\mathbb{E}_x[x] + \mathbb{E}_x\left[\mathbb{E}_x[x]^2\right]$$

$$\mathbb{E}_x[x_i] := \sum_{i=1}^{s} x_i \rho_i$$

If x is a constant?

$$\mathbb{E}_x[c] := \sum_{i=1}^{s} c \rho_i$$

$$\mathbb{E}_x[c] := c \sum_{i=1}^{s} \rho_i = c$$

Probability & statistics

The variance of a random variable X is defined as

$$\text{Var}_x[x] := \mathbb{E}_x\left[\left(x - \mathbb{E}_x[x]\right)^2\right] = \mathbb{E}_x[x^2] - \mathbb{E}_x[x]^2$$

$$\text{Var}_x[x] = \mathbb{E}_x\left[x^2 - 2x\mathbb{E}_x[x] + \mathbb{E}_x[x]^2\right]$$

$$= \mathbb{E}_x[x^2] - 2\mathbb{E}_x[x]\mathbb{E}_x[x] + \mathbb{E}_x\left[\mathbb{E}_x[x]^2\right]$$

$$= \mathbb{E}_x[x^2] - 2\mathbb{E}_x[x]^2 + \mathbb{E}_x[x]^2$$

$$\mathbb{E}_x[x_i] := \sum_{i=1}^{s} x_i \rho_i$$

If x is a constant?

$$\mathbb{E}_x[c] := \sum_{i=1}^{s} c \rho_i$$

$$\mathbb{E}_x[c] := c \sum_{i=1}^{s} \rho_i = c$$

Probability & statistics

The variance of a random variable X is defined as

$$\text{Var}_x[x] := \mathbb{E}_x\left[\left(x - \mathbb{E}_x[x]\right)^2\right] = \mathbb{E}_x[x^2] - \mathbb{E}_x[x]^2$$

$$\text{Var}_x[x] = \mathbb{E}_x\left[x^2 - 2x\mathbb{E}_x[x] + \mathbb{E}_x[x]^2\right]$$

$$= \mathbb{E}_x[x^2] - 2\mathbb{E}_x[x]\mathbb{E}_x[x] + \mathbb{E}_x\left[\mathbb{E}_x[x]^2\right]$$

$$= \mathbb{E}_x[x^2] - 2\mathbb{E}_x[x]^2 + \mathbb{E}_x[x]^2$$

$$= \mathbb{E}_x[x^2] - \mathbb{E}_x[x]^2$$

$$\mathbb{E}_x[x_i] := \sum_{i=1}^{s} x_i \rho_i$$

If x is a constant?

$$\mathbb{E}_x[c] := \sum_{i=1}^{s} c \rho_i$$

$$\mathbb{E}_x[c] := c \sum_{i=1}^{s} \rho_i = c$$

Probability & statistics

The variance of a random variable X is defined as

$$\text{Var}_x[x] := \mathbb{E}_x\left[\left(x - \mathbb{E}_x[x]\right)^2\right]$$

$$= \mathbb{E}_x[x^2] - \mathbb{E}_x[x]^2$$

Its square-root

$$\sigma_x := \sqrt{\text{Var}_x[x]}$$ is known as standard deviation

Probability & statistics

Two random variables X and Y are independent if their joint PDF factors, i.e.

$$\rho(x, y) = \rho_X(x)\,\rho_Y(y)$$

Probability & statistics

Two random variables X and Y are independent if their joint PDF factors, i.e.

$$\rho(x, y) = \rho_X(x)\,\rho_Y(y)$$

An arbitrary no. of n random variables $\{X_i\}_{i=1}^{n}$ is independent if

$$\rho(x_1, \ldots, x_n) = \prod_{i=1}^{n} \rho_{X_i}(x_i)$$

Probability & statistics

Two random variables X and Y are independent if their joint PDF factors, i.e.

$$\rho(x, y) = \rho_X(x)\,\rho_Y(y)$$

An arbitrary no. of n random variables $\{X_i\}_{i=1}^{n}$ is independent if

$$\rho(x_1, \ldots, x_n) = \prod_{i=1}^{n} \rho_{X_i}(x_i)$$

The collection of random variables is independent and identically distributed (i.i.d.) if in addition we have

$$\rho_{X_1} = \rho_{X_2} = \cdots = \rho_{X_n}$$

Reminder

The expectation value is also known as the first moment of the distribution

$$\mathbb{E}[x] = \int_a^b x\rho(x)dx = \langle x \rangle$$

Reminder

The expectation value is also known as the first moment of the distribution

$$\mathbb{E}[x] = \int_a^b x\rho(x)dx = \langle x \rangle$$

And the expectation value of x^2 is known as the second moment

Reminder

The expectation value is also known as the first moment of the distribution

$$\mathbb{E}[x] = \int_a^b x\rho(x)dx = \langle x \rangle$$

And the expectation value of x^2 is known as the second moment

$$\mathbb{E}[x^2] = \int_a^b x^2\rho(x)dx = \langle x^2 \rangle$$

Reminder

PDFs are defined in such a way that

$$\int_a^b \rho(x)dx = 1$$

For a uniform (and absolutely continuous) random variable X in $[0, 1]$ compute the expectation of $f(X)$ for

$$f(x) := \begin{cases} -\log(x) & x \in [0, 1/5] \\ 0 & \text{otherwise} \end{cases},$$

For a uniform (and absolutely continuous) random variable X in $[0,1]$ compute the expectation of $f(X)$ for

$$f(x) := \begin{cases} -\log(x) & x \in [0, 1/5] \\ 0 & \text{otherwise} \end{cases},$$

First key point, what is the PDF?

For a uniform (and absolutely continuous) random variable X in $[0,1]$ compute the expectation of $f(X)$ for

$$f(x) := \begin{cases} -\log(x) & x \in [0, 1/5] \\ 0 & \text{otherwise} \end{cases},$$

First key point, what is the PDF?

The variable is UNIFORM in [0,1] hence

For a uniform (and absolutely continuous) random variable X in $[0, 1]$ compute the expectation of $f(X)$ for

$$f(x) := \begin{cases} -\log(x) & x \in [0, 1/5] \\ 0 & \text{otherwise} \end{cases},$$

First key point, what is the PDF?

The variable is UNIFORM in [0,1] hence

$$\rho(x) = 1 \in [0,1]$$

Now we have the PDF and we can compute the expectation value of f(x)

Now we have the PDF and we can compute the expectation value of f(x)

$$\mathbb{E}[x] = \int_a^b x\rho(x)dx = \langle x \rangle$$

Now we have the PDF and we can compute the expectation value of f(x)

$$\mathbb{E}[x] = \int_a^b x\rho(x)dx = \langle x \rangle$$

$$\mathbb{E}[f(x)] = \int_a^b f(x)\rho(x)dx = \langle f(x) \rangle$$

Now we have the PDF and we can compute the expectation value of f(x)

$$\mathbb{E}[x] = \int_a^b x\rho(x)dx = \langle x \rangle$$

$$\mathbb{E}[f(x)] = \int_a^b f(x)\rho(x)dx = \langle f(x) \rangle$$

In our case: $\rho(x) = 1 \in [0,1]$

Now we have the PDF and we can compute the expectation value of f(x)

$$\mathbb{E}[x] = \int_a^b x\rho(x)dx = \langle x \rangle$$

$$\mathbb{E}[f(x)] = \int_a^b f(x)\rho(x)dx = \langle f(x) \rangle$$

In our case: $\rho(x) = 1 \in [0,1]$

$$\mathbb{E}[f(x)] = \int_0^1 f(x)dx = \langle f(x) \rangle$$

Furthermore

$$f(x) := \begin{cases} -\log(x) & x \in [0, 1/5] \\ 0 & \text{otherwise} \end{cases},$$

Furthermore

$$f(x) := \begin{cases} -\log(x) & x \in [0, 1/5] \\ 0 & \text{otherwise} \end{cases},$$

$$\mathbb{E}[f(x)] = -\int_0^{\frac{1}{5}} \log(x) dx$$

Furthermore

$$f(x) := \begin{cases} -\log(x) & x \in [0, 1/5] \\ 0 & \text{otherwise} \end{cases},$$

$$\mathbb{E}[f(x)] = -\int_0^{\frac{1}{5}} \log(x)dx$$

$$= -x\log(x) + x\,\Big|_0^{1/5}$$

1. Let X be a random variable with expectation μ and variance σ^2. Show that the variance of $aX + b$, where $a, b \in \mathbb{R}$, is

$$\mathrm{Var}_x[ax + b] = a^2\sigma^2 \,.$$

1. Let X be a random variable with expectation μ and variance σ^2. Show that the variance of $aX + b$, where $a, b \in \mathbb{R}$, is

$$\mathrm{Var}_x[ax + b] = a^2\sigma^2 \, .$$

$$Var_x[x] = \mathbb{E}\left[(x - \mathbb{E}[x])^2\right]$$

1. Let X be a random variable with expectation μ and variance σ^2. Show that the variance of $aX + b$, where $a, b \in \mathbb{R}$, is

$$\mathrm{Var}_x[ax + b] = a^2\sigma^2 \,.$$

$$Var_x[x] = \mathbb{E}\left[(x - \mathbb{E}[x])^2\right] = \mathbb{E}\left[x^2\right] - \mathbb{E}^2[x]$$

1. Let X be a random variable with expectation μ and variance σ^2. Show that the variance of $aX + b$, where $a, b \in \mathbb{R}$, is

$$\mathrm{Var}_x[ax + b] = a^2\sigma^2 \,.$$

$$Var_x[x] = \mathbb{E}\left[(x - \mathbb{E}[x])^2\right] = \mathbb{E}\left[x^2\right] - \mathbb{E}^2[x]$$

$$Var_x[ax + b] = \mathbb{E}[(ax + b)^2] - \mathbb{E}[ax + b]^2$$

1. Let X be a random variable with expectation μ and variance σ^2. Show that the variance of $aX + b$, where $a, b \in \mathbb{R}$, is

$$\mathrm{Var}_x[ax + b] = a^2\sigma^2 \,.$$

$$Var_x[x] = \mathbb{E}\left[(x - \mathbb{E}[x])^2\right] = \mathbb{E}\left[x^2\right] - \mathbb{E}^2[x]$$

$$Var_x[ax + b] = \mathbb{E}[(ax + b)^2] - \mathbb{E}[ax + b]^2$$

$$= \mathbb{E}[a^2x^2 + b^2 + 2axb] - \mathbb{E}[ax + b]^2$$

1. Let X be a random variable with expectation μ and variance σ^2. Show that the variance of $aX + b$, where $a, b \in \mathbb{R}$, is

$$\mathrm{Var}_x[ax + b] = a^2\sigma^2 \,.$$

$$Var_x[x] = \mathbb{E}\left[(x - \mathbb{E}[x])^2\right] = \mathbb{E}\left[x^2\right] - \mathbb{E}^2[x]$$

$$\begin{aligned} Var_x[ax + b] &= \mathbb{E}[(ax + b)^2] - \mathbb{E}[ax + b]^2 \\ &= \mathbb{E}[a^2x^2 + b^2 + 2axb] - \mathbb{E}[ax + b]^2 \\ &= a^2\mathbb{E}[x^2] + b^2 + 2ab\mathbb{E}[x] - \mathbb{E}[ax + b]^2 \end{aligned}$$

1. Let X be a random variable with expectation μ and variance σ^2. Show that the variance of $aX + b$, where $a, b \in \mathbb{R}$, is

$$\mathrm{Var}_x[ax + b] = a^2\sigma^2 \, .$$

$$Var_x[x] = \mathbb{E}\left[(x - \mathbb{E}[x])^2\right] = \mathbb{E}\left[x^2\right] - \mathbb{E}^2[x]$$

$$\begin{aligned} Var_x[ax+b] &= \mathbb{E}[(ax+b)^2] - \mathbb{E}[ax+b]^2 \\ &= \mathbb{E}[a^2x^2 + b^2 + 2axb] - \mathbb{E}[ax+b]^2 \\ &= a^2\mathbb{E}[x^2] + b^2 + 2ab\mathbb{E}[x] - \mathbb{E}[ax+b]^2 \\ &= a^2\mathbb{E}[x^2] + b^2 + 2ab\mathbb{E}[x] - (a\mathbb{E}[x]+b)^2 \end{aligned}$$

1. Let X be a random variable with expectation μ and variance σ^2. Show that the variance of $aX+b$, where $a, b \in \mathbb{R}$, is

$$\mathrm{Var}_x[ax+b] = a^2\sigma^2 \,.$$

$$Var_x[x] = \mathbb{E}\left[(x - \mathbb{E}[x])^2\right] = \mathbb{E}\left[x^2\right] - \mathbb{E}^2[x]$$

$$\begin{aligned} Var_x[ax+b] &= \mathbb{E}[(ax+b)^2] - \mathbb{E}[ax+b]^2 \\ &= \mathbb{E}[a^2x^2 + b^2 + 2axb] - \mathbb{E}[ax+b]^2 \\ &= a^2\mathbb{E}[x^2] + b^2 + 2ab\mathbb{E}[x] - \mathbb{E}[ax+b]^2 \\ &= a^2\mathbb{E}[x^2] + b^2 + 2ab\mathbb{E}[x] - (a\mathbb{E}[x]+b)^2 \\ &= a^2\mathbb{E}[x^2] + b^2 + 2ab\mathbb{E}[x] - a^2\mathbb{E}[x]^2 - b^2 - 2ab\mathbb{E}[x] \end{aligned}$$

1. Let X be a random variable with expectation μ and variance σ^2. Show that the variance of $aX + b$, where $a, b \in \mathbb{R}$, is

$$\mathrm{Var}_x[ax + b] = a^2\sigma^2 .$$

$$Var_x[x] = \mathbb{E}\left[(x - \mathbb{E}[x])^2\right] = \mathbb{E}\left[x^2\right] - \mathbb{E}^2[x]$$

$$\begin{aligned} Var_x[ax+b] &= \mathbb{E}[(ax+b)^2] - \mathbb{E}[ax+b]^2 \\ &= \mathbb{E}[a^2x^2 + b^2 + 2axb] - \mathbb{E}[ax+b]^2 \\ &= a^2\mathbb{E}[x^2] + b^2 + 2ab\mathbb{E}[x] - \mathbb{E}[ax+b]^2 \\ &= a^2\mathbb{E}[x^2] + b^2 + 2ab\mathbb{E}[x] - (a\mathbb{E}[x]+b)^2 \\ &= a^2\mathbb{E}[x^2] + b^2 + 2ab\mathbb{E}[x] - a^2\mathbb{E}[x]^2 - b^2 - 2ab\mathbb{E}[x] \\ &= a^2\mathbb{E}[x^2] - a^2\mathbb{E}[x]^2 \end{aligned}$$

1. Let X be a random variable with expectation μ and variance σ^2. Show that the variance of $aX + b$, where $a, b \in \mathbb{R}$, is

$$\text{Var}_x[ax + b] = a^2\sigma^2 .$$

$$Var_x[x] = \mathbb{E}\left[(x - \mathbb{E}[x])^2\right] = \mathbb{E}\left[x^2\right] - \mathbb{E}^2[x]$$

$$\begin{aligned}
Var_x[ax + b] &= \mathbb{E}[(ax + b)^2] - \mathbb{E}[ax + b]^2 \\
&= \mathbb{E}[a^2x^2 + b^2 + 2axb] - \mathbb{E}[ax + b]^2 \\
&= a^2\mathbb{E}[x^2] + b^2 + 2ab\mathbb{E}[x] - \mathbb{E}[ax + b]^2 \\
&= a^2\mathbb{E}[x^2] + b^2 + 2ab\mathbb{E}[x] - (a\mathbb{E}[x] + b)^2 \\
&= a^2\mathbb{E}[x^2] + b^2 + 2ab\mathbb{E}[x] - a^2\mathbb{E}[x]^2 - b^2 - 2ab\mathbb{E}[x] \\
&= a^2\mathbb{E}[x^2] - a^2\mathbb{E}[x]^2 = a^2(\mathbb{E}[x^2] - \mathbb{E}[x]^2)
\end{aligned}$$

(a) Consider the following probability density function of a (continuous) random variable x: $p(x|\alpha) = Ax^{-\alpha}$ for $x \geqslant 1$ where $A \in \mathbb{R}$ and $\alpha \in \mathbb{R}$. Compute the value A as function of α and discuss where it is defined.

Since the PDF need to be normalized and the variable is continuous we can write

$$\int_1^\infty p(x|\alpha)dx = 1$$

or

$$A\int_1^\infty x^{-\alpha}dx = 1 \rightarrow A = \frac{1}{\int_1^\infty x^{-\alpha}dx}$$

(a) Consider the following probability density function of a (continuous) random variable x: $p(x|\alpha) = Ax^{-\alpha}$ for $x \geqslant 1$ where $A \in \mathbb{R}$ and $\alpha \in \mathbb{R}$. Compute the value A as function of α and discuss where it is defined.

Since the PDF need to be normalized and the variable is continuous we can write

$$\int_1^\infty p(x|\alpha)dx = 1$$

or

$$A\int_1^\infty x^{-\alpha}dx = 1 \rightarrow A = \frac{1}{\int_1^\infty x^{-\alpha}dx}$$

$$\int_1^\infty x^{-\alpha} = \frac{1}{1-\alpha}x^{1-\alpha}|_1^\infty$$

(a) Consider the following probability density function of a (continuous) random variable x: $p(x|\alpha) = Ax^{-\alpha}$ for $x \geqslant 1$ where $A \in \mathbb{R}$ and $\alpha \in \mathbb{R}$. Compute the value A as function of α and discuss where it is defined.

Since the PDF need to be normalized and the variable is continuous we can write

$$\int_1^\infty p(x|\alpha)dx = 1$$

or

$$A\int_1^\infty x^{-\alpha}dx = 1 \to A = \frac{1}{\int_1^\infty x^{-\alpha}dx}$$

$$\int_1^\infty x^{-\alpha} = \frac{1}{1-\alpha} x^{1-\alpha}|_1^\infty$$

Which is defined only for $a > 1$

(a) Consider the following probability density function of a (continuous) random variable x: $p(x|\alpha) = Ax^{-\alpha}$ for $x \geqslant 1$ where $A \in \mathbb{R}$ and $\alpha \in \mathbb{R}$. Compute the value A as function of α and discuss where it is defined.

Since the PDF need to be normalized and the variable is continuous we can write

$$\int_1^\infty p(x|\alpha)dx = 1$$

or

$$A\int_1^\infty x^{-\alpha}dx = 1 \rightarrow A = \frac{1}{\int_1^\infty x^{-\alpha}dx}$$

$$\int_1^\infty x^{-\alpha} = \frac{1}{1-\alpha}x^{1-\alpha}|_1^\infty$$

Which is defined only for $a > 1$

In this case $A = 1 - \alpha$

(b) Compute the expectation value $\mathbb{E}[x]$ and the second moment of the distribution $\mathbb{E}[x^2]$ as function of α and discuss where they are defined.

(b) Compute the expectation value $\mathbb{E}[x]$ and the second moment of the distribution $\mathbb{E}[x^2]$ as function of α and discuss where they are defined.

(b) By definition the expectation value is

$$\begin{aligned} \mathbb{E}[x] &= \int x p(x|\alpha) dx = (\alpha - 1) \int_1^{\infty} x^{1-\alpha} dx \\ &= (\alpha - 1) \frac{1}{2-\alpha} x^{2-\alpha} |_1^{\infty} \end{aligned} \tag{24}$$

It can be easily seen how for any $\alpha \leqslant 2$ the expectation value is divergent. For $\alpha > 2$ we get $\mathbb{E}[x] = \frac{\alpha-1}{\alpha-2}$

(b) Compute the expectation value $\mathbb{E}[x]$ and the second moment of the distribution $\mathbb{E}[x^2]$ as function of α and discuss where they are defined.

(b) By definition the expectation value is

$$\begin{aligned} \mathbb{E}[x] &= \int x p(x|\alpha) dx = (\alpha - 1) \int_1^\infty x^{1-\alpha} dx \\ &= (\alpha - 1) \frac{1}{2-\alpha} x^{2-\alpha} |_1^\infty \end{aligned} \tag{24}$$

It can be easily seen how for any $\alpha \leqslant 2$ the expectation value is divergent. For $\alpha > 2$ we get $\mathbb{E}[x] = \frac{\alpha-1}{\alpha-2}$

(b) Compute the expectation value $\mathbb{E}[x]$ and the second moment of the distribution $\mathbb{E}[x^2]$ as function of α and discuss where they are defined.

By definition the second moment of the distribution reads

$$\begin{aligned} \mathbb{E}[x^2] &= \int x^2 p(x|\alpha)dx = (\alpha - 1)\int_1^{\infty} x^{2-\alpha} dx \\ &= (\alpha - 1)\frac{1}{3-\alpha} x^{3-\alpha}|_1^{\infty} \end{aligned}$$

It can be easily seen how for any $\alpha \leqslant 3$ the second moment is divergent. For $\alpha > 3$ we get $\mathbb{E}[x^2] = \frac{\alpha-1}{\alpha-3}$