# Machine Learning with Python

## MTH786U/P 2023/24

## Week 1: Introduction and maths preliminaries

Nicola Perra, Queen Mary University of London (QMUL)

n.perra@qmul.ac.uk

# Organisation

- Lecturer: Dr Nicola Perra
- Email: n.perra@qmul.ac.uk
- Office: Maths: MB-G26
- Office hours: Tuesdays 15:00-16:00 by request

Timetable

- Lectures: Tuesday 13:00-15:00
- Laboratory sessions: Friday 10:00-11:00

All relevant information also available on QMPlus page!

# Organisation

There will be

- non-examinable weekly coursework (theory first, practice from week 4 on)
- a mid-term exam (theory questions)
- a final coding project

# Organisation

There will be

- non-examinable weekly coursework (theory first, practice from week 4 on)
- a mid-term exam (theory questions)
- a final coding project

Coding project objectives:

- Implementation of classic machine learning tasks
- Submission of well-documented Jupyter notebook script
- Submission of a written report

# Organisation

There will be

- non-examinable weekly coursework (theory first, practice from week 4 on)
- a mid-term exam (theory questions)
- a final coding project

Coding project objectives:

- Implementation of classic machine learning tasks
- Submission of well-documented Jupyter notebook script
- Submission of a written report

$$\text{Final Grade} = 0.6 \times \text{Project Grade} + 0.4 \times \text{Exam Grade}$$

# Learning outcomes

# Learning outcomes

- understand basic & advanced concepts in linear algebra, calculus, probability, statistical interference, mathematical modelling and optimisation

# Learning outcomes

- understand basic & advanced concepts in linear algebra, calculus, probability, statistical interference, mathematical modelling and optimisation
- describe the concept of supervised machine learning

# Learning outcomes

- understand basic & advanced concepts in linear algebra, calculus, probability, statistical interference, mathematical modelling and optimisation
- describe the concept of supervised machine learning
- formulate representative problems for supervised machine learning and derive mathematical models to solve them

# Learning outcomes

# Learning outcomes

- write algorithms to solve real-world supervised machine learning problems

# Learning outcomes

- write algorithms to solve real-world supervised machine learning problems
- process and interpret data

# Learning outcomes

- write algorithms to solve real-world supervised machine learning problems
- process and interpret data
- work as part of a team on the solution of real-world, supervised machine learning problems

# Learning outcomes

- write algorithms to solve real-world supervised machine learning problems
- process and interpret data
- work as part of a team on the solution of real-world, supervised machine learning problems
- write scientific reports

# Weekly procedure

# Weekly procedure

Take part to the weekly tutorials

- One to discuss previous week’s coursework, one to work on this week’s coursework

# Weekly procedure

Take part to the weekly tutorials

One to discuss previous week's coursework, one to work on this week's coursework

Take part to the weekly lectures

General introduction to various topics and  time to ask questions

# Weekly procedure

## Take part to the weekly tutorials

One to discuss previous week's coursework, one to work on this week's coursework

## Take part to the weekly lectures

General introduction to various topics and time to ask questions

## Complete the coursework ahead of the following week

To the best of your abilities; coursework is teamwork, so please engage with your peers

# Weekly procedure

Take part to the weekly tutorials

One to discuss previous week's coursework, one to work on this week's coursework

Take part to the weekly lectures

General introduction to various topics and  time to ask questions

Complete the coursework ahead of the following week

To the best of your abilities; coursework is teamwork, so please  engage with your peers

Post your questions in the student forum

# WHAT IS MACHINE LEARNING?

# What is machine learning?

# What is machine learning?

In a nutshell:

ML is about building data-driven models that can make predictions or decisions without being explicitly programmed to do so

# What is machine learning?

In a nutshell:

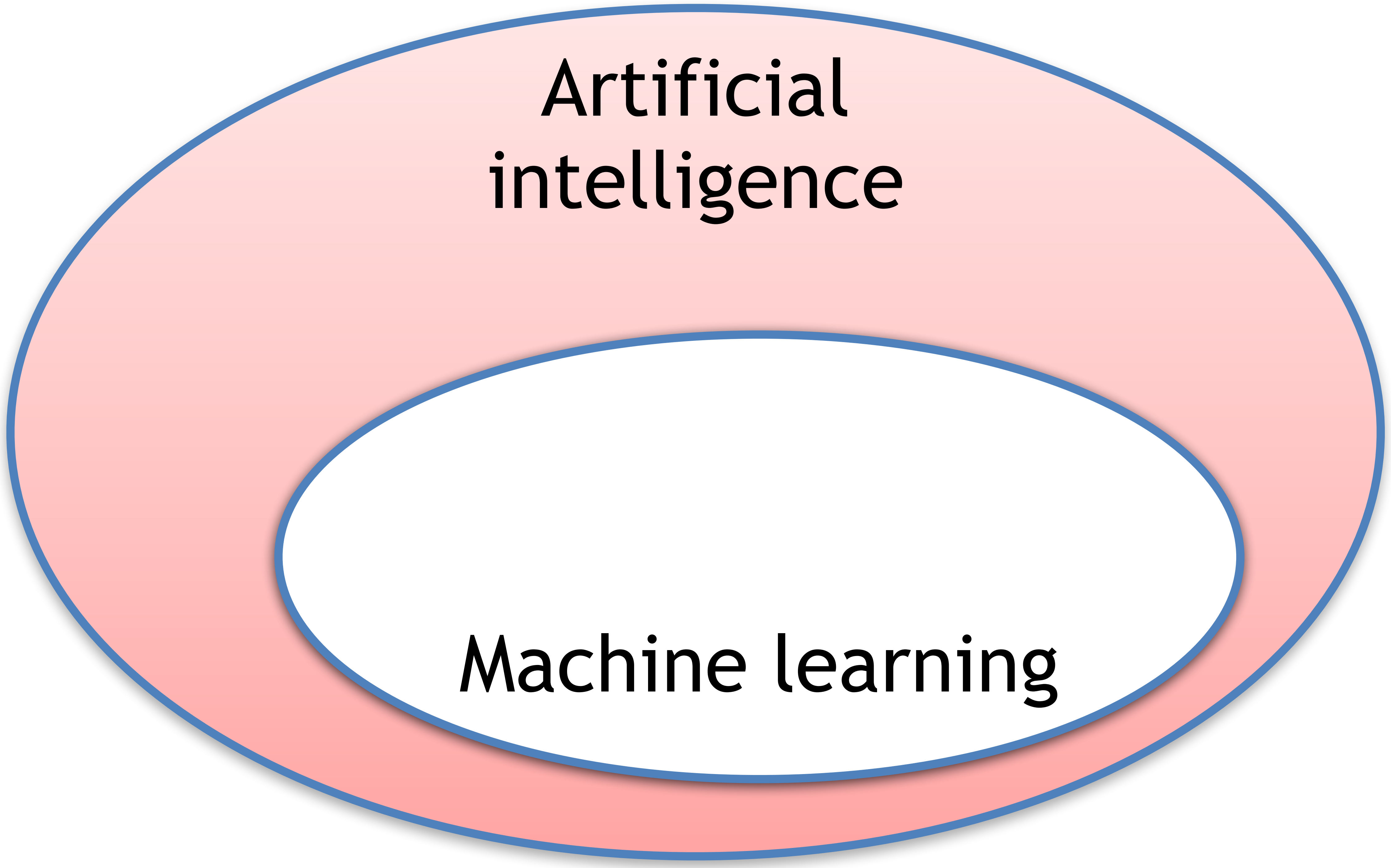
ML is about building data-driven models that can make predictions or decisions without being explicitly programmed to do so

Is it the same as artificial intelligence?

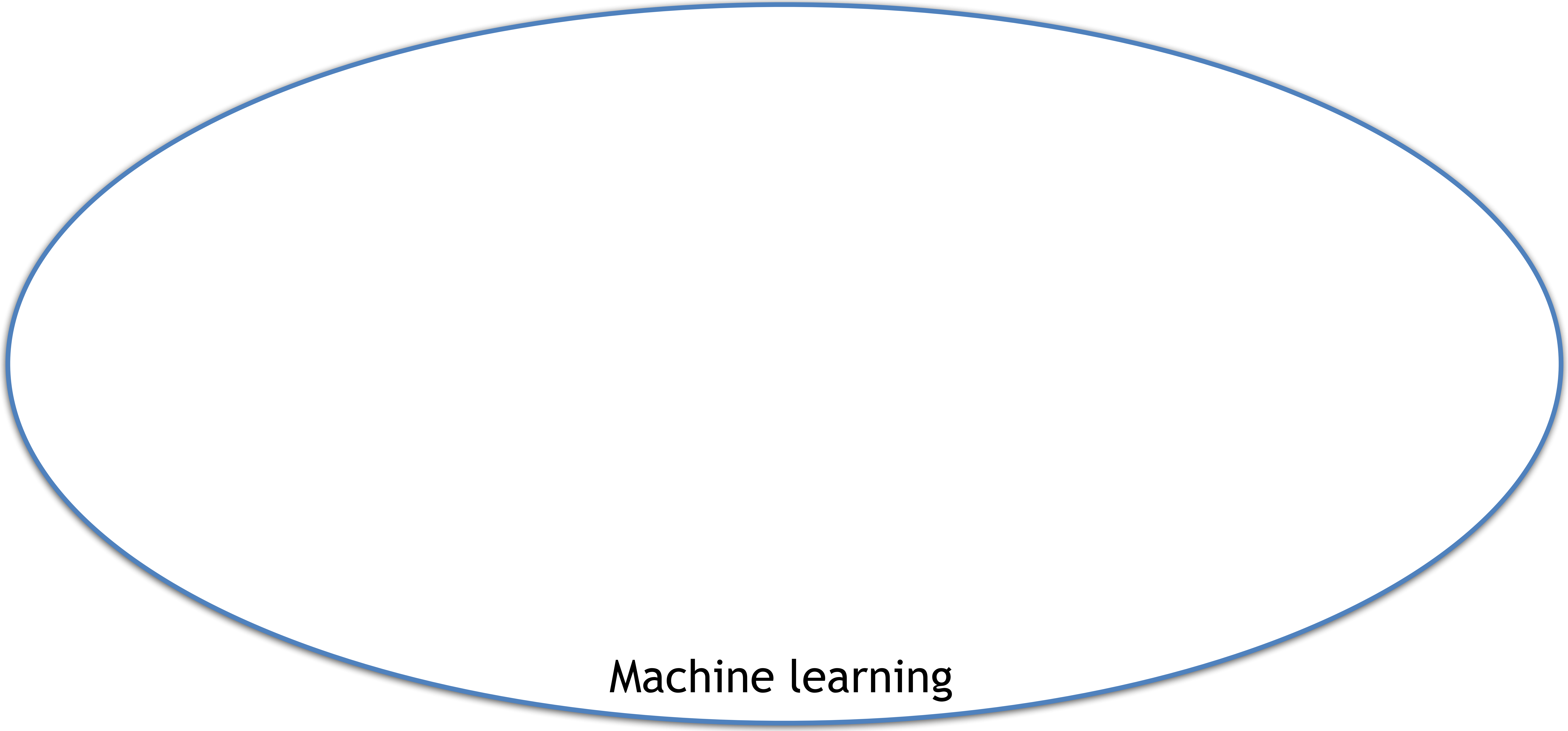
Machine learning $\neq$ artificial intelligence...

Machine learning $\neq$ artificial intelligence...

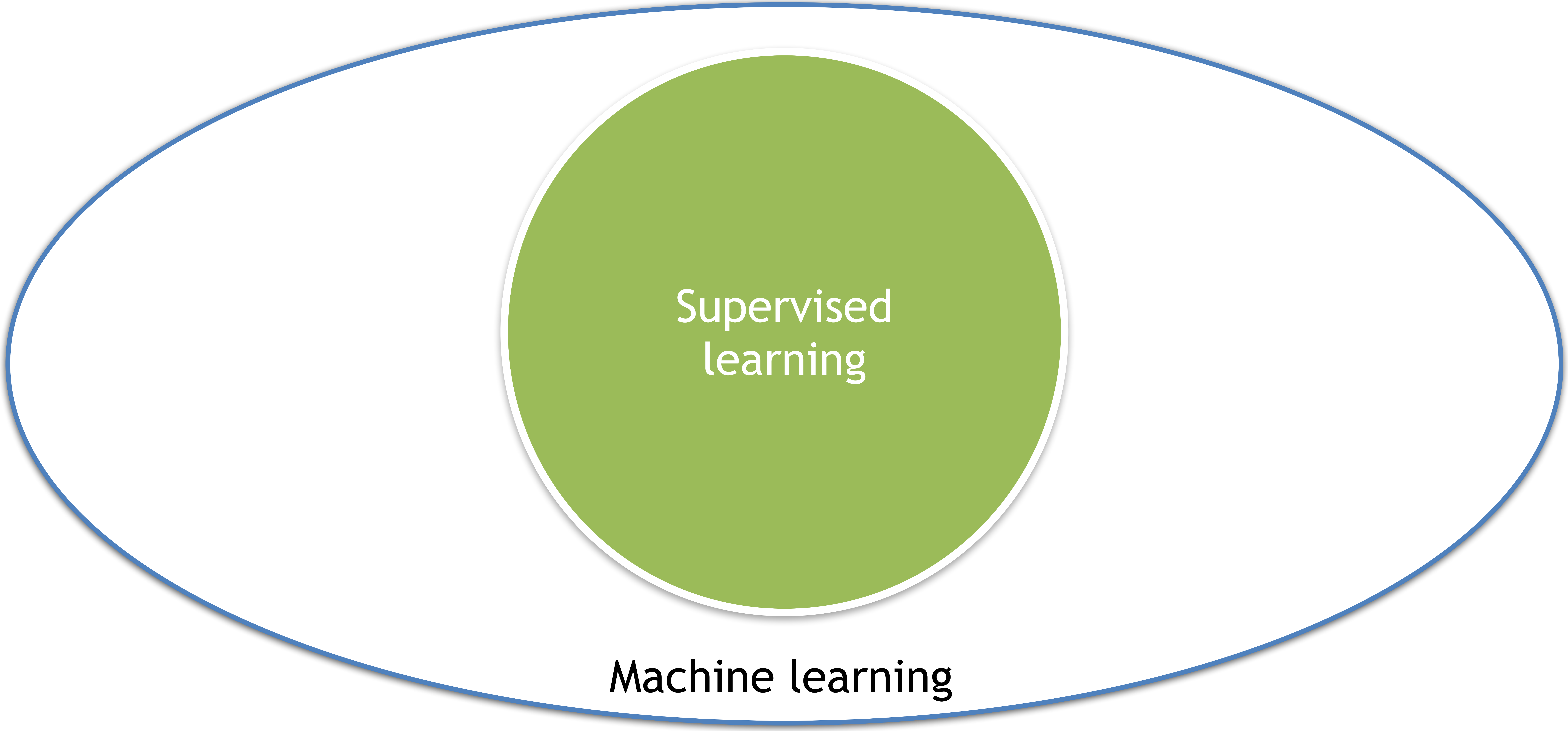
...but machine learning $\subsetneq$ artificial intelligence



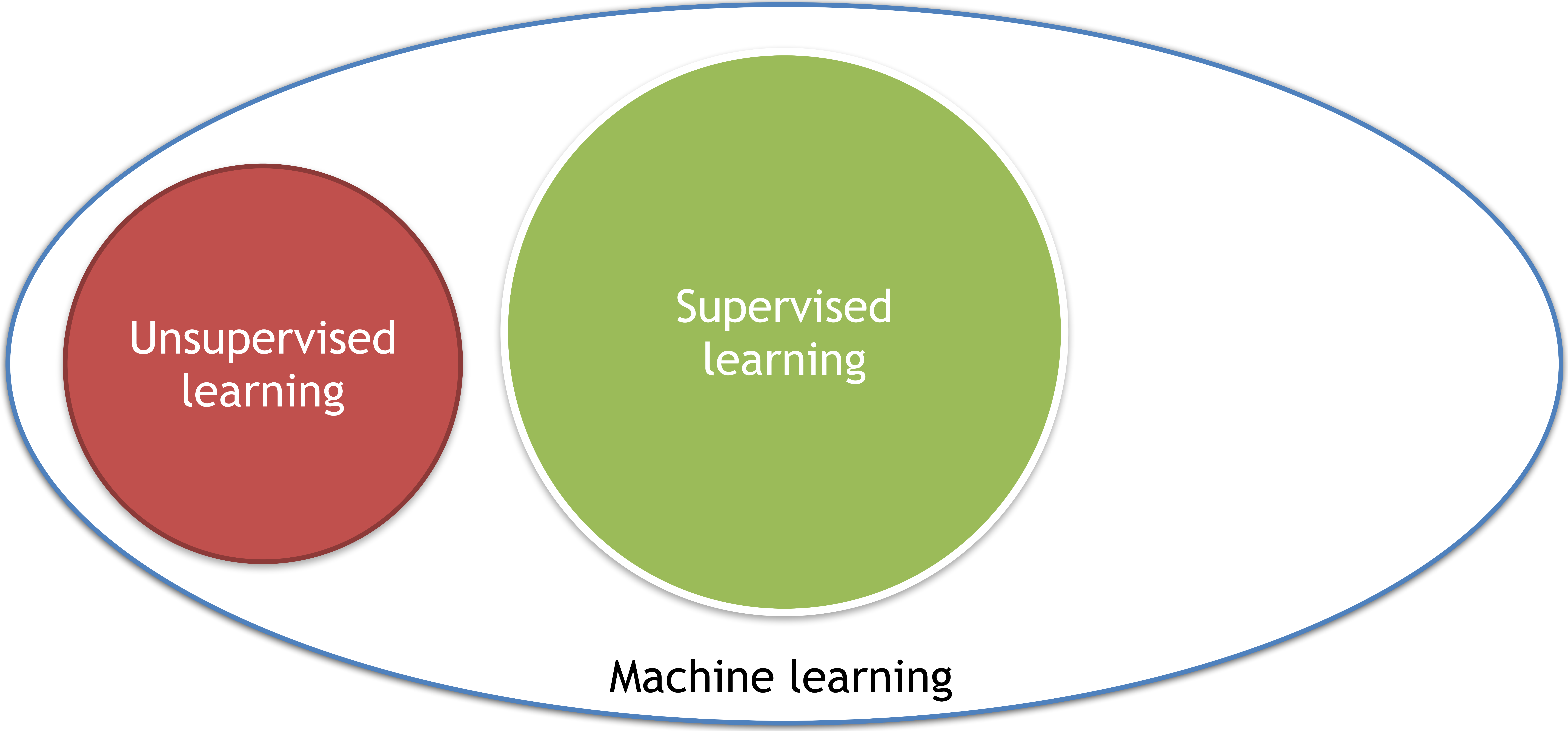
# What is machine learning?



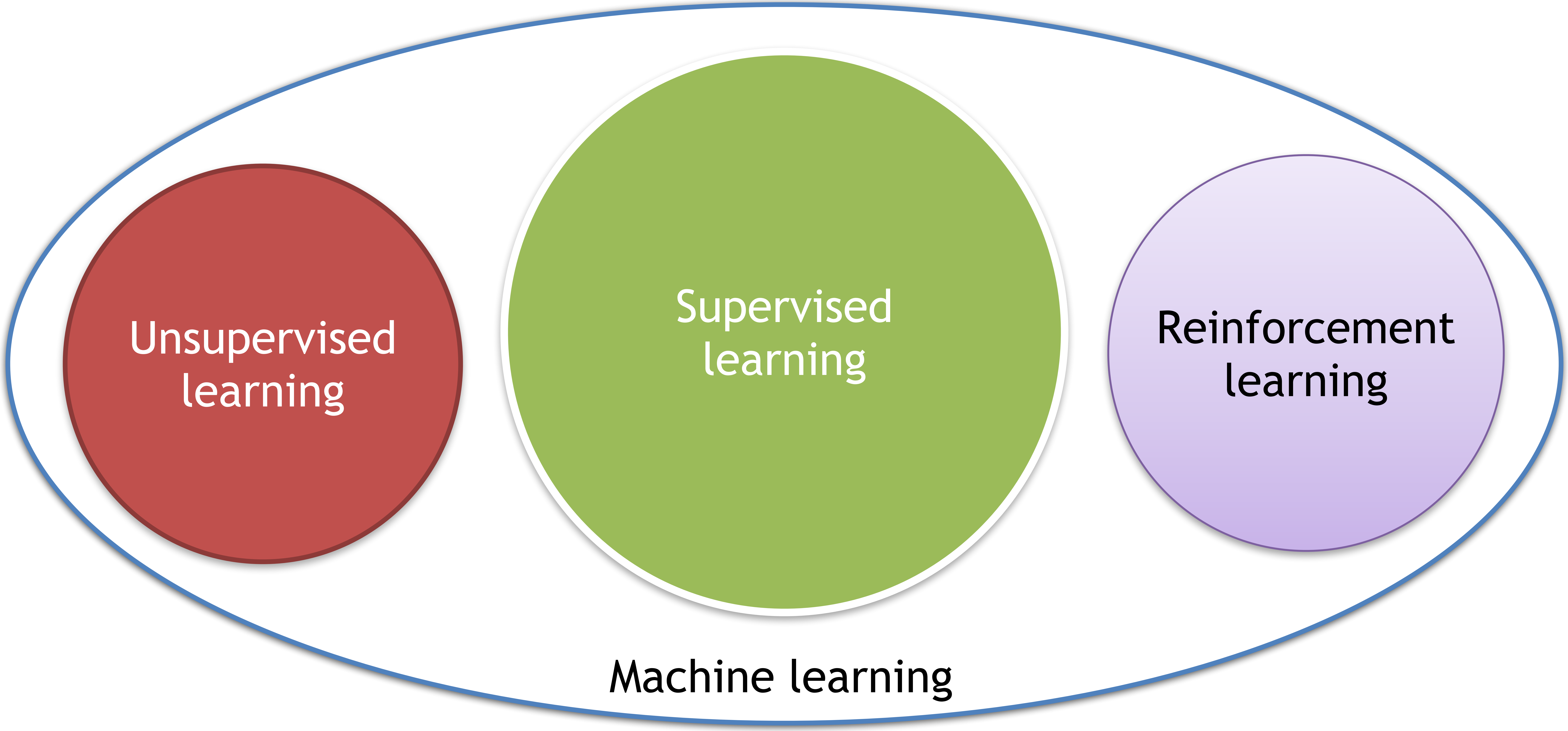
# What is machine learning?



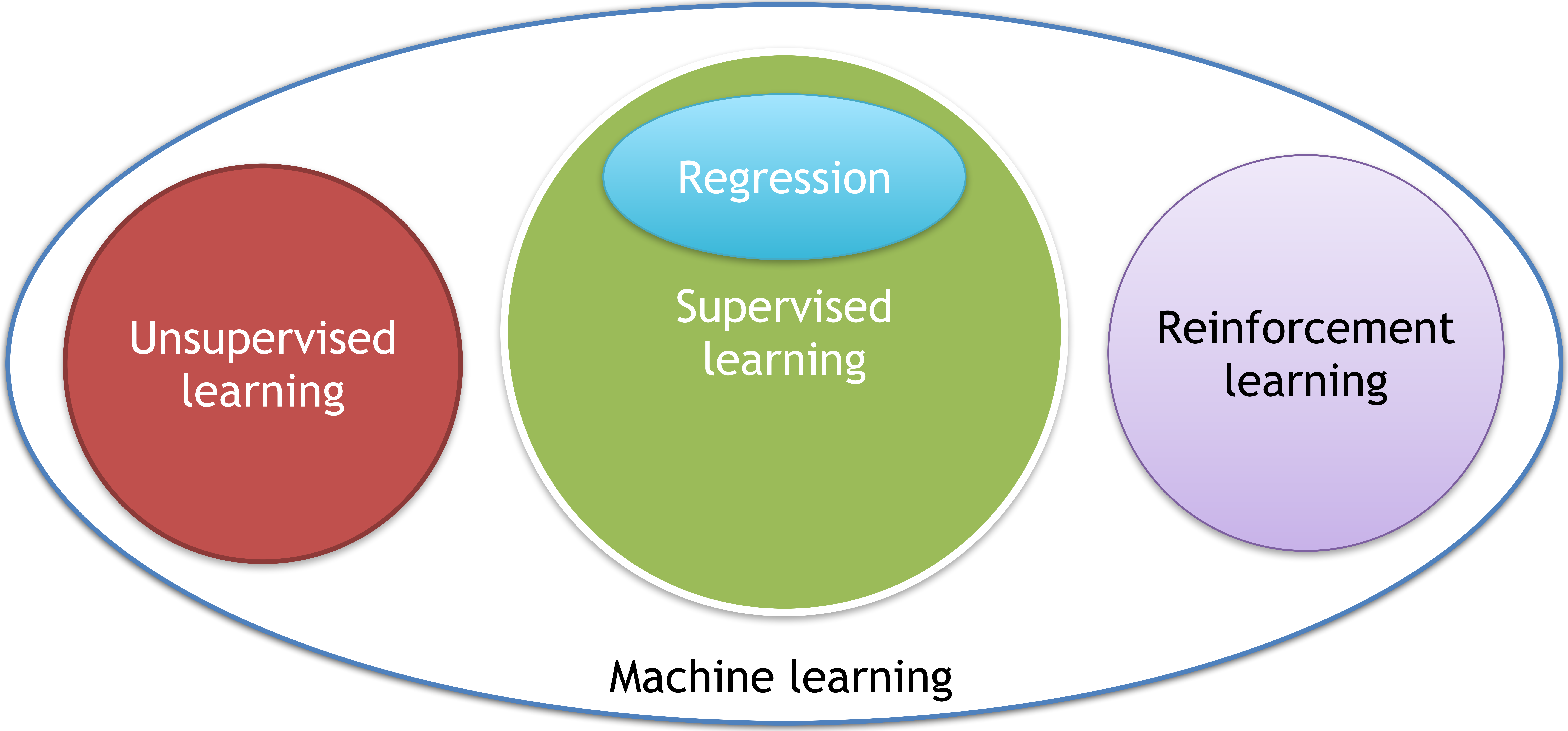
# What is machine learning?



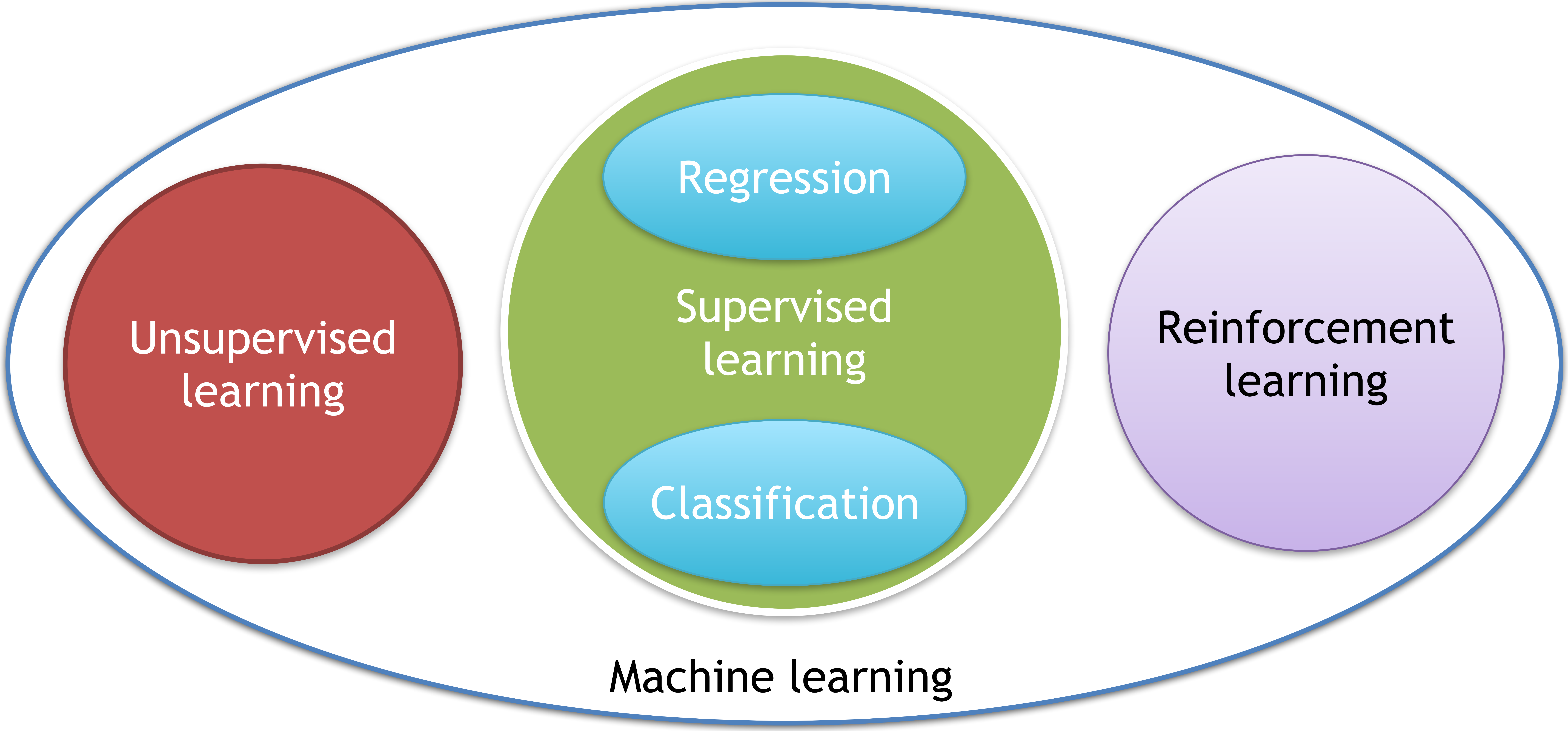
# What is machine learning?



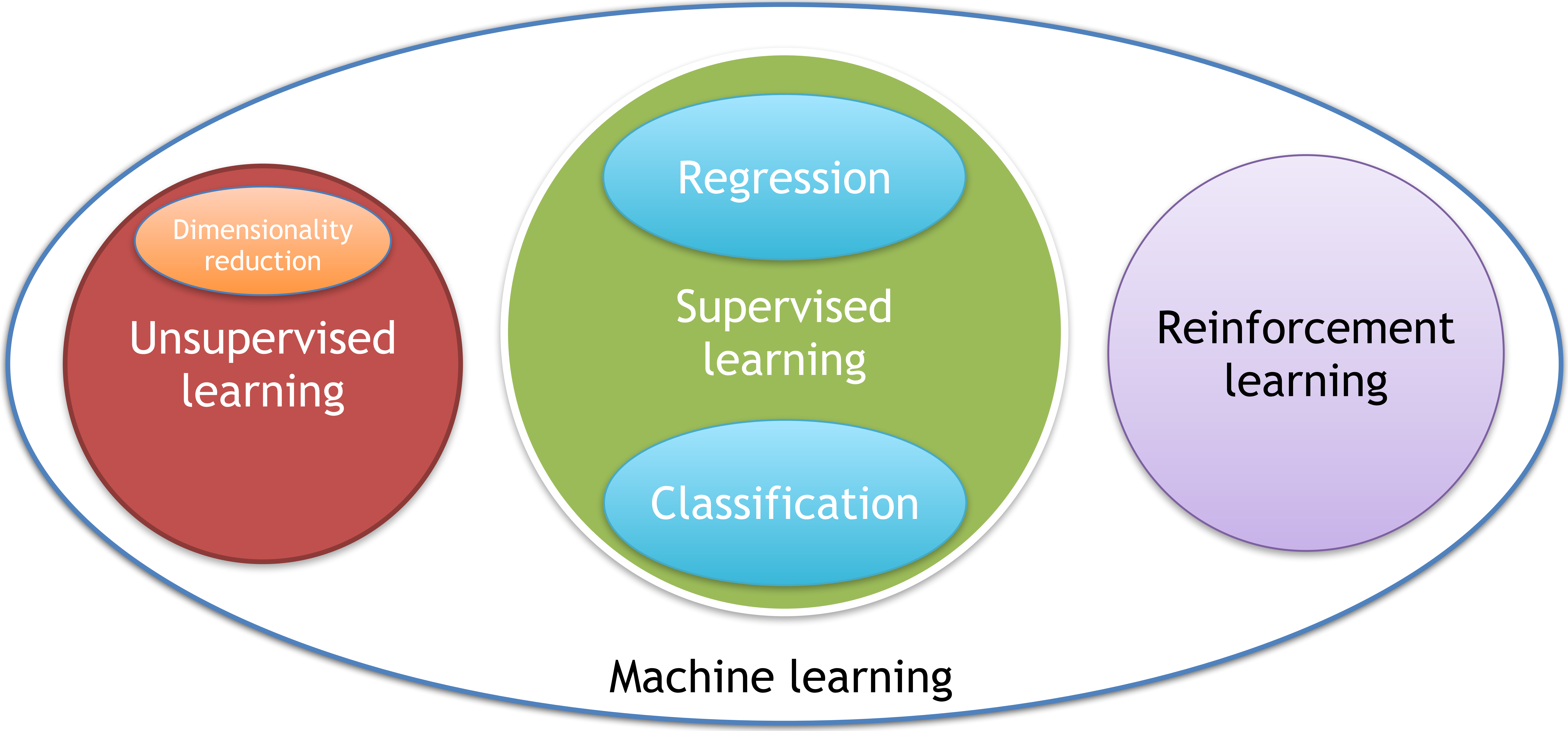
# What is machine learning?



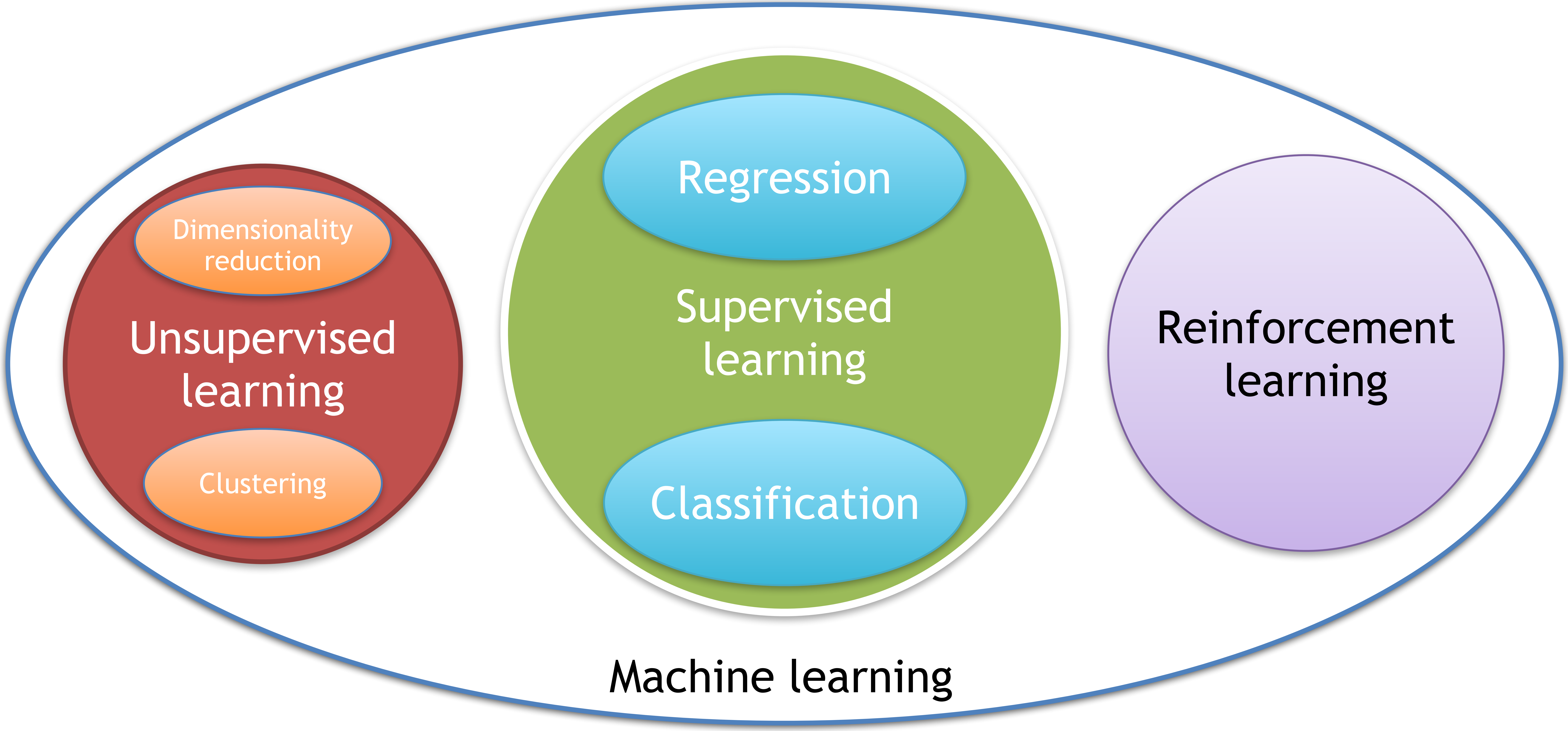
# What is machine learning?



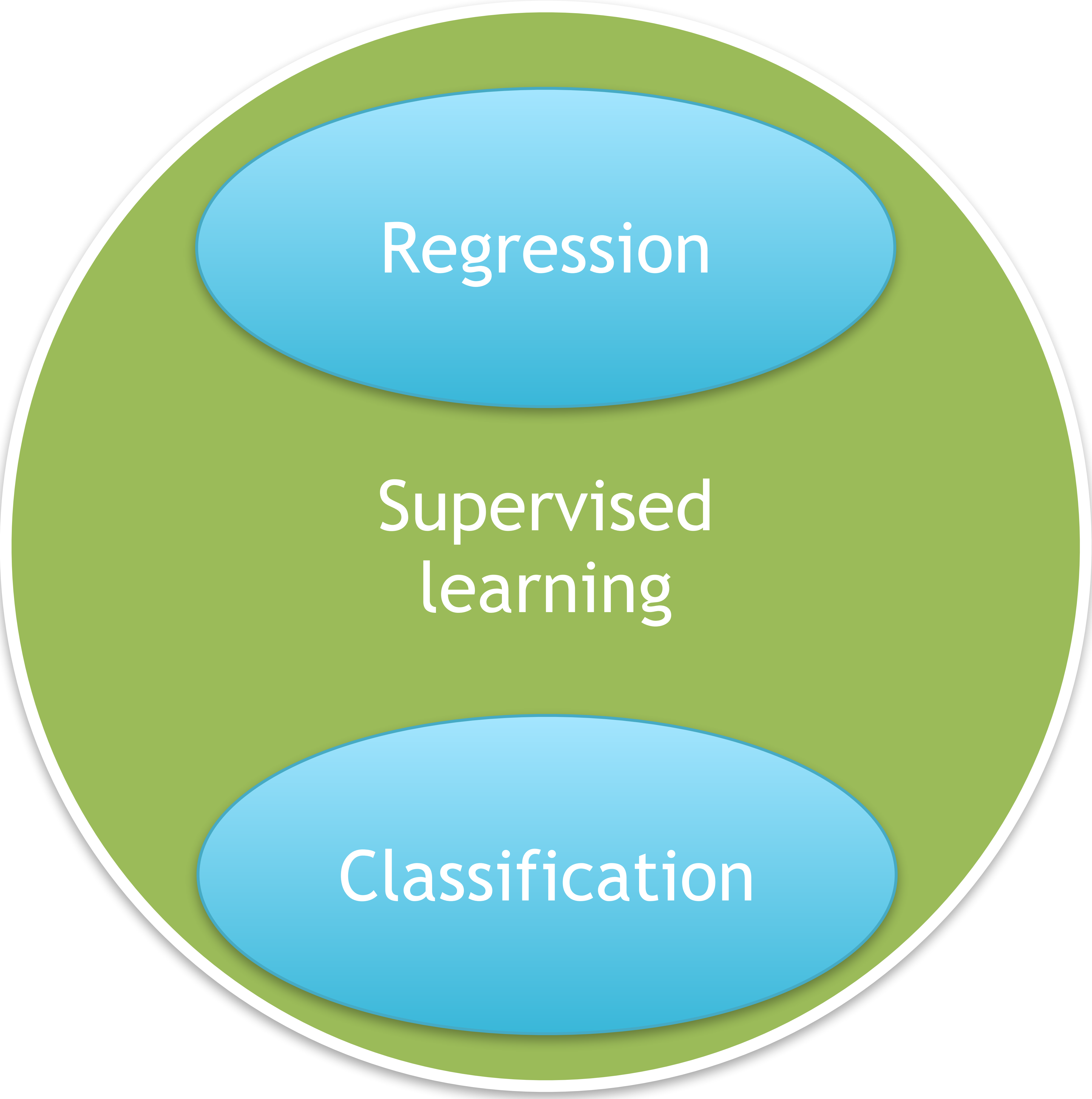
# What is machine learning?



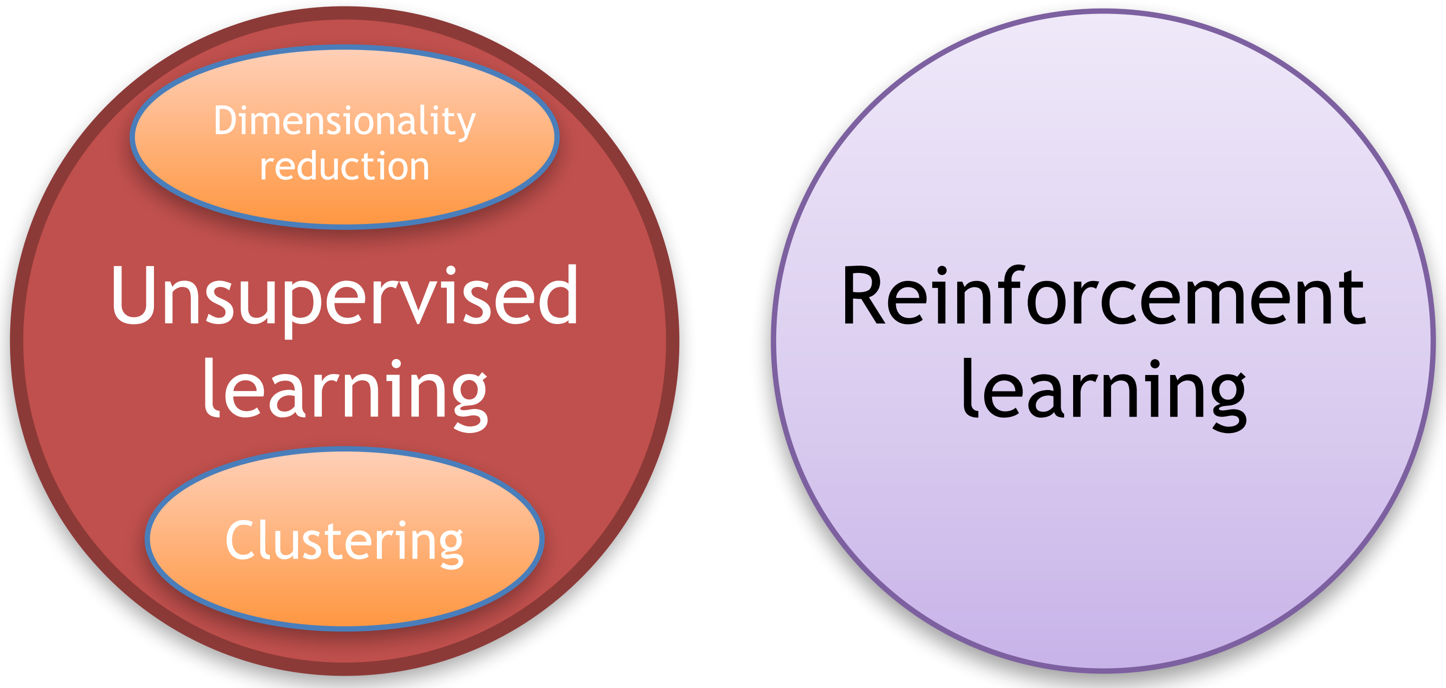
# What is machine learning?



# What is machine learning?




In this module, we solely focus on supervised machine learning &




will be covered in MTH793P

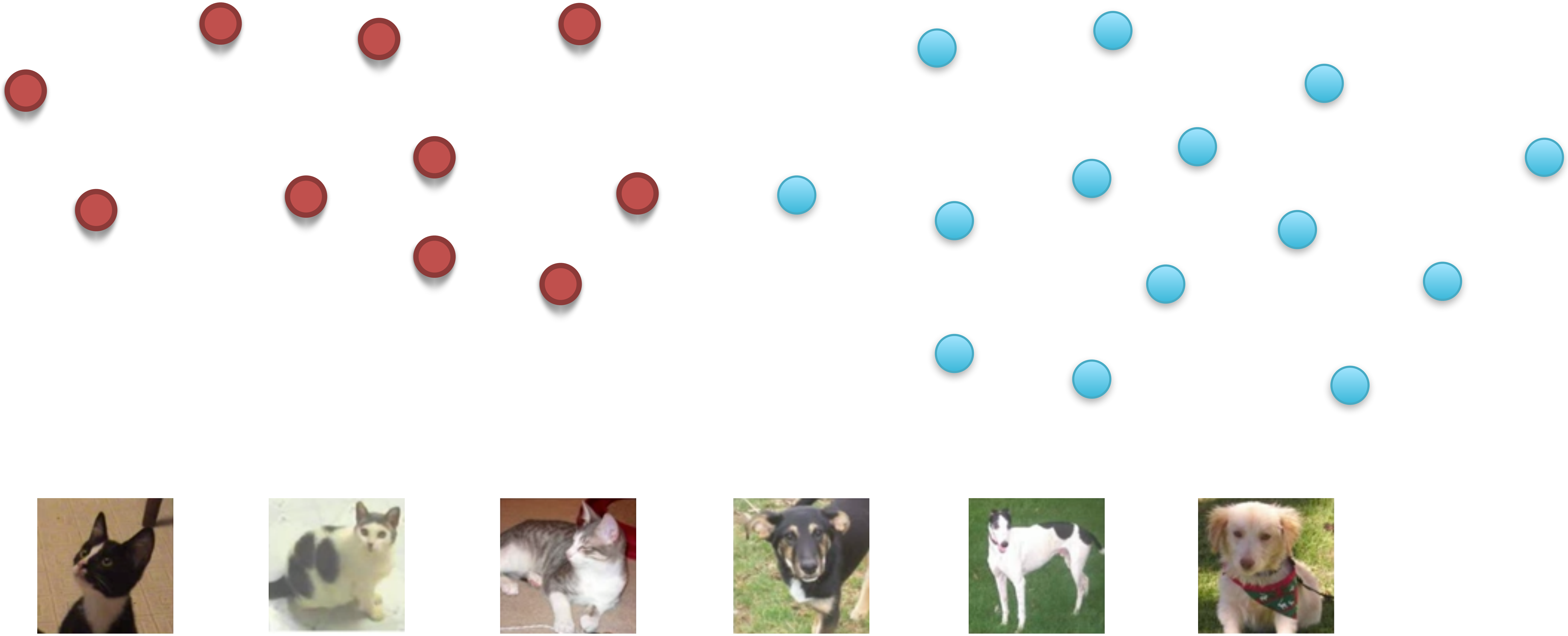
n.perra@qmul.ac.uk

# What is supervised machine learning?

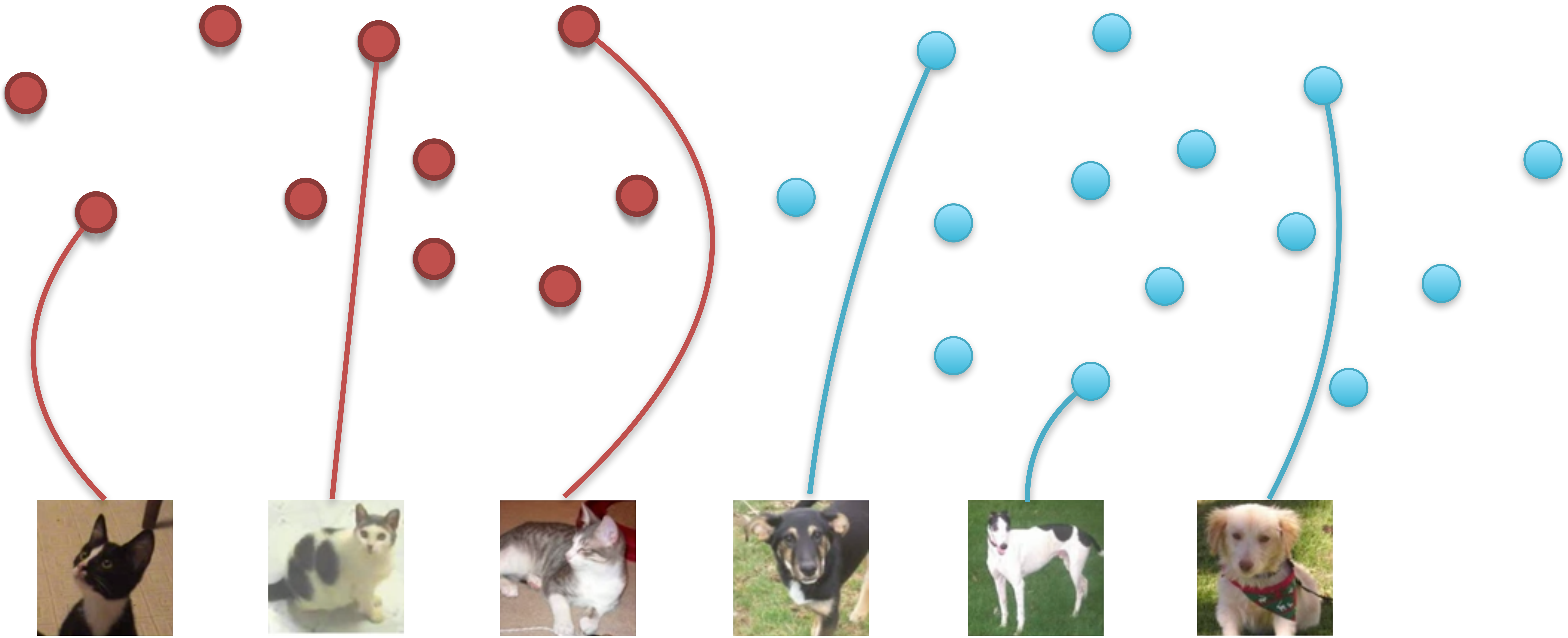
Classification

Example:

# What is supervised machine learning?

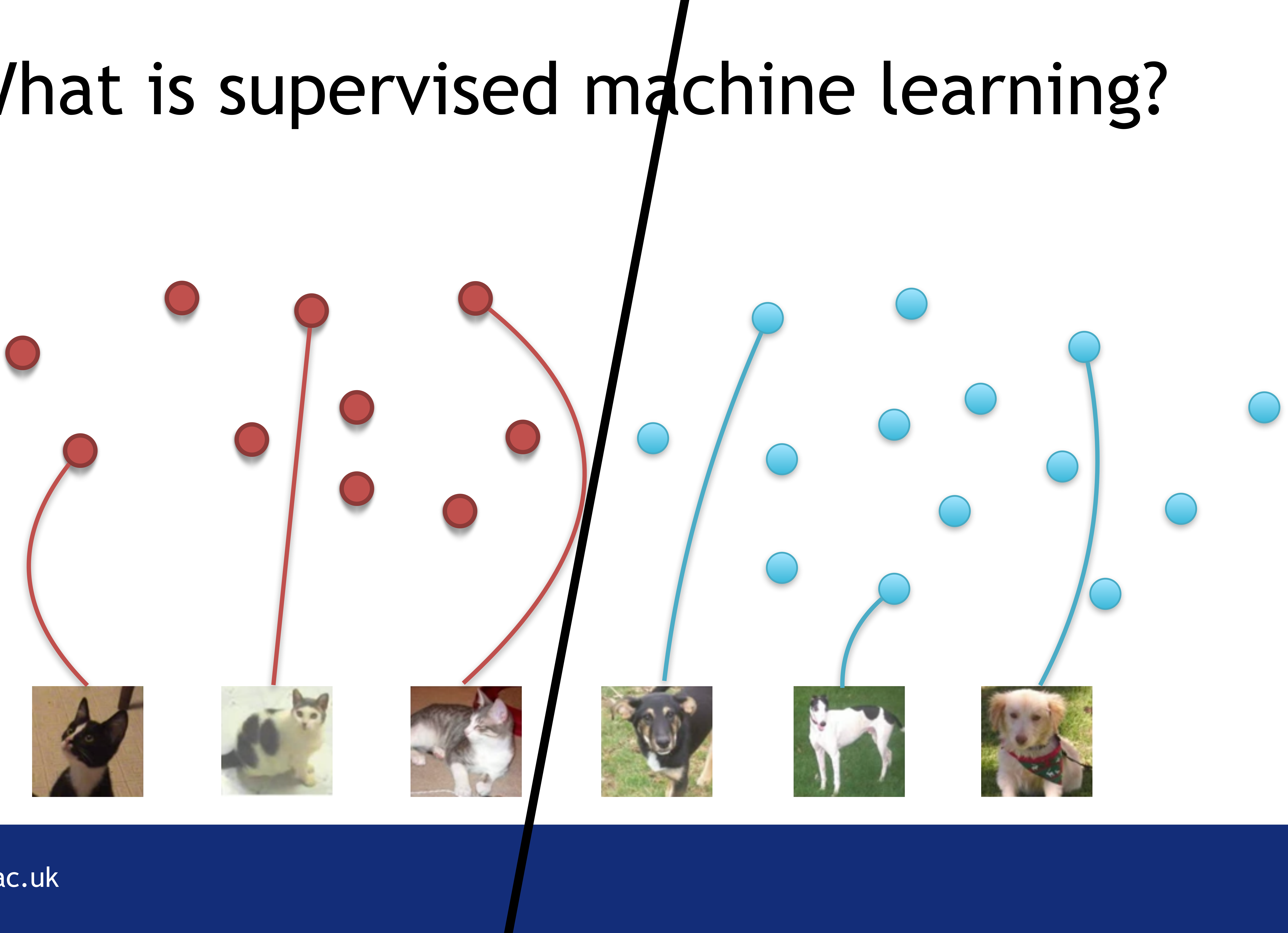


# What is supervised machine learning?

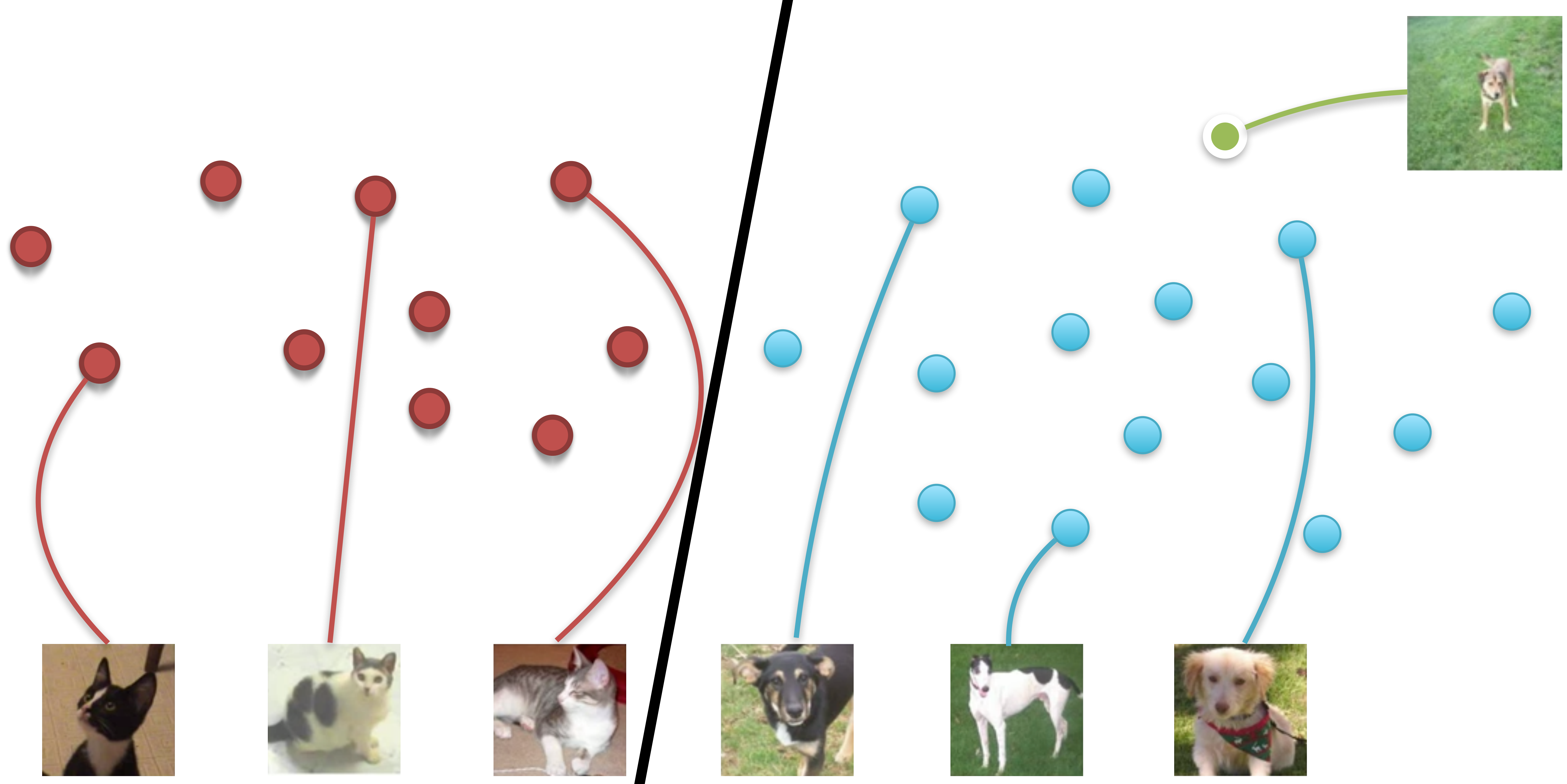


# What is supervised machine learning?

# What is supervised machine learning?



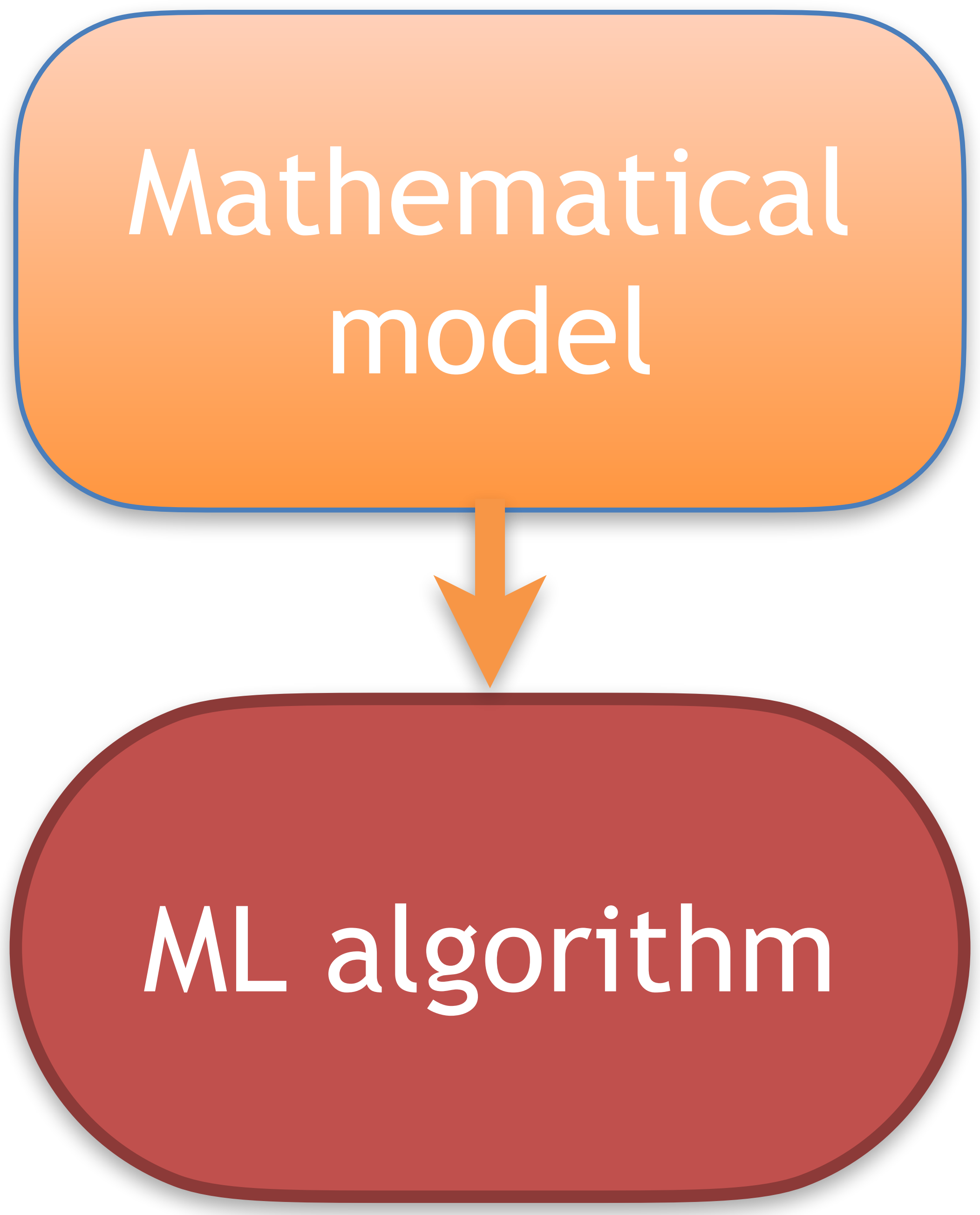
# What is supervised machine learning?



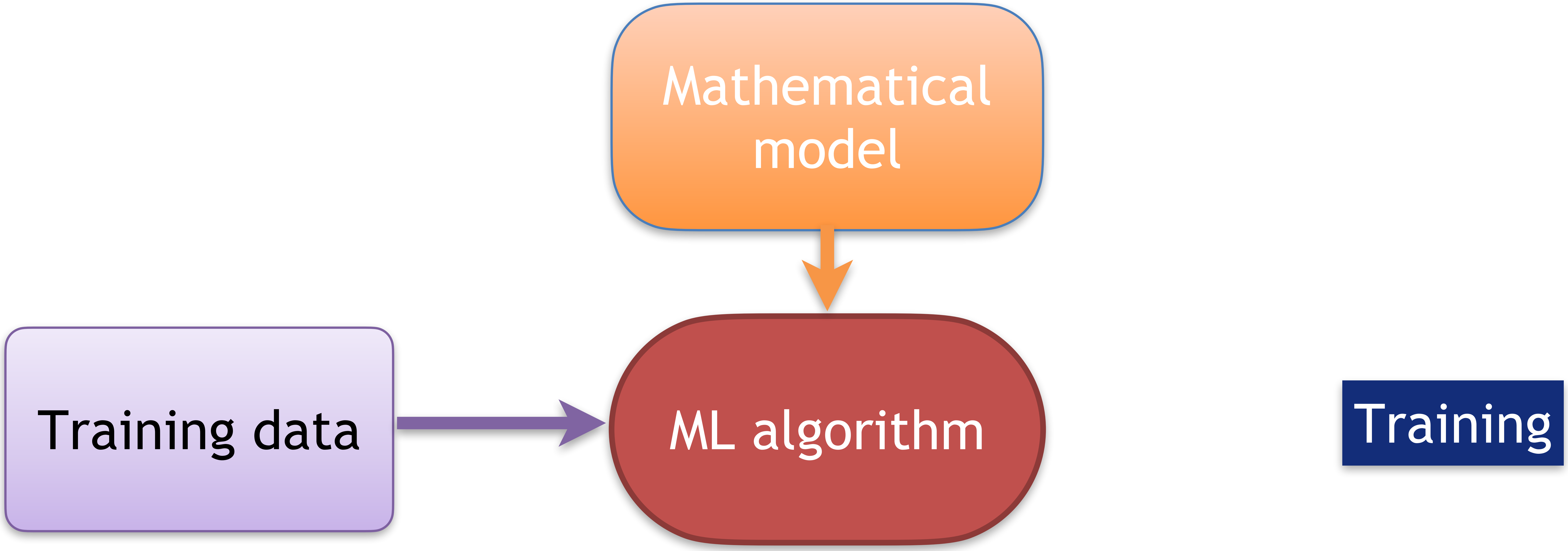
# What is supervised machine learning?

Mathematical model

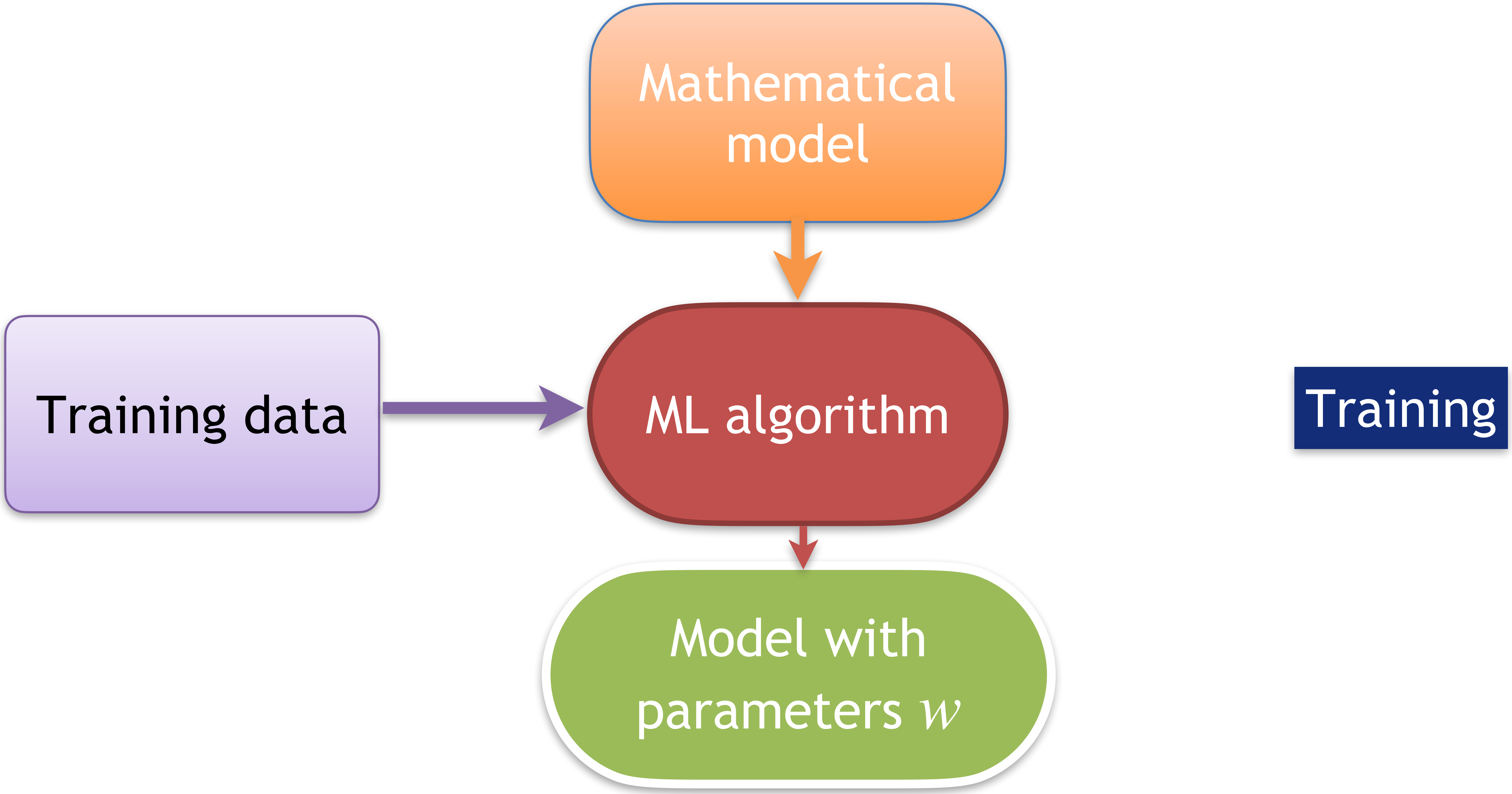
# What is supervised machine learning?



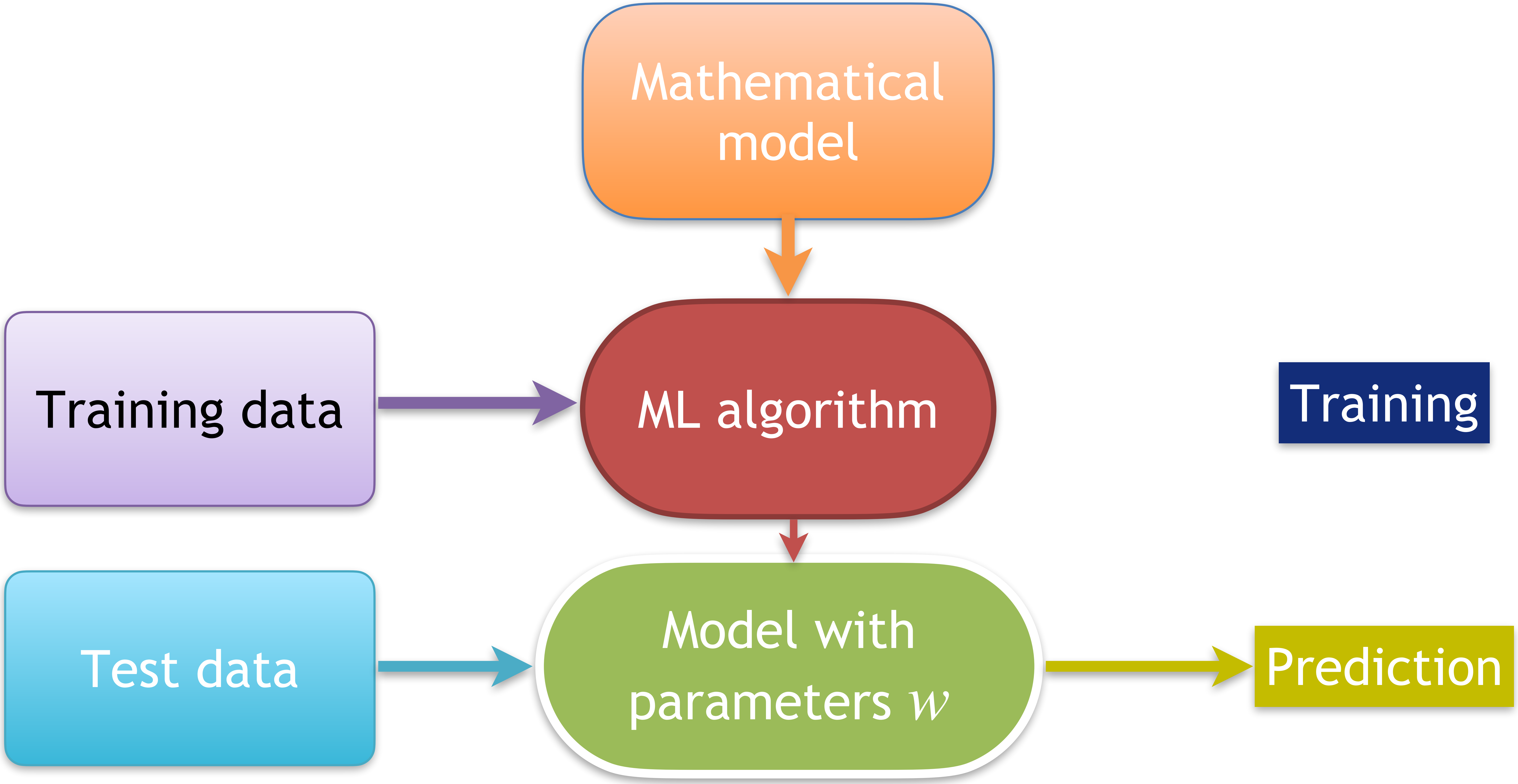
# What is supervised machine learning?



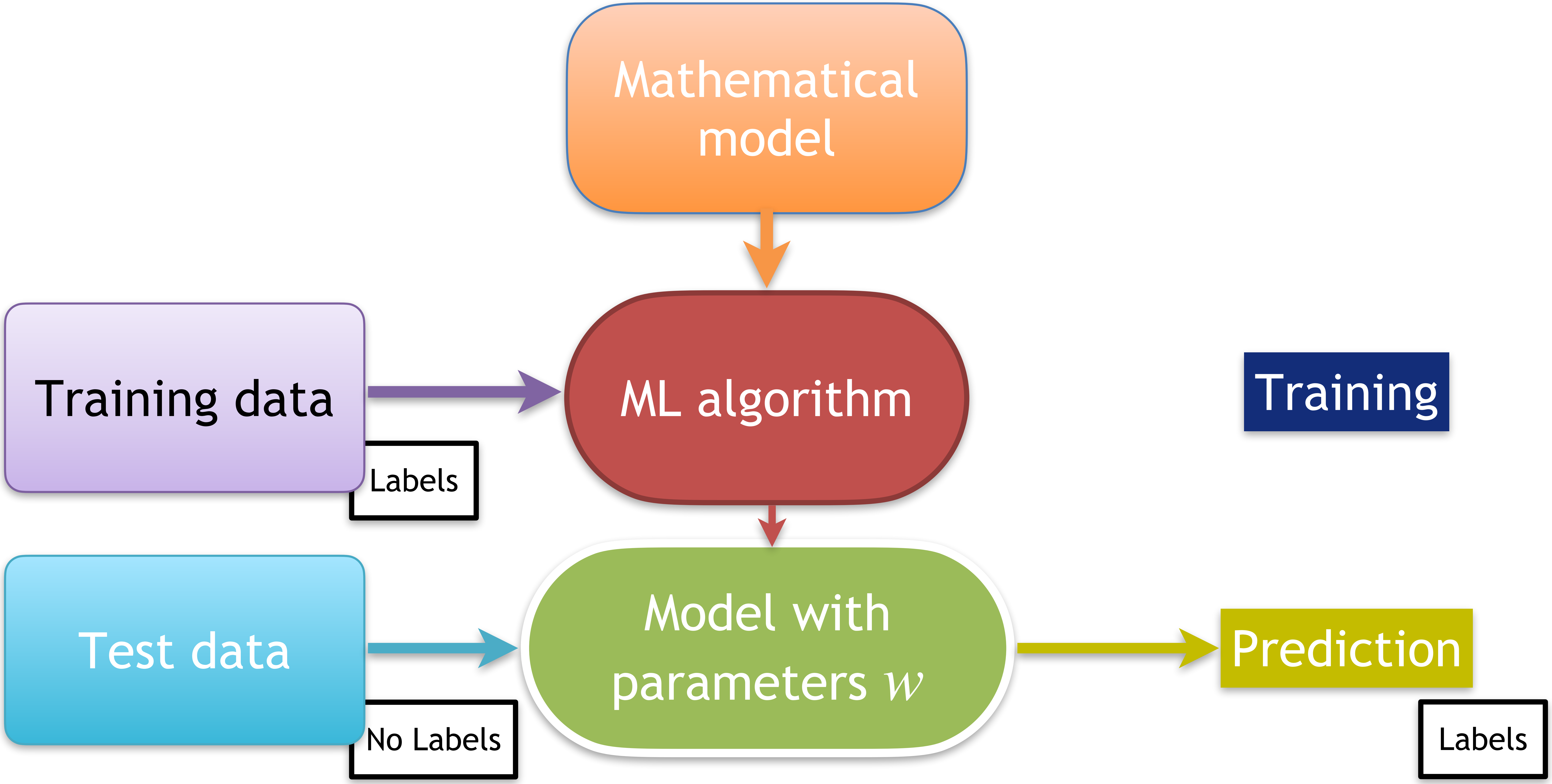
# What is supervised machine learning?



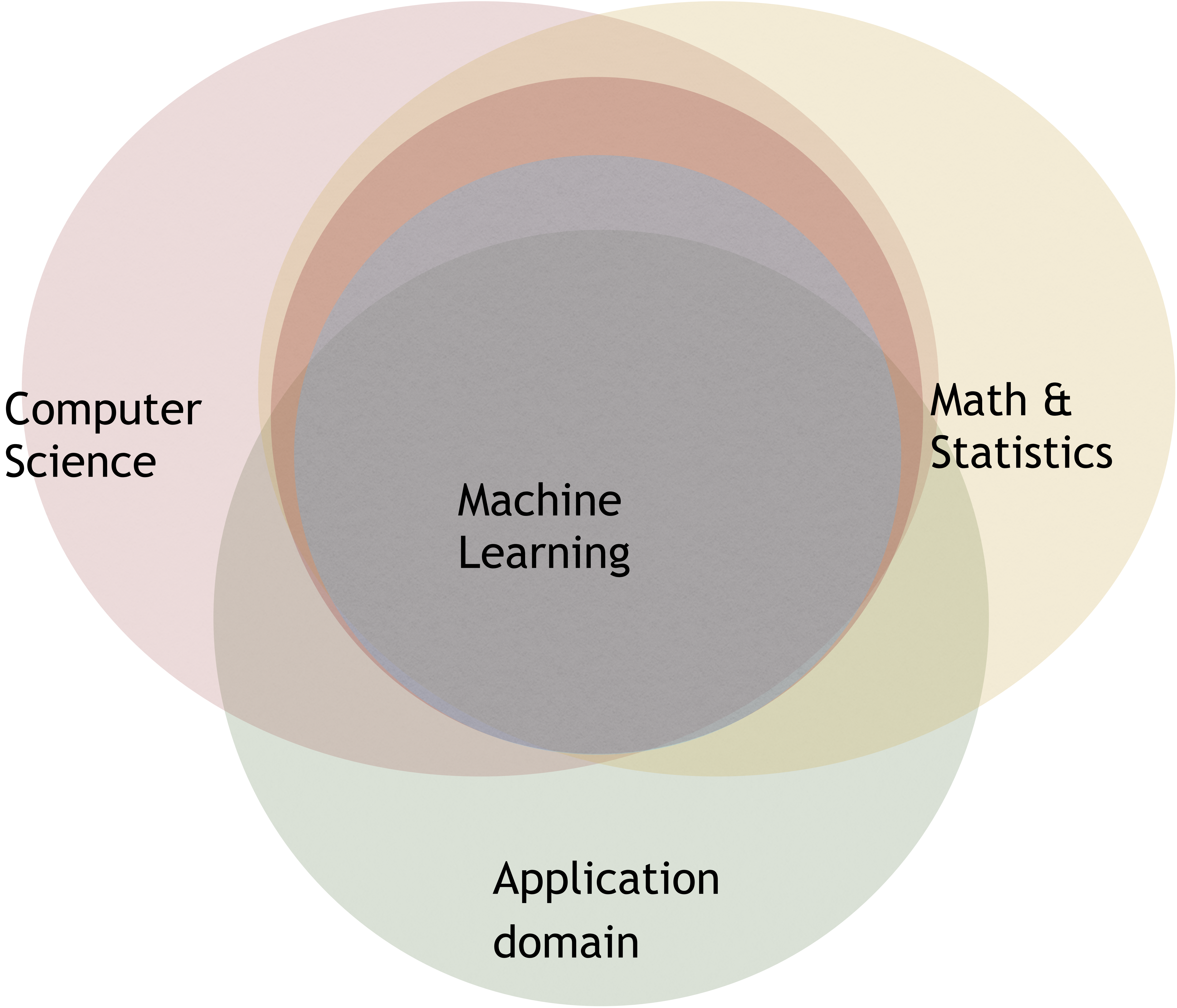
# What is supervised machine learning?



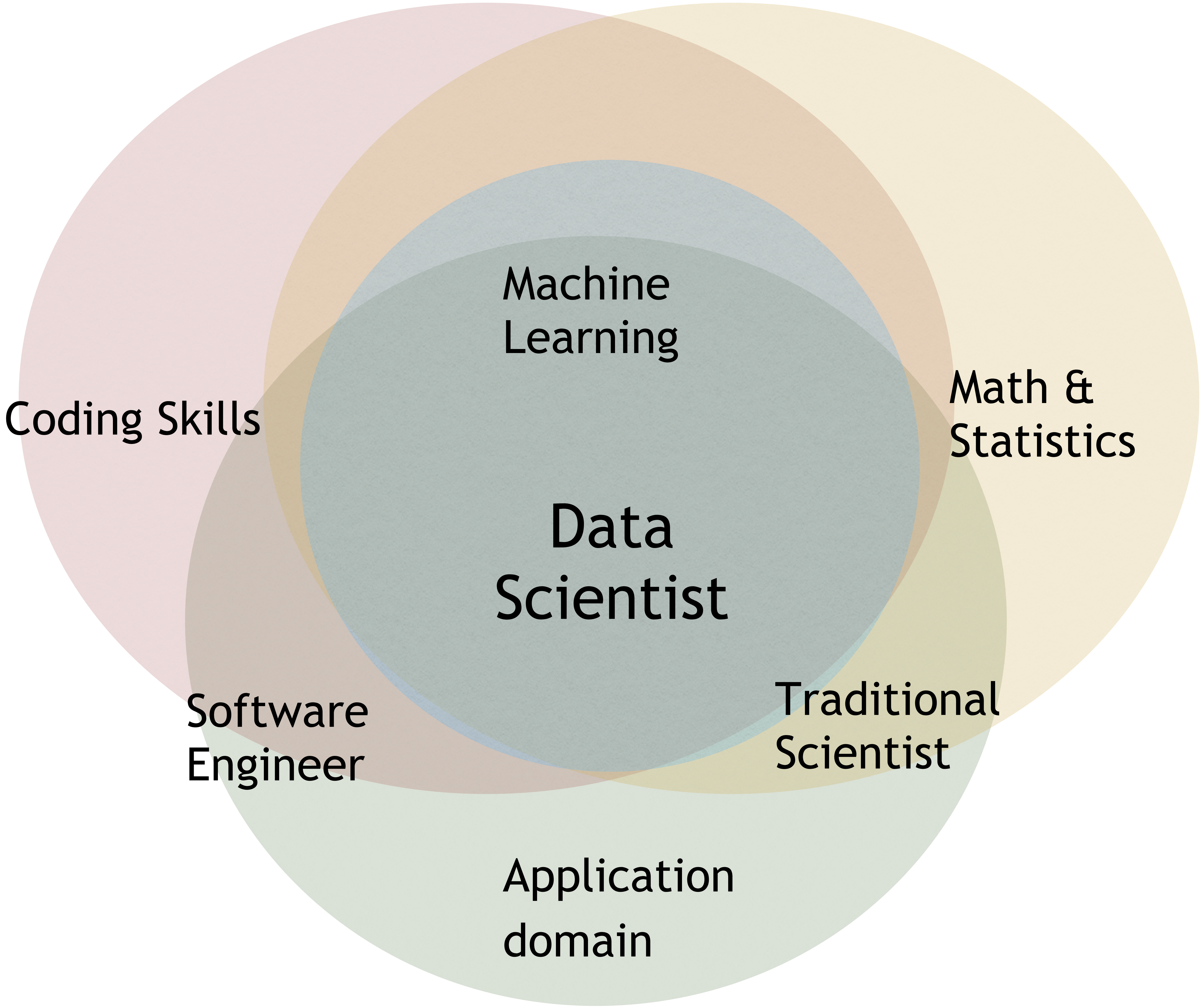
# What is supervised machine learning?



# ML and its broader context



# ML and jobs


Machine Learning
Coding Skills
Math & Statistics
Data Scientist
Software Engineer
Traditional Scientist
Application domain

# ML and jobs

Harvard Business Review

Coronavirus Magazine Popular Topics Podcasts Video Store The Big Idea Visual Library

Artwork: Tamar Cohen, Andrew J Buboltz, 2011, silk screen on a page from a high school yearbook, 8.5" x 12"

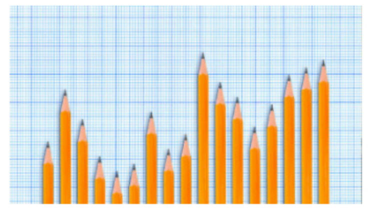
DATA

## Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

From the October 2012 Issue

WHAT TO READ NEXT

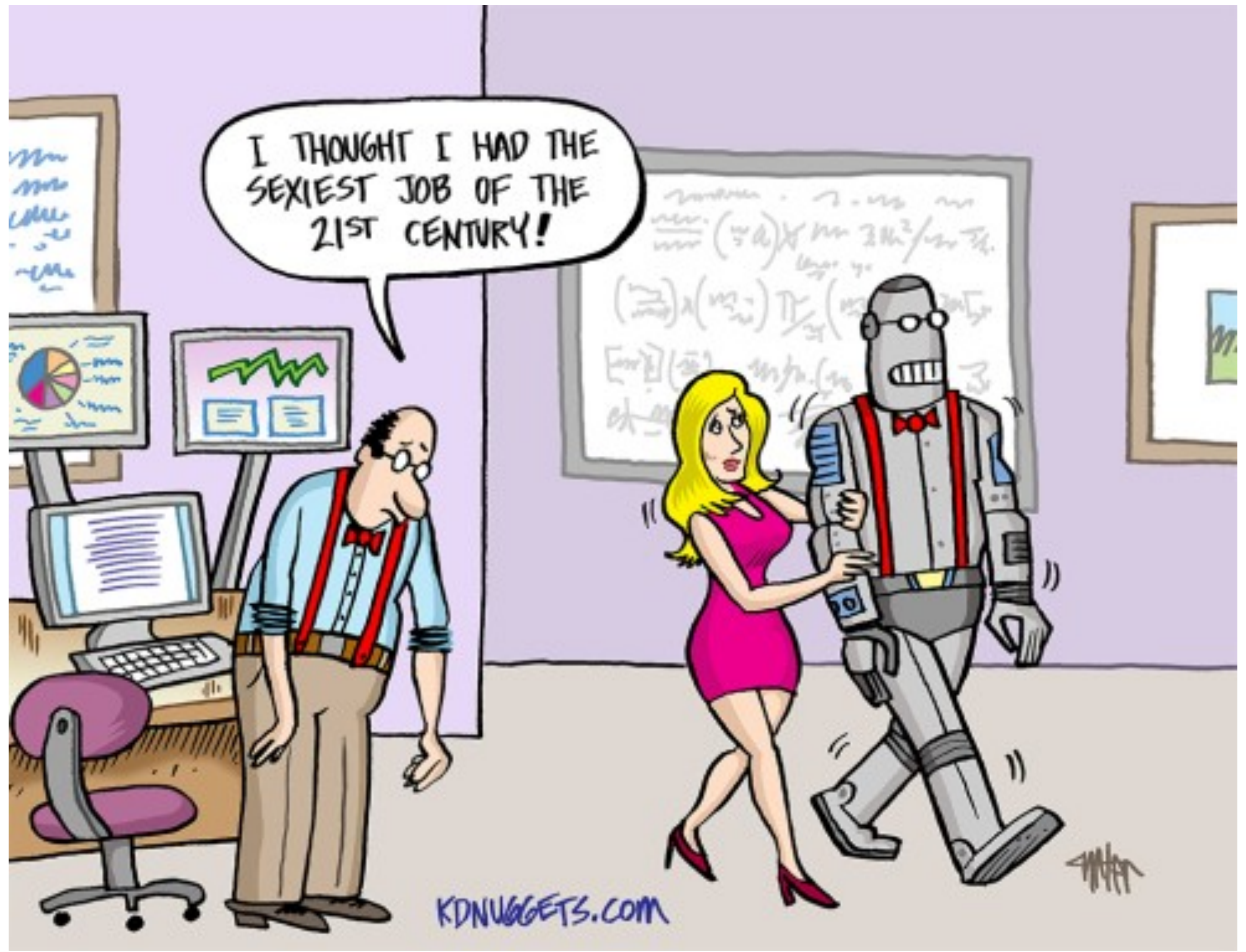


**What Data Scientists Really Do, According to 35 Data Scientists**

DATA SCIENTISTS

## Is 'Data Scientist' the 'Sexiest Job of the 21st Century'? And How Do You Get One of Your Own?

*Even if you're not versed in advanced analytics and data science, you can understand the thought process data scientists go through.*



# ML and jobs

## Royal Society: Dynamics of Data Science

Ensuring a Healthy Data Science Skills Landscape Across All Sectors in the UK

Britain is facing "explosive demand" for data science skills and its education system needs to change to keep up, according to a Burning Glass Technologies analysis commissioned by the Royal Society.

Demand for data scientists and data engineers tripled over the past five years, rising 231%. That's much faster than job postings overall in the UK, which rose 36%, according to the report, "Dynamics of data science skills."

Different regions saw different growth rates, from 79% in Wales to 269% in the North West and 563% in Northern Ireland. On average, salaries for these roles is £64,376, up 22% over the same period.

Job of the 21st Century

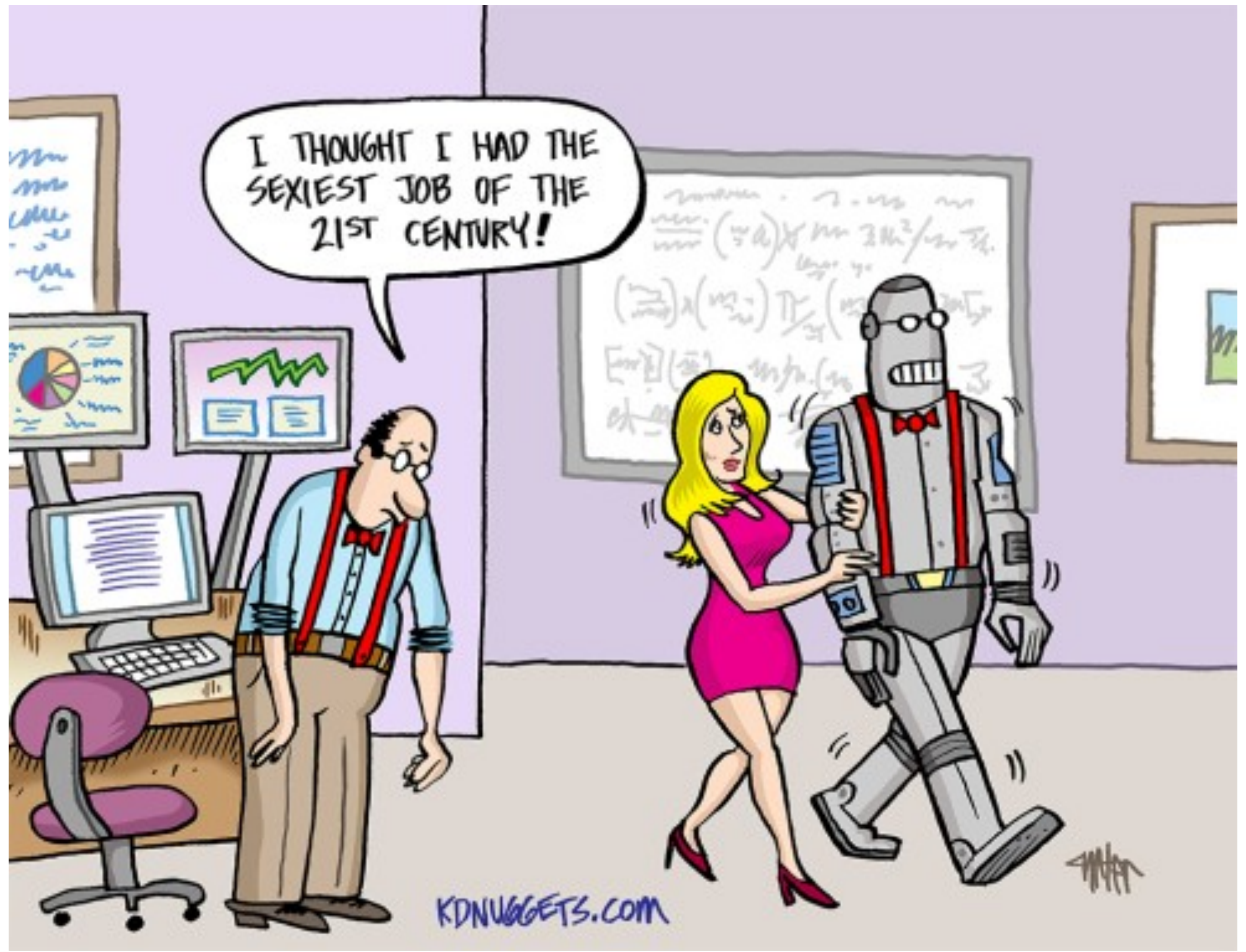
by Thomas H. Davenport and D.J. Patil

From the October 2012 Issue

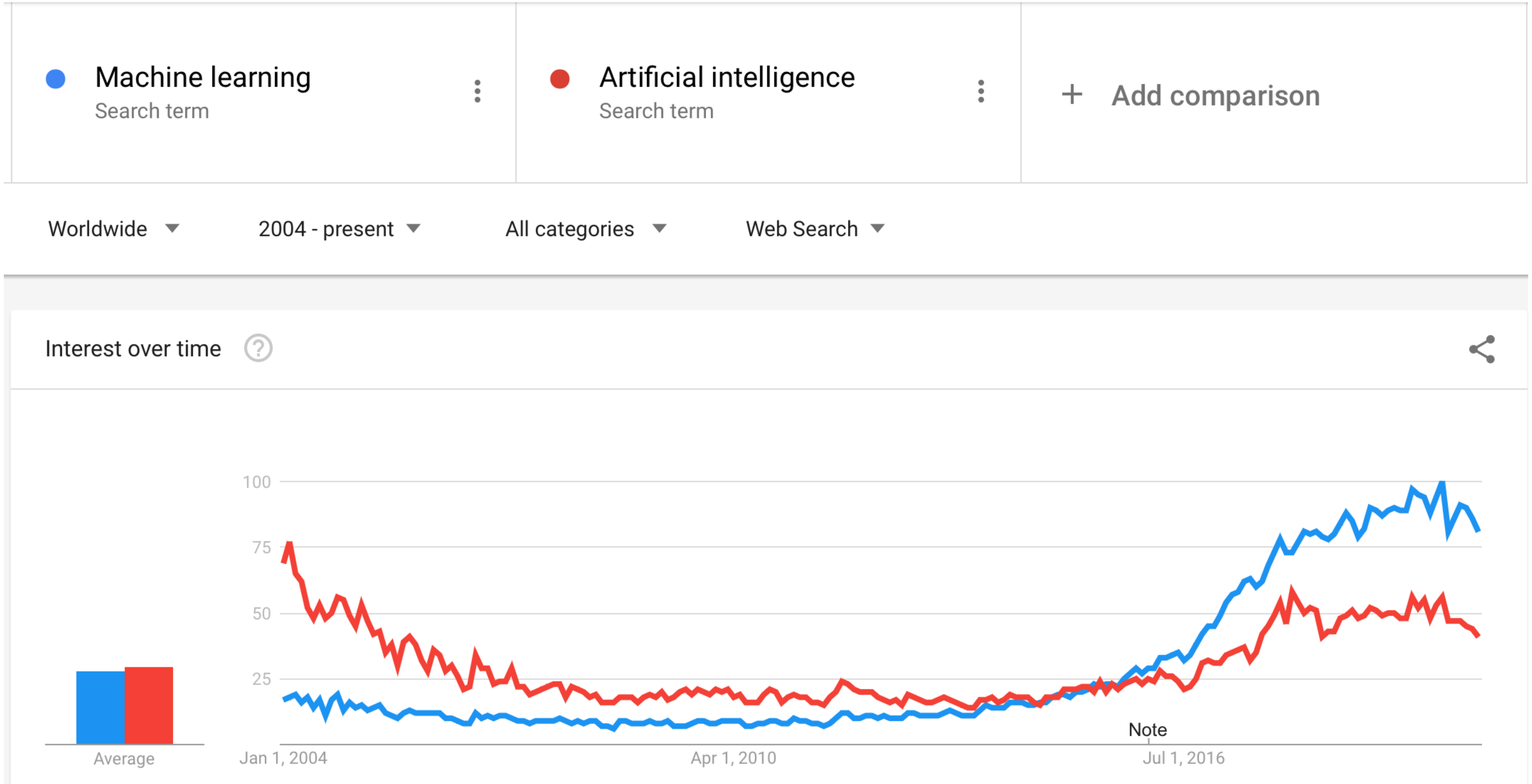
What Data Scientists Really Do, According to 35 Data Scientists

ientist' the 'Sexiest Job of the ury'? And How Do You Get One wn?

advanced analytics and data science, you can understand the thought process data



# ML popularity



# ML and jobs

The usual suspects

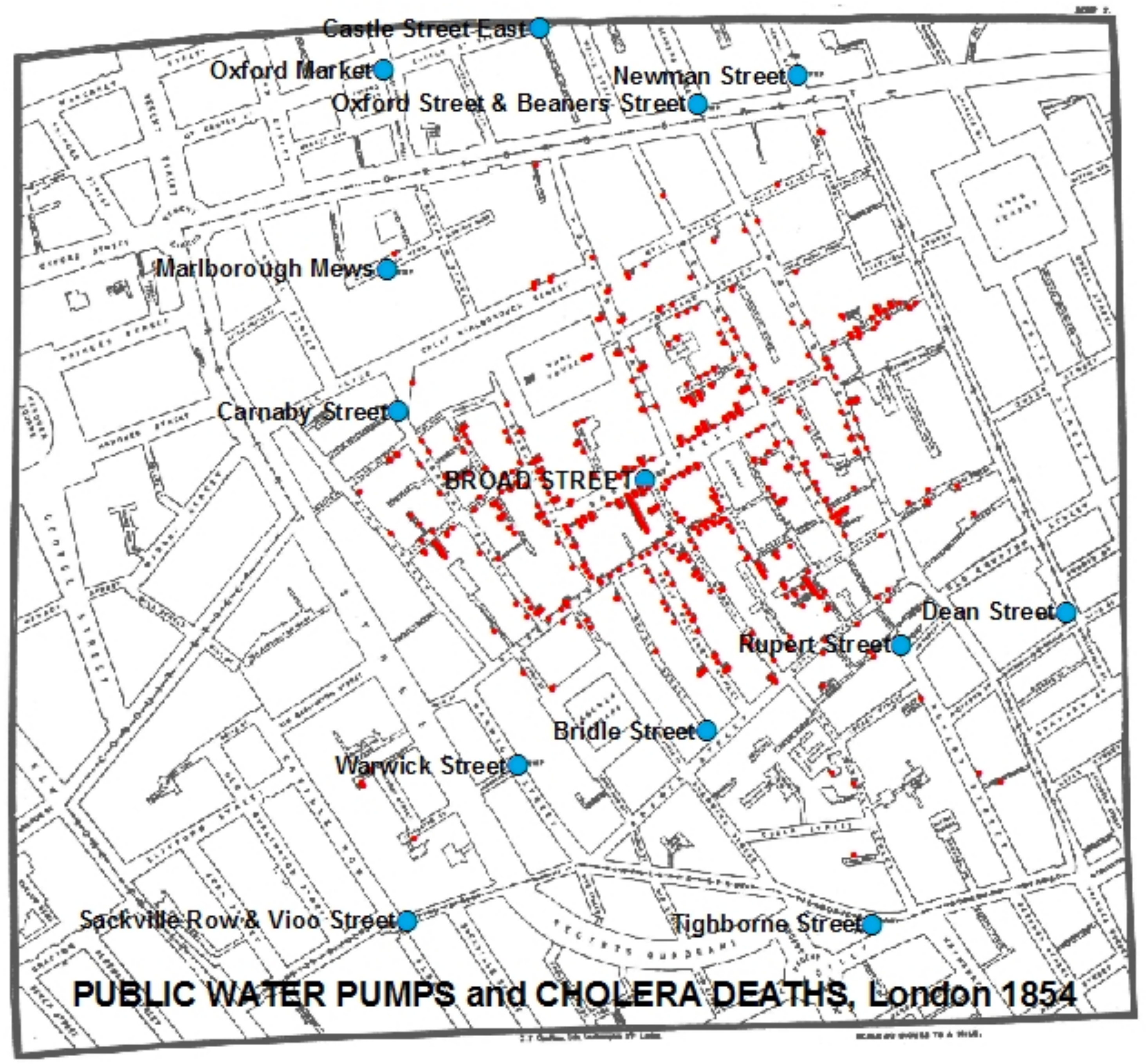
# ML and jobs

The usual suspects

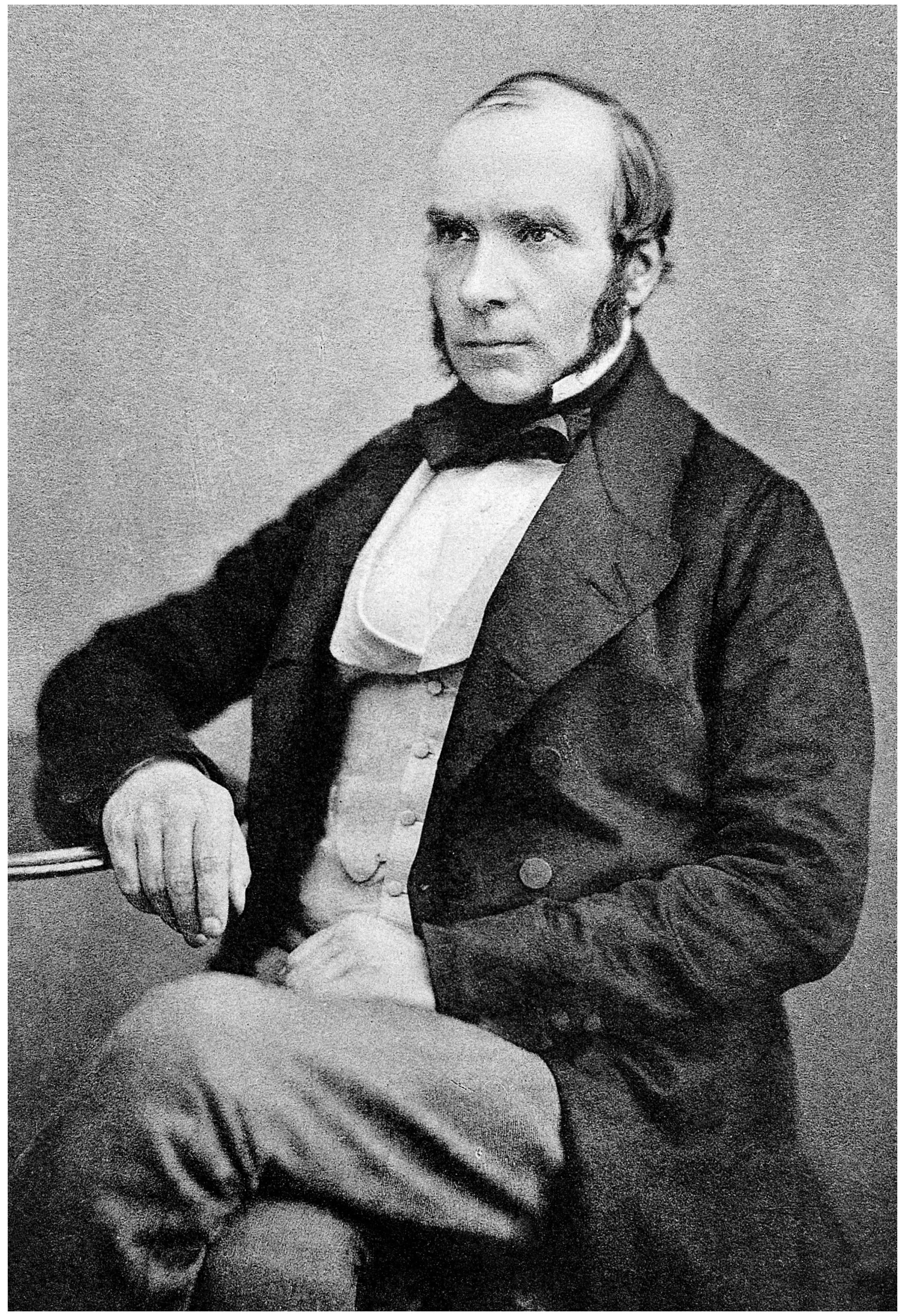
No longer just them

Majority of industries hiring data scientists

# Machine learning is not new



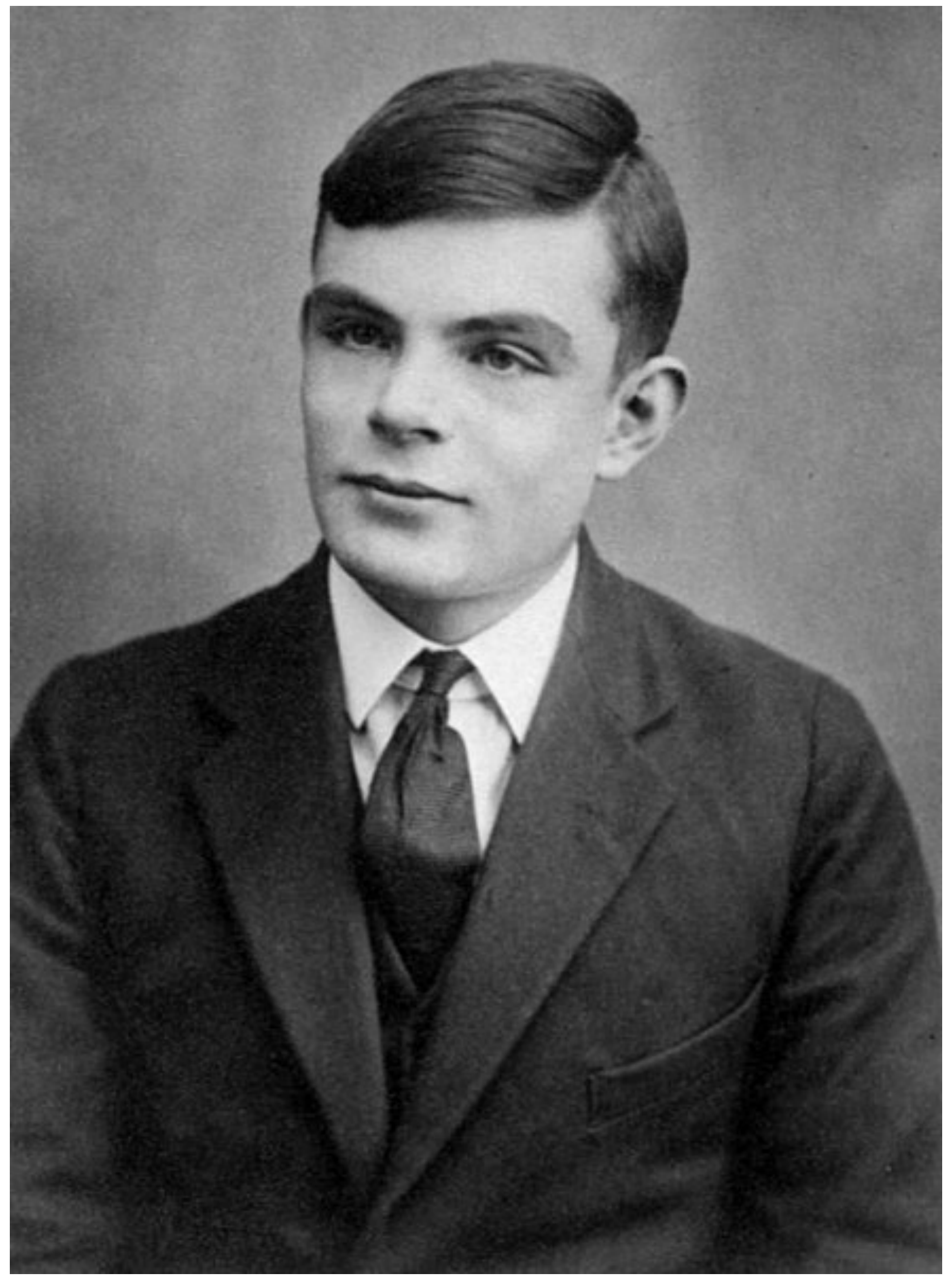

1854 Broad Street cholera outbreak



John Snow. ©Wikimedia commons

©Wikimedia commons

# Machine learning is not new



Turing

Rosenblatt

# What has changed?

*1950s:* $10^3$ FLOPS

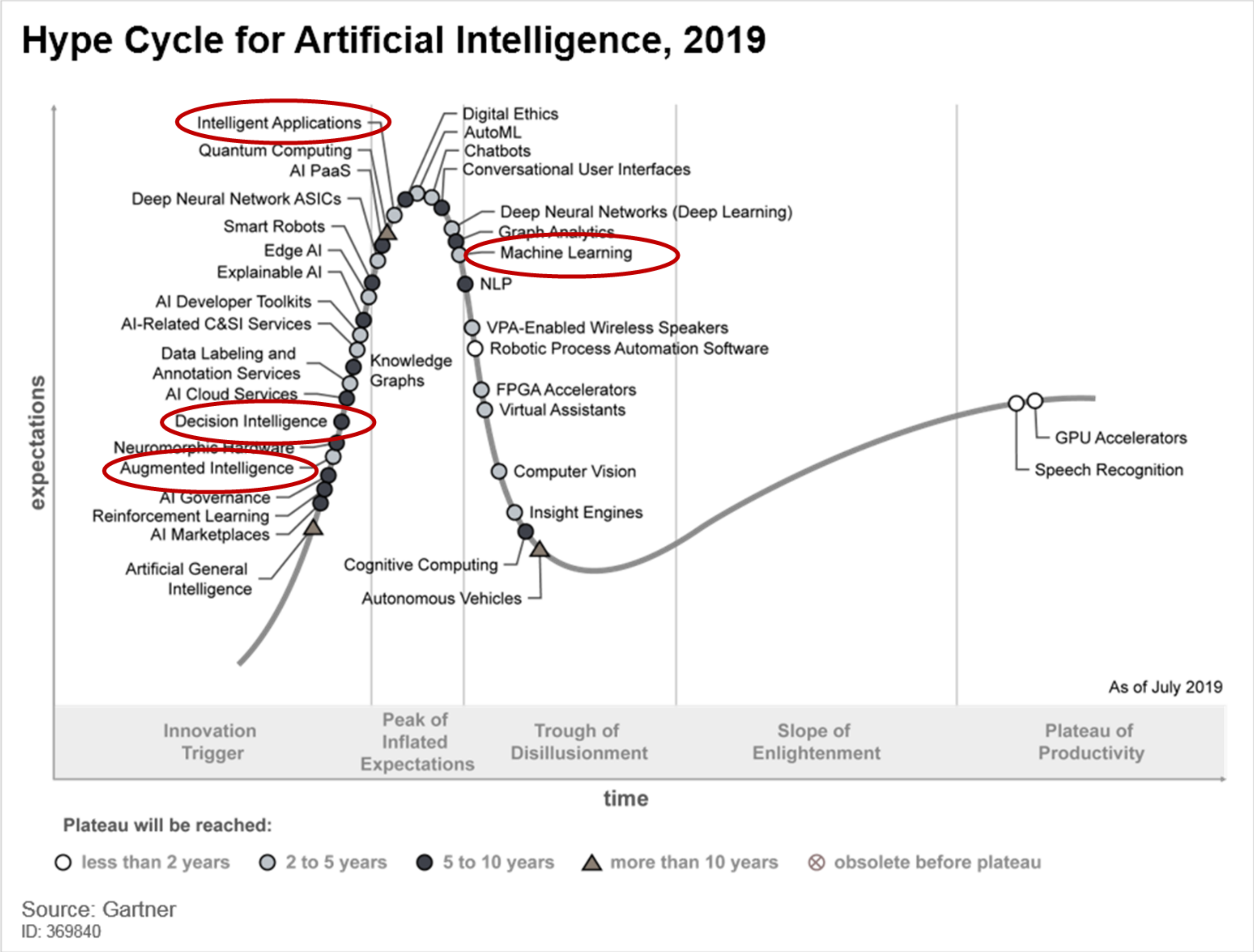
*2020:* $10^{18}$ FLOPS

FLOPS -> floating point operations per second

# Challenges

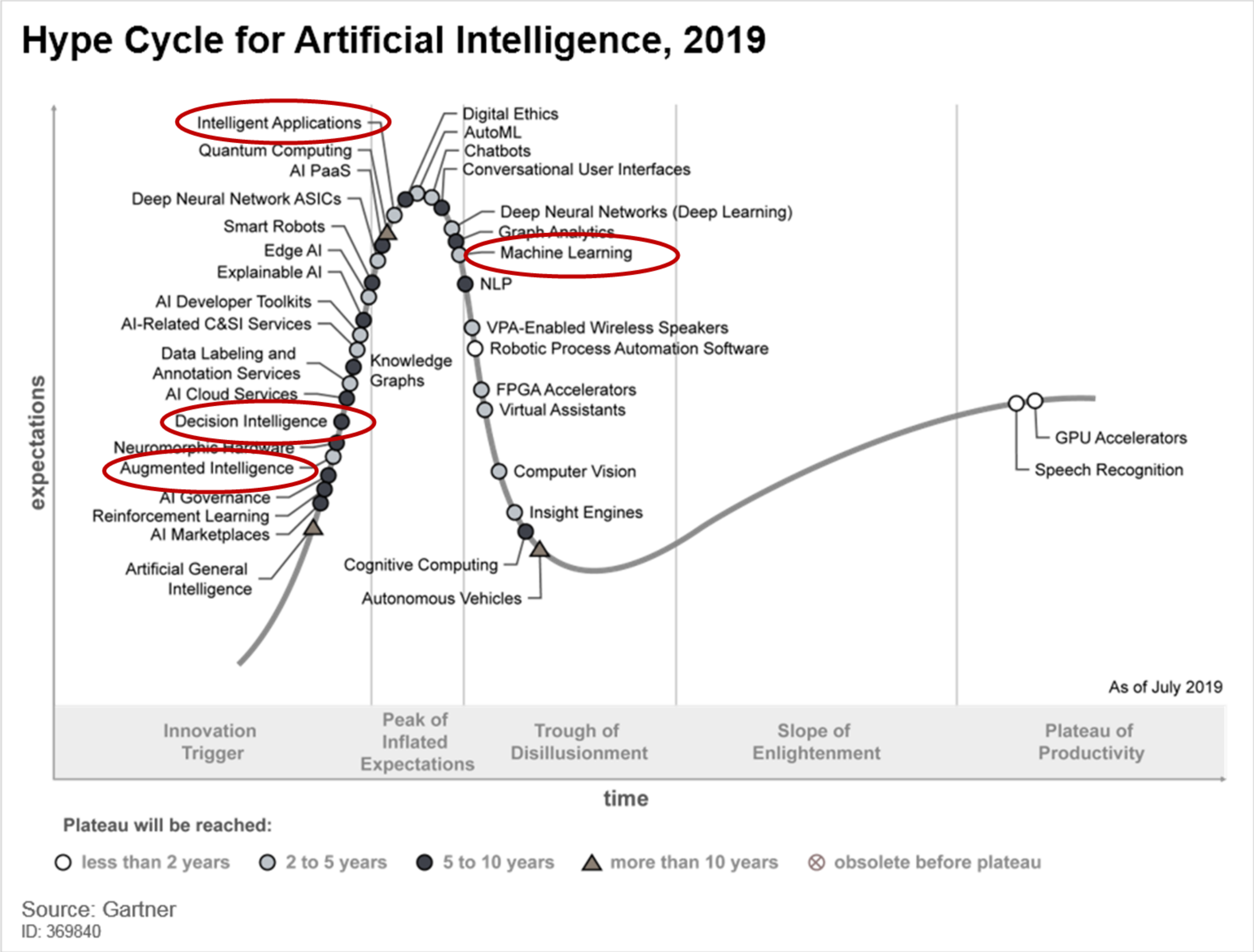
# Challenges

- Hype, i.e. cycles of AI and ML popularity



# Challenges

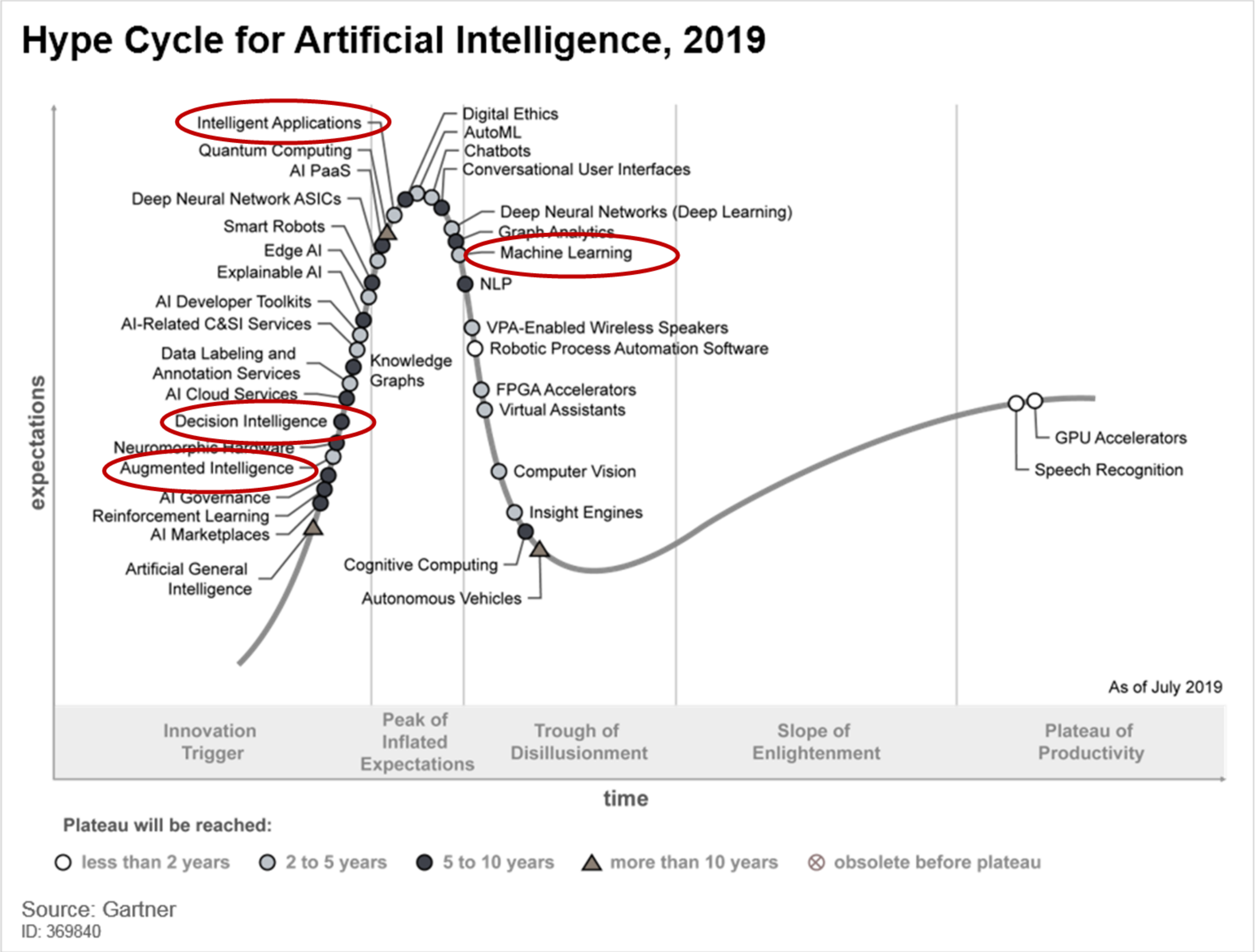
- Hype, i.e. cycles of AI and ML popularity
- Data Ethics, Privacy, Fairness




n.perra@qmul.ac.uk

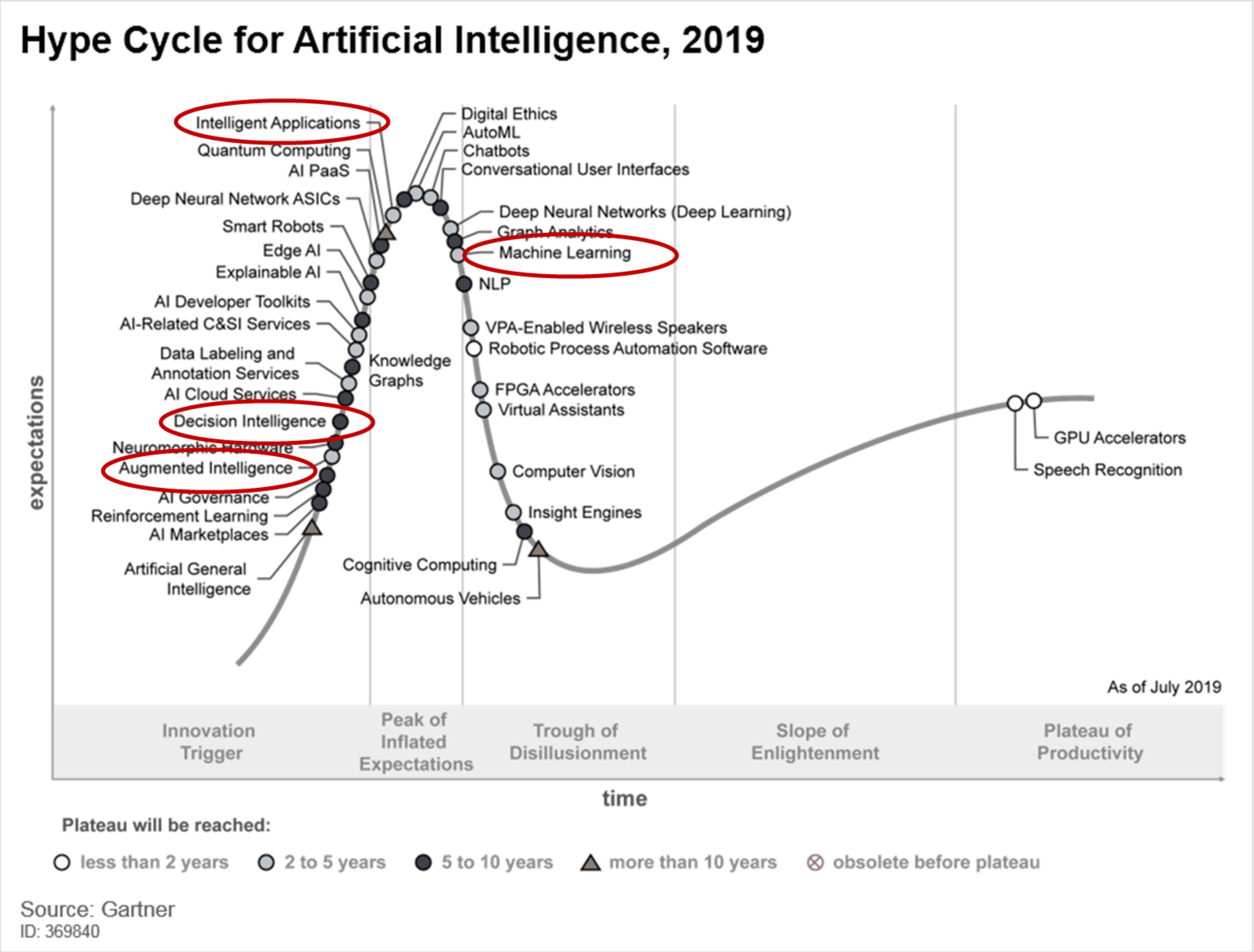
# Challenges

- Hype, i.e. cycles of AI and ML popularity
- Data Ethics, Privacy, Fairness
- Lack of Interpretability



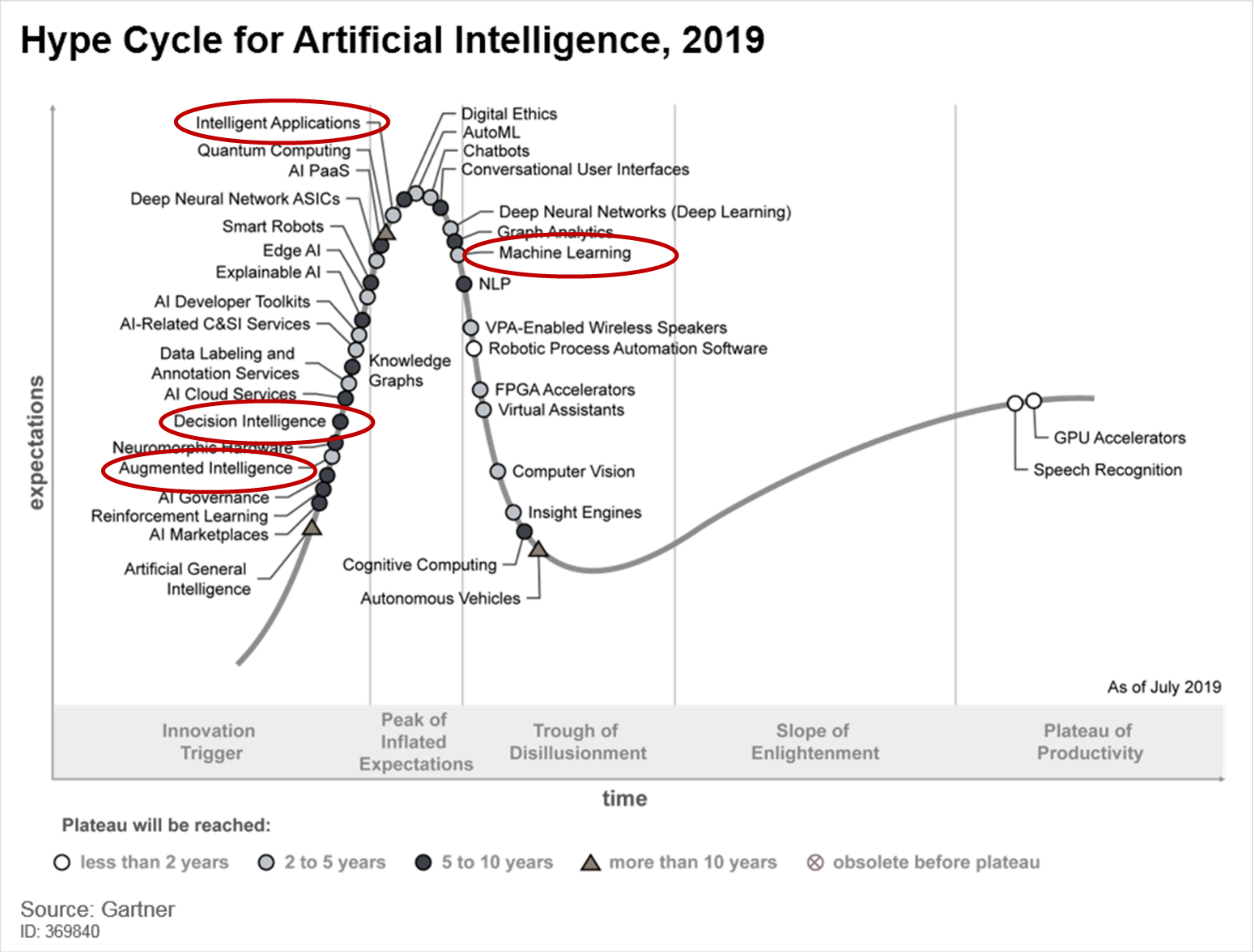
# Challenges

- Hype, i.e. cycles of AI and ML popularity
- Data Ethics, Privacy, Fairness
- Lack of Interpretability
- Social Implications: threats from super-human AI? Job losses?



# Challenges

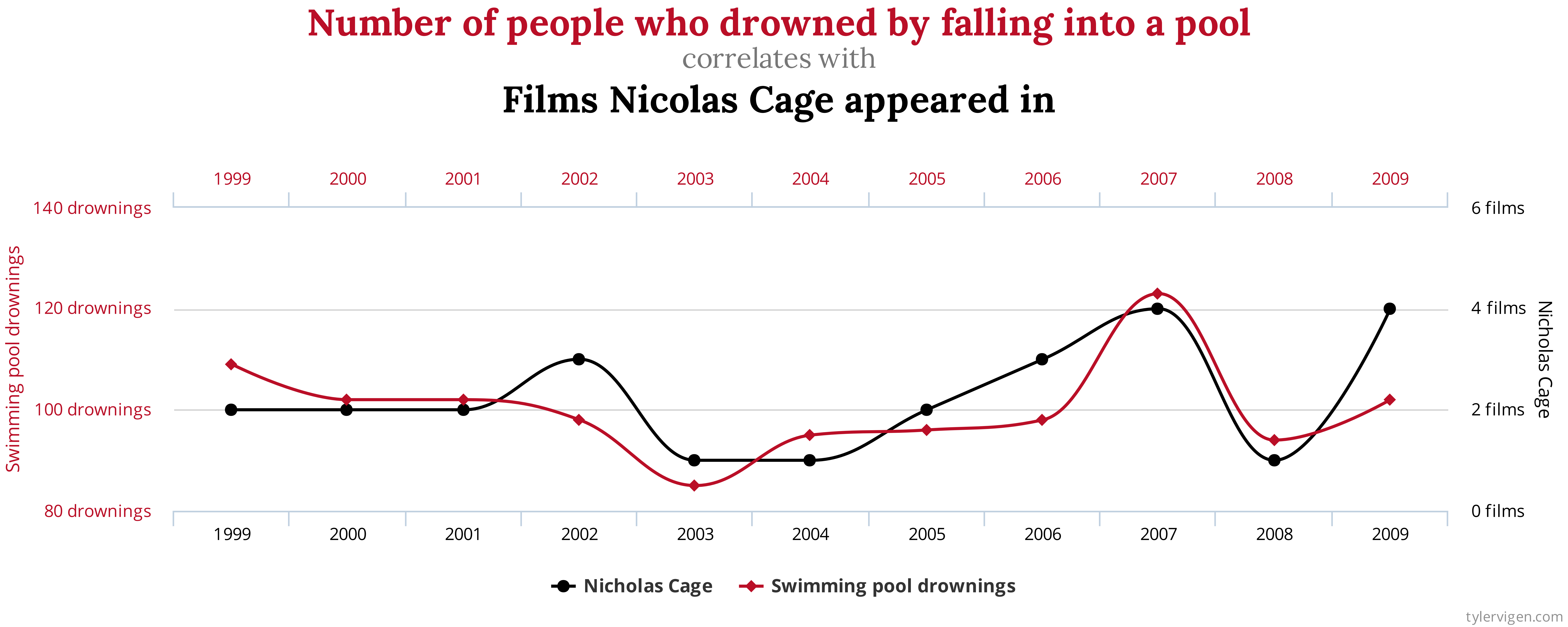
- Hype, i.e. cycles of AI and ML popularity
- Data Ethics, Privacy, Fairness
- Lack of Interpretability
- Social Implications: threats from super-human AI? Job losses?




We need: scientific method, reproducible research, open source and open data

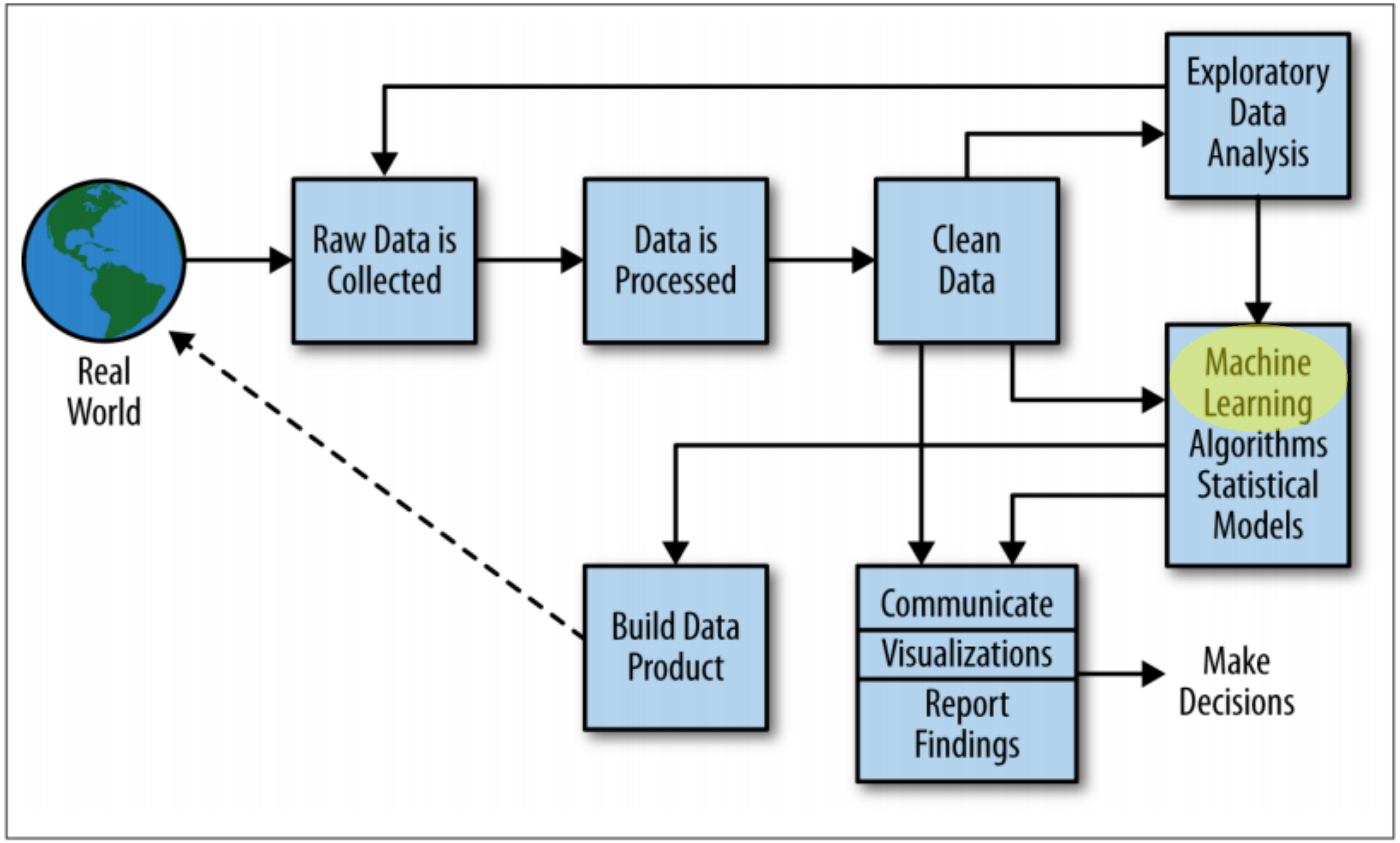
# Correlation ≠ causation*

What we not cover in this module: causal inference**




*in general, and without additional prior assumptions
**Pearl, Judea, Madelyn Glymour, and Nicholas P. Jewell. *Causal inference in statistics: A primer*. John Wiley & Sons, 2016.

# Machine learning is only a small part




From Schutt and O'Neil "Doing Data Science"

# LINEAR ALGEBRA

# Matrices and Vectors

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

# 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! $s \times d \quad d \times 1 \to s \times 1$

n.perra@qmul.ac.uk

# Matrices and Vectors

Example

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

# Matrices and Vectors

Example

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

What is the expected dimension of the outcome of this multiplication?

# Matrices and Vectors

Example

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

What is the expected dimension of the outcome of this multiplication?

$$3 \times 3$$

# Matrices and Vectors

Example

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

What is the expected dimension of the outcome of this multiplication?

$$3 \times 3 \quad 3 \times 1$$

# Matrices and Vectors

Example

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

What is the expected dimension of the outcome of this multiplication?

$$3 \times 3 \quad 3 \times 1 \rightarrow 3 \times 1$$

n.perra@qmul.ac.uk

# Matrices and Vectors

How do we compute the result of the multiplication?

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

# Matrices and Vectors

How do we compute the result of the multiplication?

General rule: rows times columns

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

n.perra@qmul.ac.uk

# Matrices and Vectors

How do we compute the result of the multiplication?

General rule: rows times columns

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

# Matrices and Vectors

How do we compute the result of the multiplication?

General rule: rows times columns

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

# Matrices and Vectors

How do we compute the result of the multiplication?

General rule: rows times columns

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

# Matrices and Vectors

How do we compute the result of the multiplication?

General rule: rows times columns

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

# Matrices and Vectors

How do we compute the result of the multiplication?

General rule: rows times columns

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

# Matrices and Vectors

How do we compute the result of the multiplication?

General rule: rows times columns

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

n.perra@qmul.ac.uk

# Matrices and Vectors

How do we compute the result of the multiplication?

General rule: rows times columns

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

# Matrices and Vectors

How do we compute the result of the multiplication?

General rule: rows times columns

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

# Matrices and Vectors

How do we compute the result of the multiplication?

General rule: rows times columns

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

# Matrices and Vectors

How do we compute the result of the multiplication?

General rule: rows times columns

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 4 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \times 1 + 2 \times 4 + 3 \times 0 \\ 0 \times 1 + 1 \times 4 + 2 \times 0 \\ 1 \times 1 - 1 \times 4 + 1 \times 0 \end{pmatrix} = \begin{pmatrix} 9 \\ 4 \\ -3 \end{pmatrix}$$

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

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

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

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

n.perra@qmul.ac.uk

# Matrices and Vectors

The row times column rule can be written as

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

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

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

# 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

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

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

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

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

# Matrices and Vectors

Matrix-matrix multiplication:

# Matrices and Vectors

Matrix-matrix multiplication: $\mathbf{XW} = \mathbf{Y}$ $\Leftrightarrow$ $\sum_{j=1}^{d} x_{ij} w_{jk} = y_{ik}$ for

$$X \in \mathbb{R}^{s \times d}, W \in \mathbb{R}^{d \times m}, Y \in \mathbb{R}^{s \times m}$$

# Matrices and Vectors

Matrix-matrix multiplication: $\mathbf{XW} = \mathbf{Y}$ $\Leftrightarrow$ $\sum_{j=1}^{d} x_{ij} w_{jk} = y_{ik}$ for

$$X \in \mathbb{R}^{s \times d}, W \in \mathbb{R}^{d \times m}, Y \in \mathbb{R}^{s \times m}$$

$$s \times d \;\; d \times m \rightarrow s \times m$$

# Matrices and Vectors

Matrix-matrix multiplication: $\mathbf{XW} = \mathbf{Y}$ $\quad\Leftrightarrow\quad$ $\sum_{j=1}^{d} x_{ij} w_{jk} = y_{ik}$ for

$$X \in \mathbb{R}^{s \times d}, W \in \mathbb{R}^{d \times m}, Y \in \mathbb{R}^{s \times m}$$

$$s \times d \;\; d \times m \to s \times m$$

# Matrices and Vectors

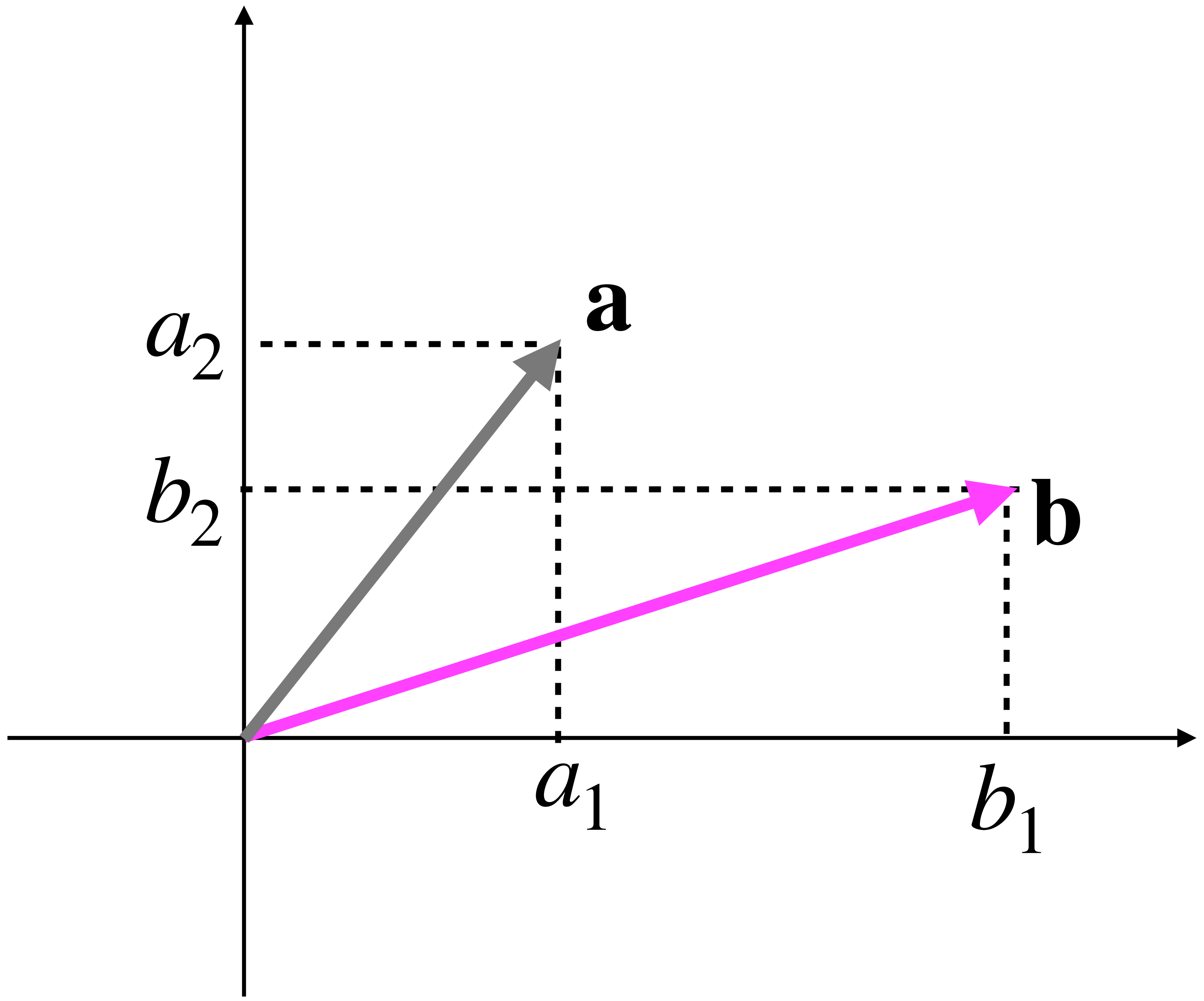
Matrix: $$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & & \ddots & \vdots \\ x_{s1} & x_{s2} & \cdots & x_{sd} \end{pmatrix} \in \mathbb{R}^{s \times d}$$

# Matrices and Vectors

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

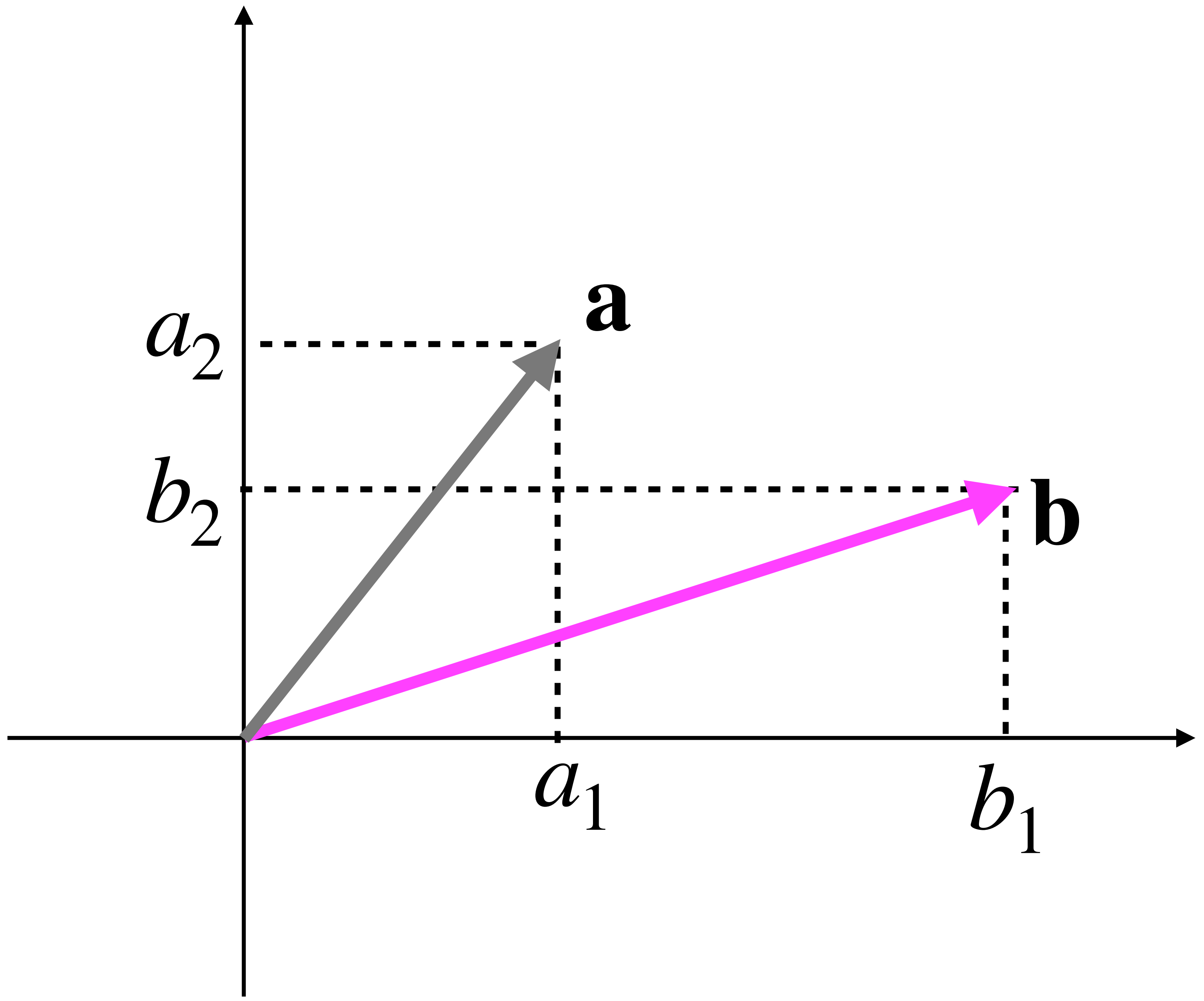
The transpose $X^\top$ of a matrix $X$ is defined as $$X^\top = \begin{pmatrix} x_{11} & x_{21} & \dots & x_{s1} \\ x_{12} & x_{22} & \dots & x_{s2} \\ \vdots & & \ddots & \vdots \\ x_{1d} & x_{2d} & \dots & x_{sd} \end{pmatrix} \in \mathbb{R}^{d \times s}$$

# Inner/dot/scalar product




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

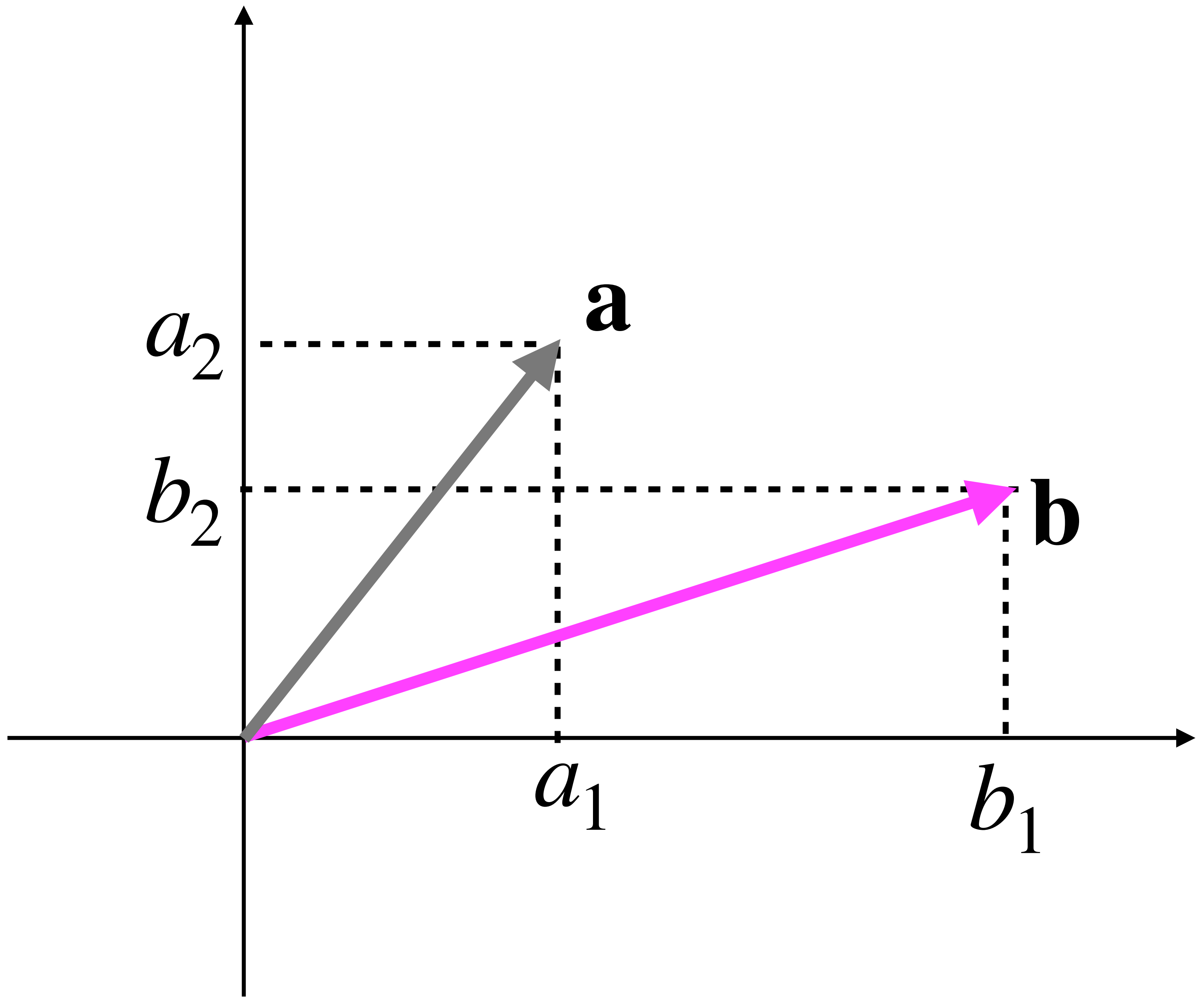
# Inner/dot/scalar product




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

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

# Inner/dot/scalar product

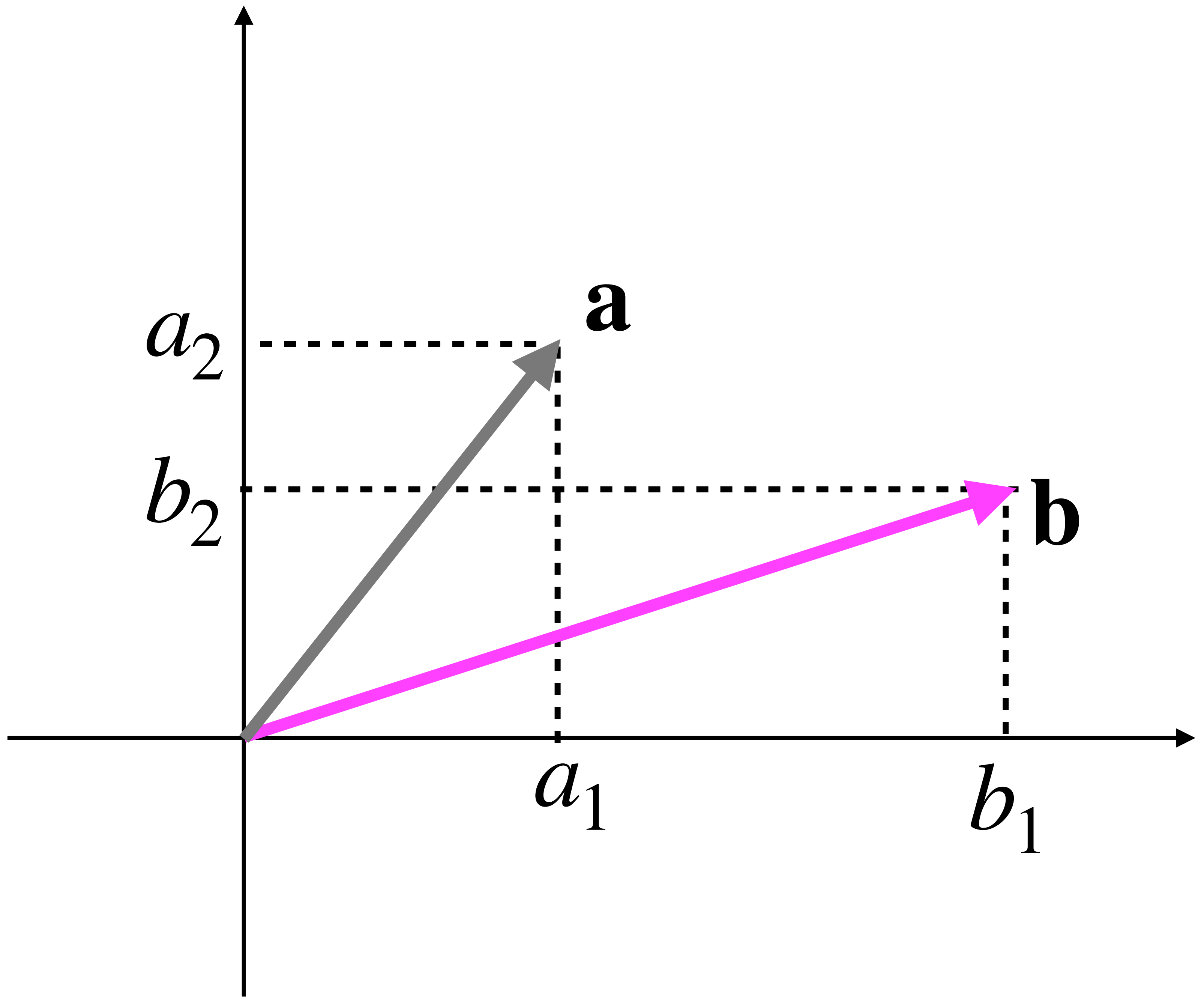



$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad \mathbf{a}, \mathbf{b} \in \mathbb{R}^{2\times 1}$$

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

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

# Inner/dot/scalar product



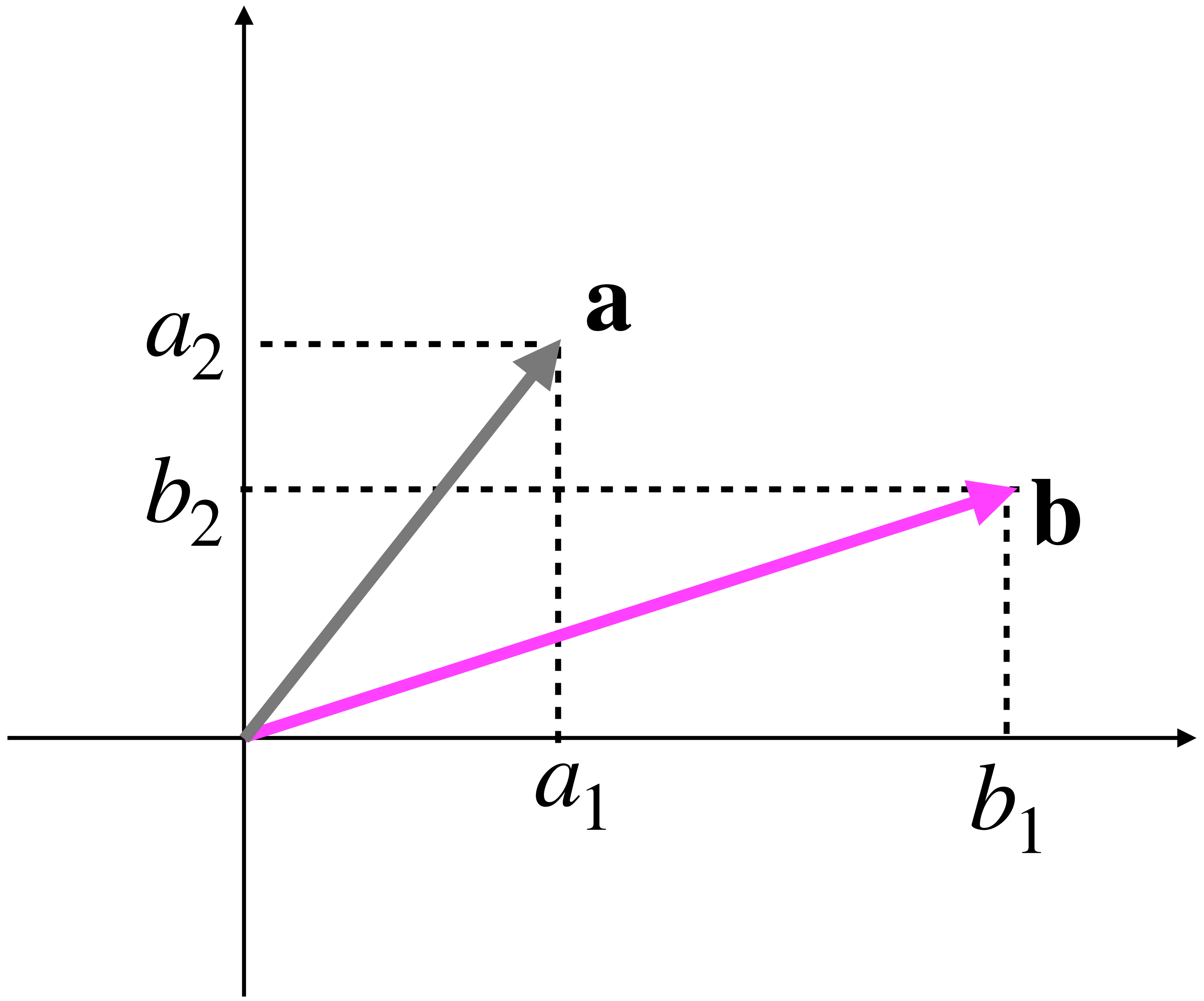

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad \mathbf{a}, \mathbf{b} \in \mathbb{R}^{2\times 1}$$

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

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

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

# Inner/dot/scalar product

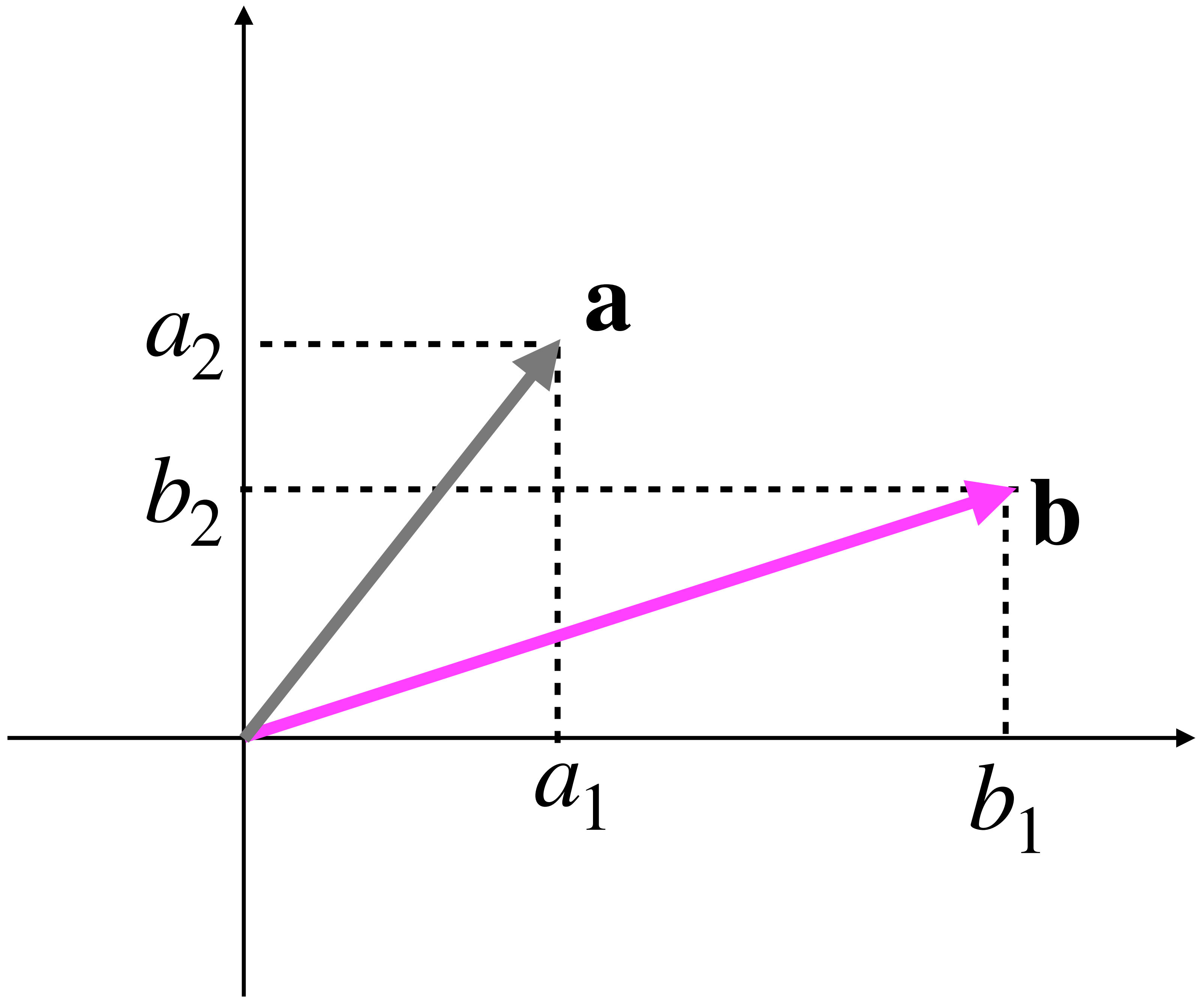



$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad \mathbf{a}, \mathbf{b} \in \mathbb{R}^{2\times 1}$$

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

n.perra@qmul.ac.uk

# Inner/dot/scalar product

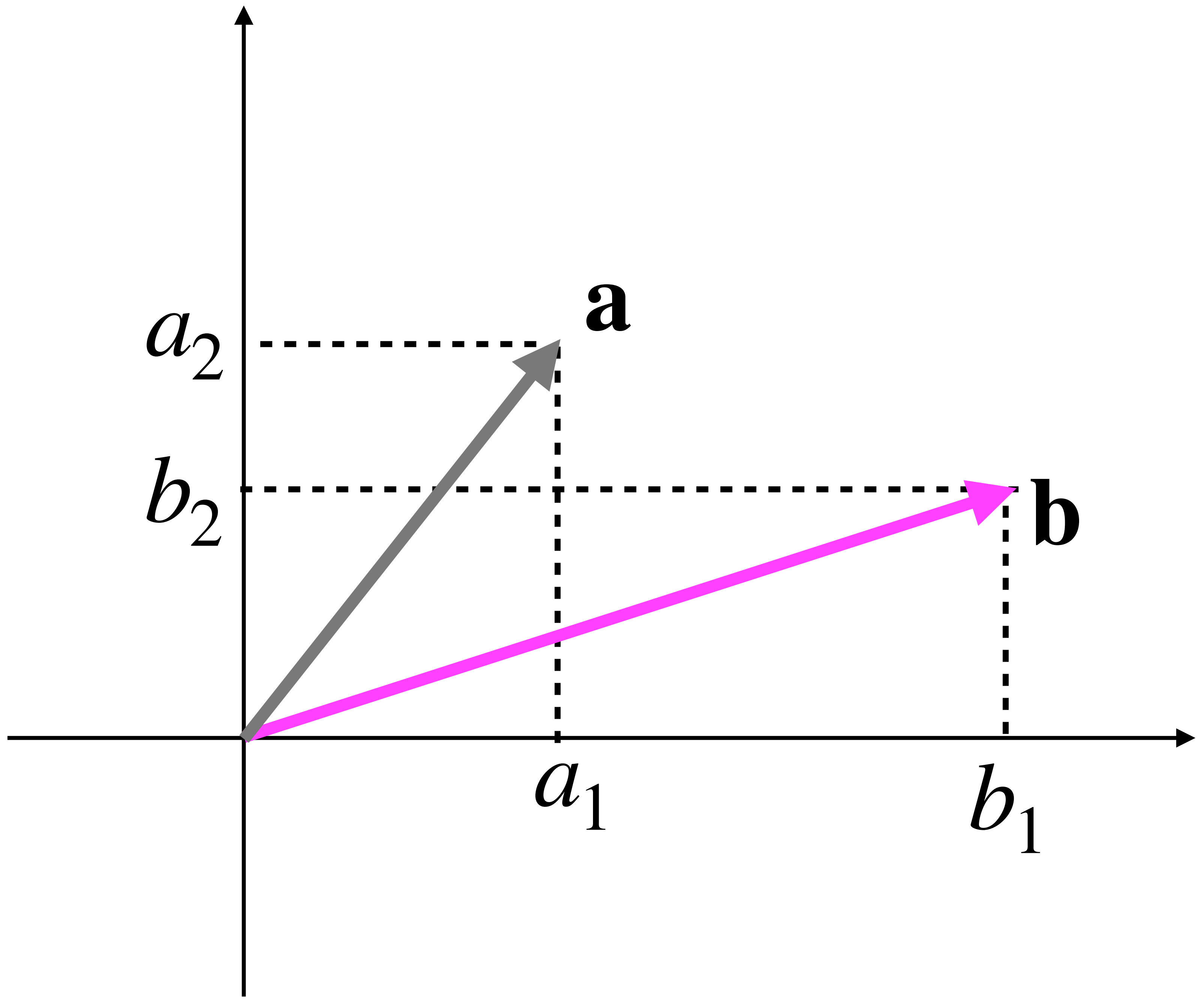



$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad \mathbf{a}, \mathbf{b} \in \mathbb{R}^{2\times 1}$$

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

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

# Inner/dot/scalar product



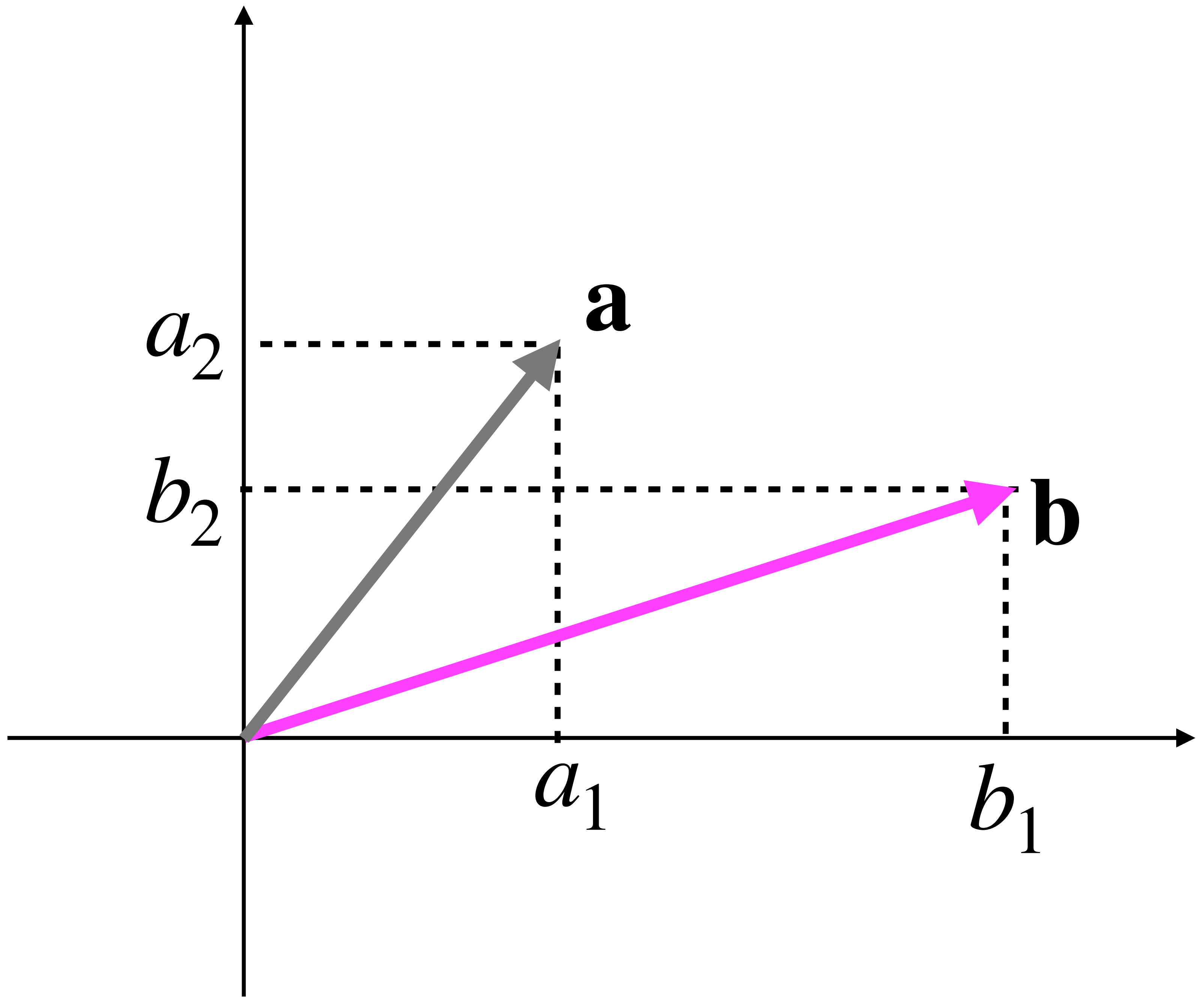

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad \mathbf{a}, \mathbf{b} \in \mathbb{R}^{2\times 1}$$

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

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

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

# Inner/dot/scalar product




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

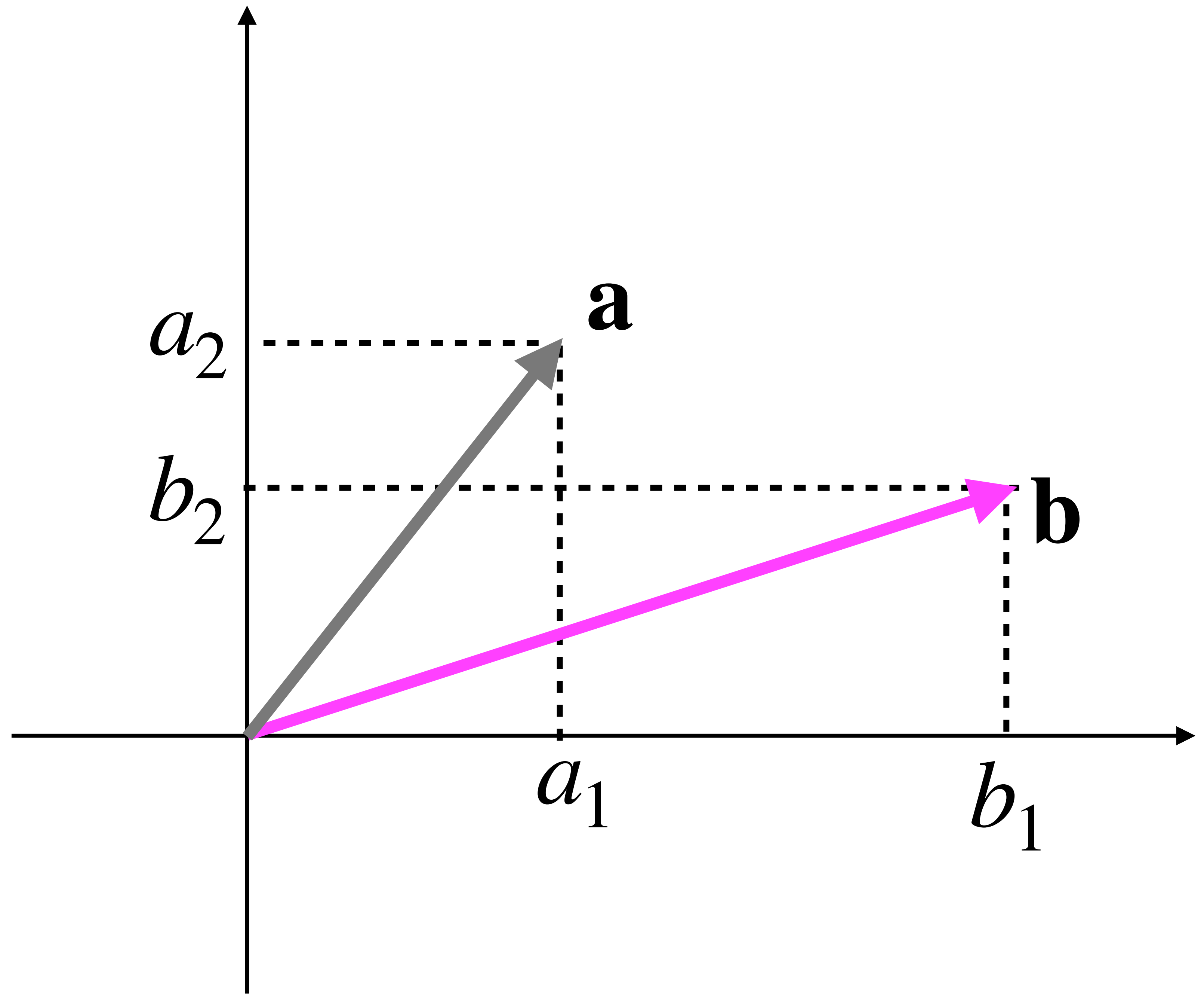
$$\mathbf{a} \cdot \mathbf{b} = \sum_{j=1}^{d} a_j b_j = a_1 b_1 + a_2 b_2$$

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

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

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

# Inner/dot/scalar product




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

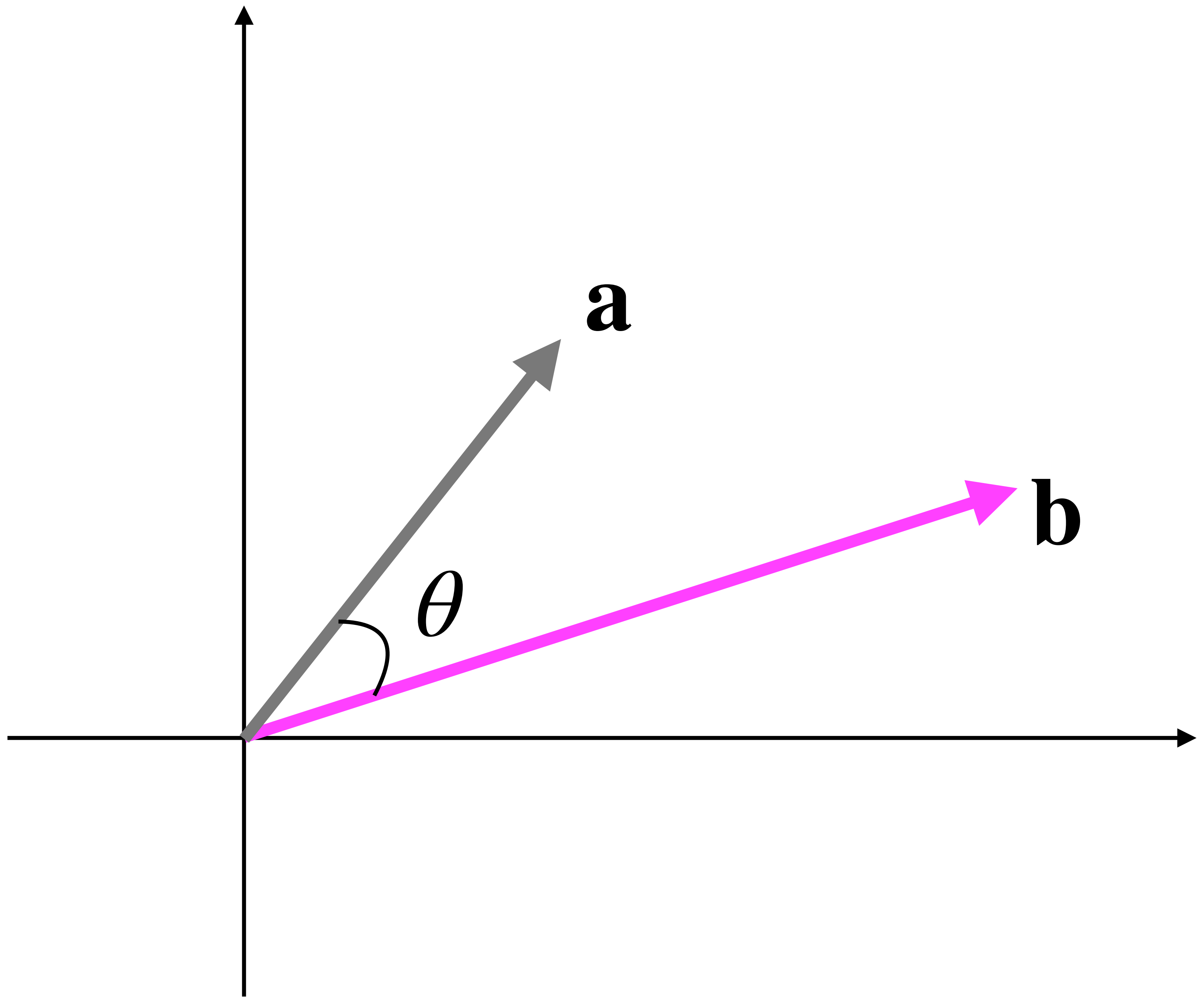
$$\mathbf{a} \cdot \mathbf{b} = \sum_{j=1}^{d} a_j b_j = a_1 b_1 + a_2 b_2$$

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

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

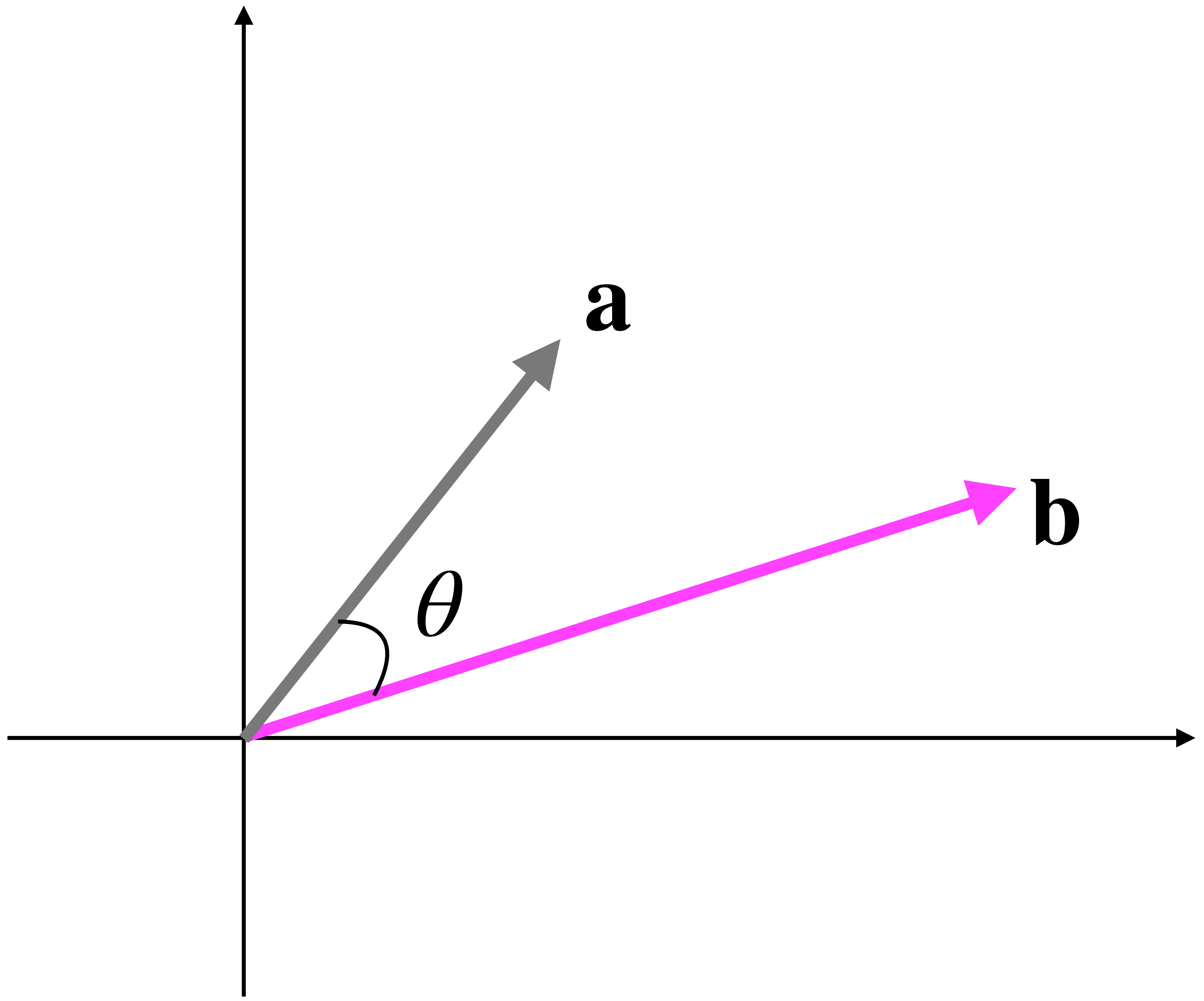
$$\mathbf{a}^\top \mathbf{b} = \begin{pmatrix} a_1 & a_2 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \mathbf{b}^\top \mathbf{a} = \begin{pmatrix} b_1 & b_2 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

# Inner/dot/scalar product




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

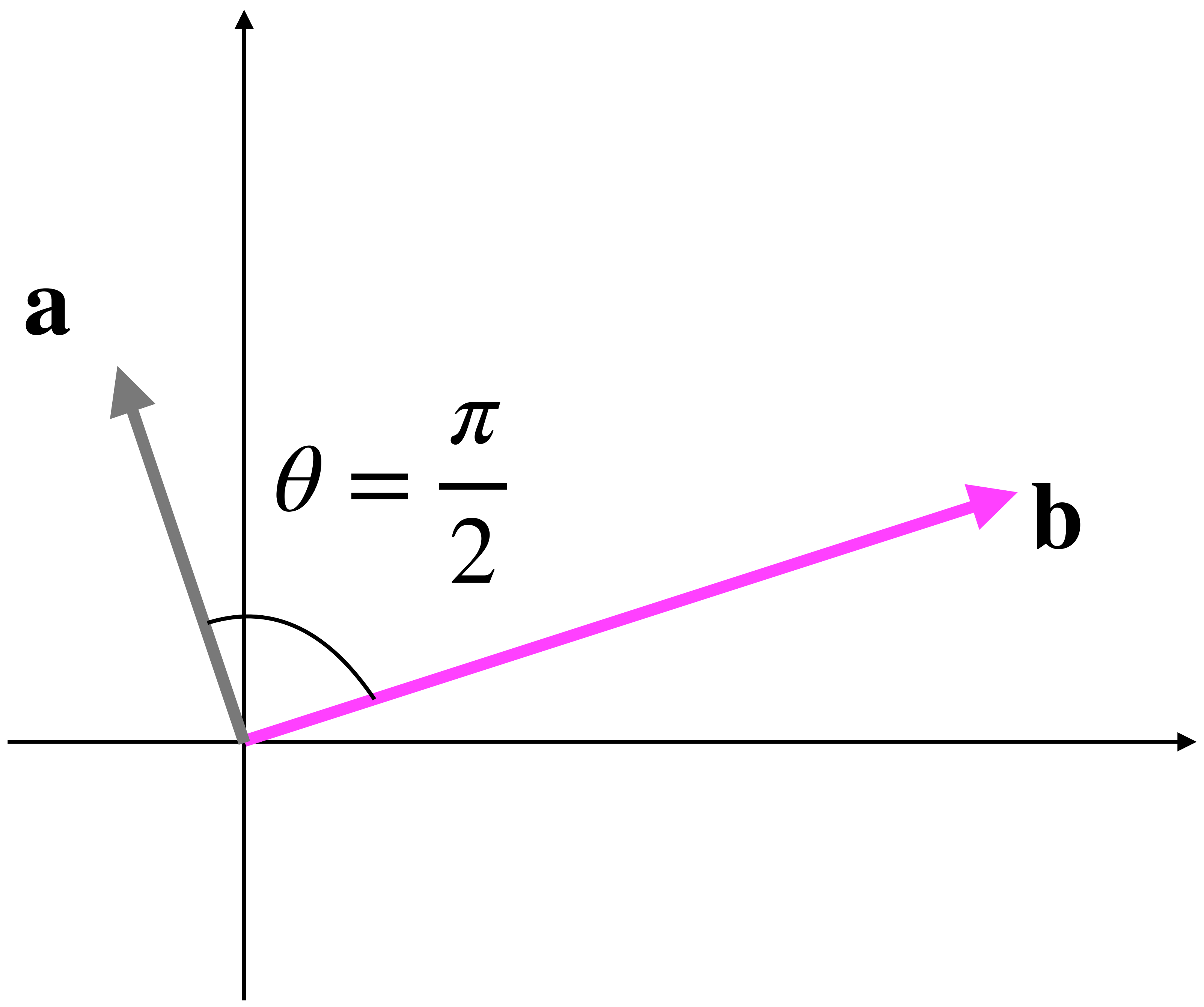
# Inner/dot/scalar product




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

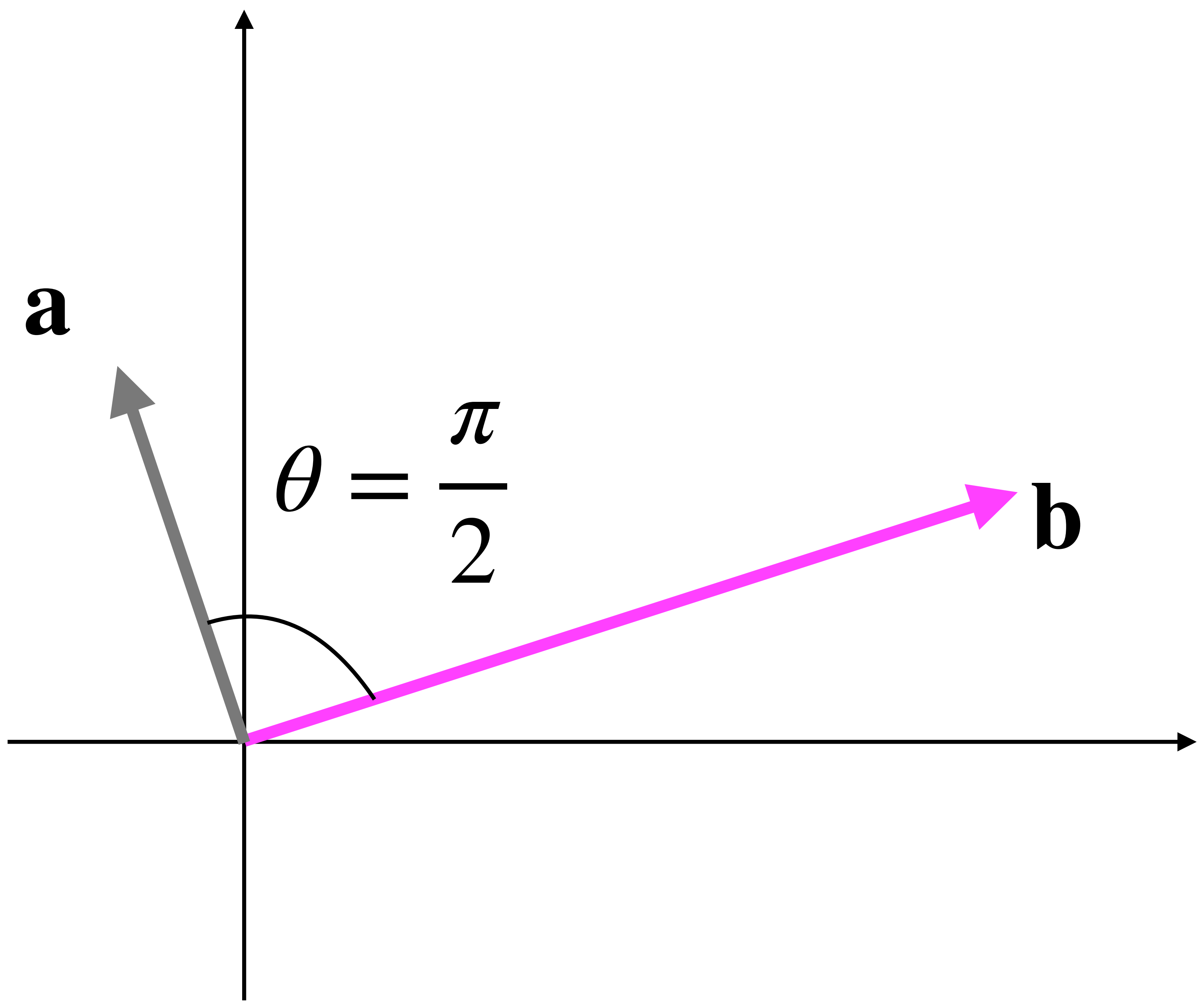
$$\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\| \|\mathbf{b}\| \cos\theta$$

# Inner/dot/scalar product



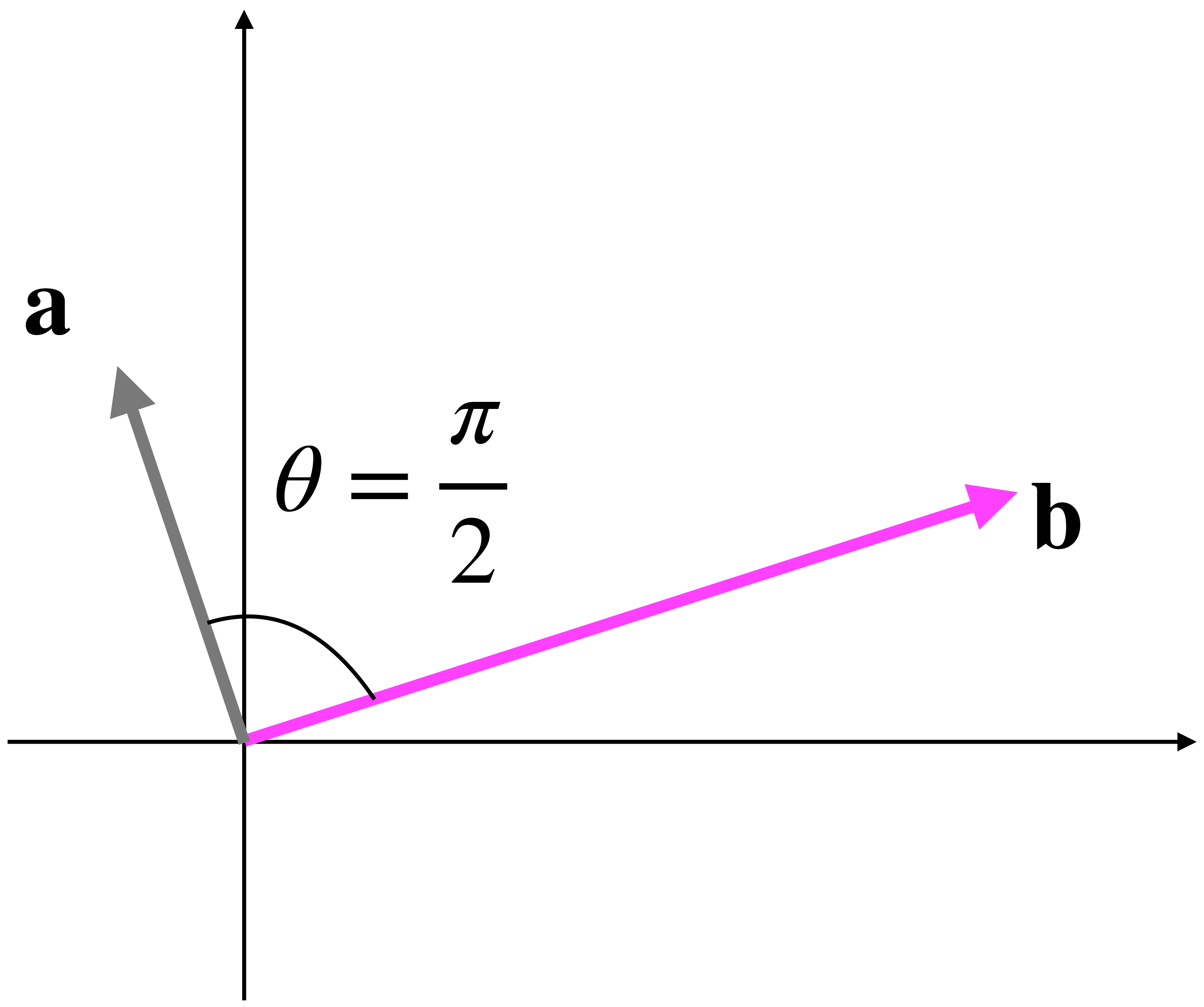
# Inner/dot/scalar product

What happens if the two vectors are orthogonal?



# Inner/dot/scalar product

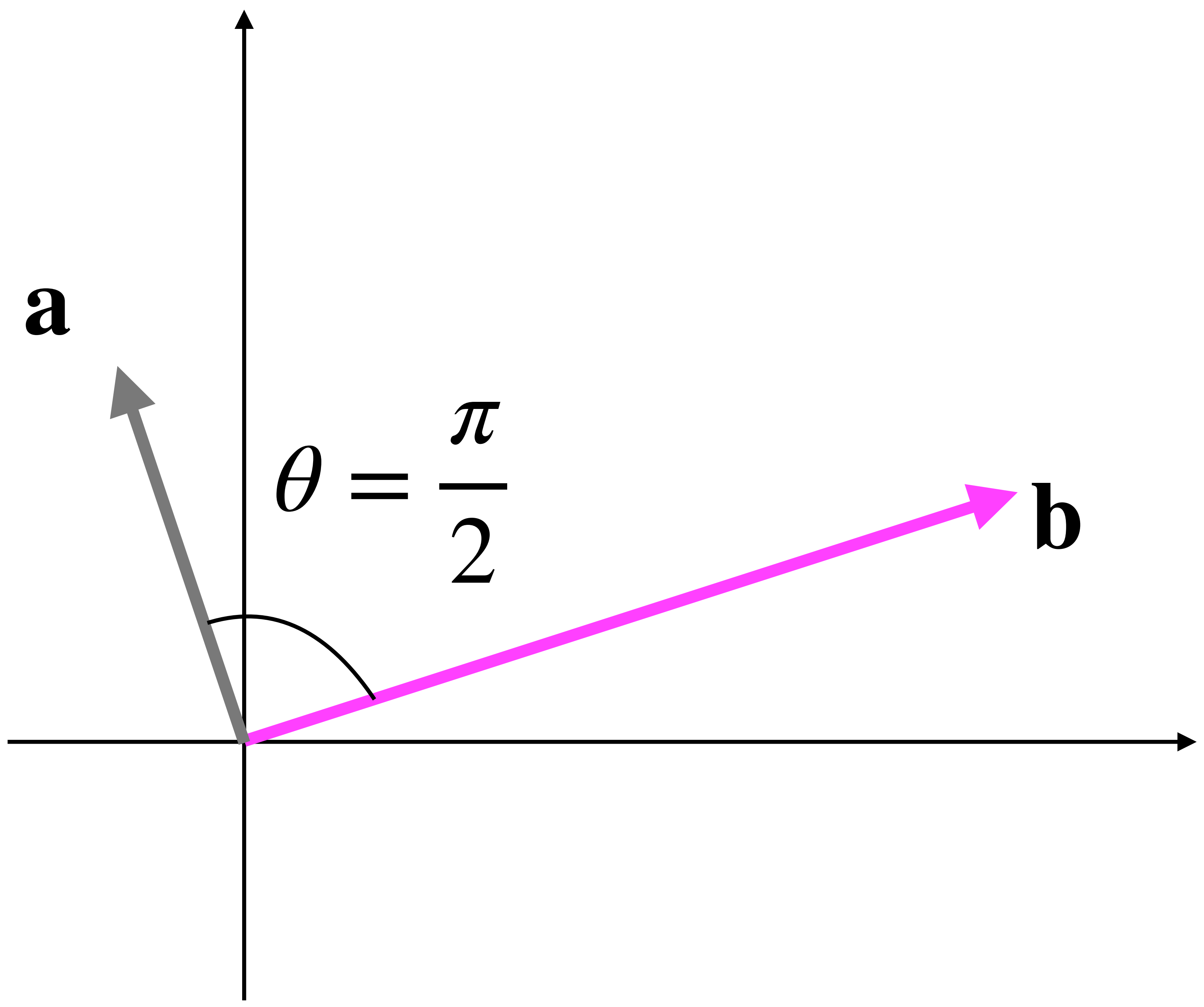
What happens if the two vectors are orthogonal?




$$\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\| \, \|\mathbf{b}\| \cos \frac{\pi}{2} = 0$$

# Inner/dot/scalar product

What happens if the two vectors are orthogonal?




$$\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\| \, \|\mathbf{b}\| \cos\frac{\pi}{2} = 0$$

The inner/dot/scalar product of two orthogonal vectors is zero!

# 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

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

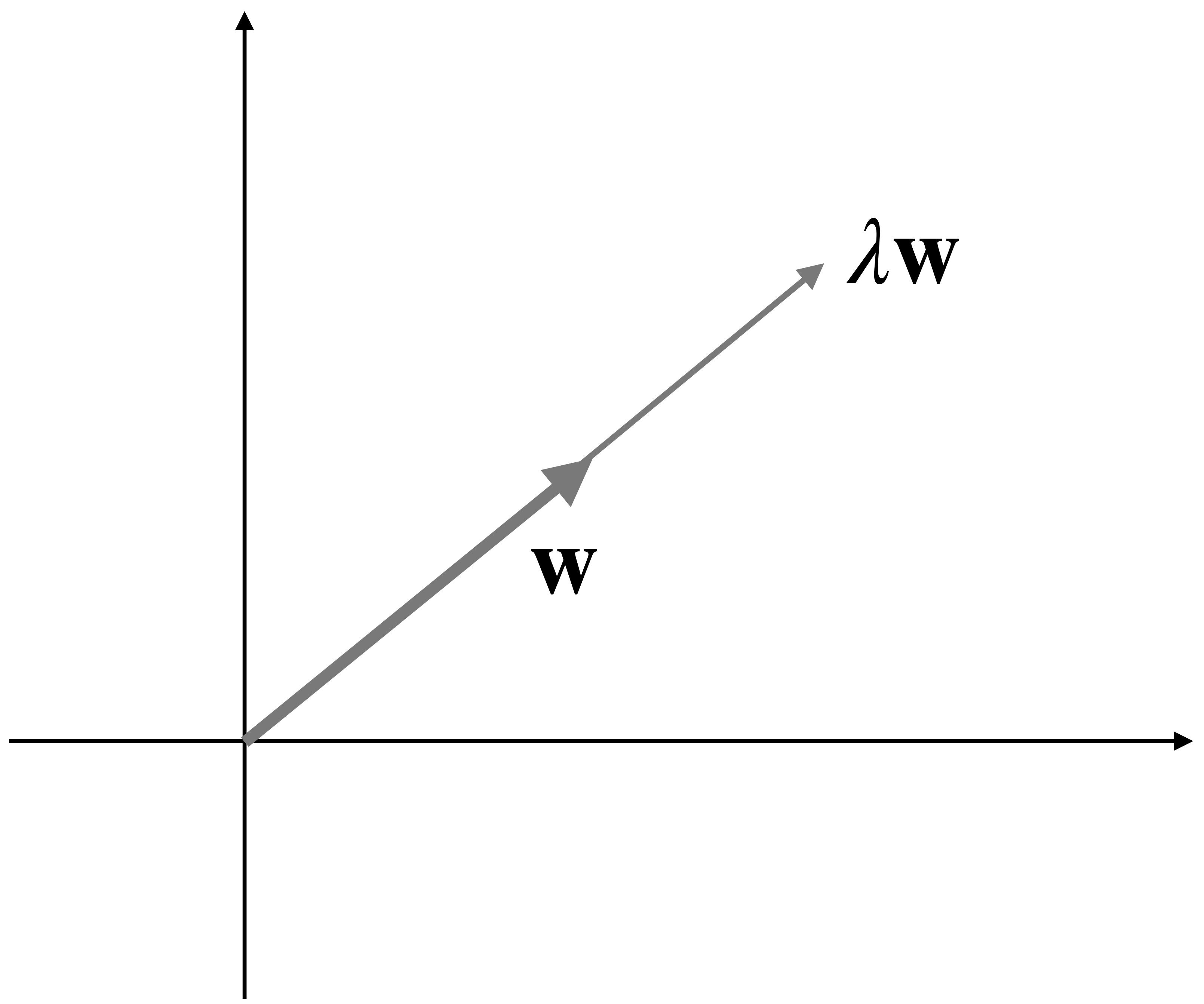
# Eigenvalues and Eigenvectors

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

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

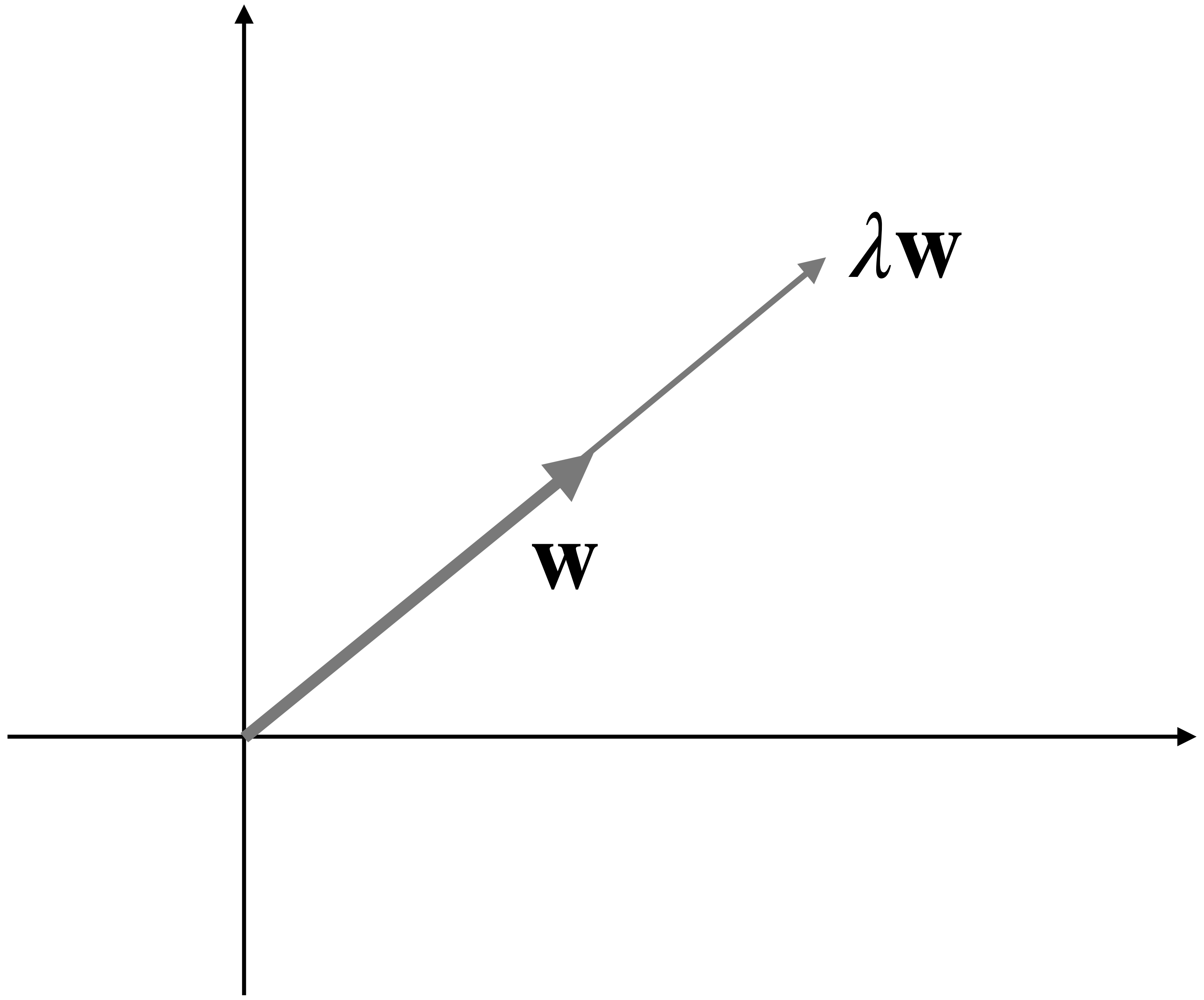
It is called eigenvector of the matrix and lambda is the correspondent eigenvalue

# Eigenvalues and Eigenvectors




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

# Eigenvalues and Eigenvectors




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

The matrix is a linear transformation that acts on w squeezing or stretching it

# Eigenvalues and Eigenvectors

The equation

$$\mathbf{X}\mathbf{w_i} = \lambda_i \mathbf{w_i}$$

# Eigenvalues and Eigenvectors

The equation

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

Can be written as

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

# Eigenvalues and Eigenvectors

The equation

$$\mathbf{X}\mathbf{w_i} = \lambda_i \mathbf{w_i}$$

Can be written as

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

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

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

Can be written as

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

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!

# Eigenvalues and Eigenvectors

Consider the matrix X such that

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

# Eigenvalues and Eigenvectors

Consider the matrix X such that

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

and let's assume that the eigenvectors are all linear independent

# Eigenvalues and Eigenvectors

Consider the matrix X such that

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

and let's assume that the eigenvectors are all linear independent

In this case, we can write the problem as

# Eigenvalues and Eigenvectors

Consider the matrix X such that

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

and let's assume that the eigenvectors are all linear independent

In this case, we can write the problem as

$$\mathbf{XW} = \mathbf{W\Lambda}$$

# Eigenvalues and Eigenvectors

Consider the matrix X such that

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

and let's assume that the eigenvectors are all linear independent

In this case, we can write the problem as

$$\mathbf{X}\mathbf{W} = \mathbf{W}\boldsymbol{\Lambda}$$

Where the matrix $\mathbf{W}$ contains, in the columns, the eigenvectors and $\boldsymbol{\Lambda}$ is a diagonal matrix whose entries are the eigenvalues

# Eigenvalues and Eigenvectors

$$\mathbf{XW} = \mathbf{W\Lambda}$$

# Eigenvalues and Eigenvectors

$$\mathbf{XW} = \mathbf{W\Lambda}$$

In these conditions, we know that the matrix $\mathbf{W}$ is invertible (all eigenvectors are linearly independent)

# Eigenvalues and Eigenvectors

$$\mathbf{XW} = \mathbf{W\Lambda}$$

In these conditions, we know that the matrix $\mathbf{W}$ is invertible (all eigenvectors are linearly independent)

$$\mathbf{X} = \mathbf{W\Lambda W}^{-1}$$

# Eigenvalues and Eigenvectors

$$\mathbf{XW} = \mathbf{W\Lambda}$$

In these conditions, we know that the matrix $\mathbf{W}$ is invertible (all eigenvectors are linearly independent)

$$\mathbf{X} = \mathbf{W\Lambda W}^{-1}$$

This is the eigendecomposition of the matrix

# Eigenvalues and Eigenvectors

We can write

$$\mathbf{X} = \mathbf{W}\boldsymbol{\Lambda}\mathbf{W}^{-1}$$

As

$$\mathbf{W}^{-1}\mathbf{X}\mathbf{W} = \boldsymbol{\Lambda}$$

# Eigenvalues and Eigenvectors

We can write

$$\mathbf{X} = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^{-1}$$

As

$$\mathbf{W}^{-1}\mathbf{X}\mathbf{W} = \mathbf{\Lambda}$$

The two matrices $\mathbf{X}$ and $\mathbf{\Lambda}$ are the same linear transformation expressed in two different basis

# Eigenvalues and Eigenvectors

We can write

$$\mathbf{X} = \mathbf{W}\boldsymbol{\Lambda}\mathbf{W}^{-1}$$

As

$$\mathbf{W}^{-1}\mathbf{X}\mathbf{W} = \boldsymbol{\Lambda}$$

The two matrices $\mathbf{X}$ and $\boldsymbol{\Lambda}$ are the same linear transformation expressed in two different basis

In the basis formed by the eigenvector the transformation is diagonal

# 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)

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 $\sigma_i^2$

# 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 $\sigma_i^2$

$\sigma_i$ are called singular values

# Singular value decomposition (SVD)

We define $\mathbf{U} \in \mathbb{R}^{s \times s}$ such that

$$\mathbf{U_i} = \sigma_{\mathbf{i}}^{-1}\mathbf{X}\mathbf{V_i}$$

# Singular value decomposition (SVD)

We define $\mathbf{U} \in \mathbb{R}^{s \times s}$ such that

$$\mathbf{U_i} = \sigma_{\mathbf{i}}^{-1} \mathbf{X} \mathbf{V_i}$$

Hence we have

$$\mathbf{X}^\top \mathbf{U_i} = \sigma_{\mathbf{i}}^{-1} \mathbf{X}^\top \mathbf{X} \mathbf{V_i} = \sigma_{\mathbf{i}} \mathbf{V_i}$$

# Singular value decomposition (SVD)

We define $\mathbf{U} \in \mathbb{R}^{s \times s}$ such that

$$\mathbf{U_i} = \sigma_\mathbf{i}^{-1}\mathbf{X}\mathbf{V_i}$$

Hence we have

$$\mathbf{X}^\top\mathbf{U_i} = \sigma_\mathbf{i}^{-1}\mathbf{X}^\top\mathbf{X}\mathbf{V_i} = \sigma_\mathbf{i}\mathbf{V_i}$$

Furthermore

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

# Singular value decomposition (SVD)

We define $\mathbf{U} \in \mathbb{R}^{s \times s}$ such that

$$\mathbf{U_i} = \sigma_\mathbf{i}^{-1}\mathbf{X}\mathbf{V_i}$$

Hence we have

$$\mathbf{X}^\top\mathbf{U_i} = \sigma_\mathbf{i}^{-1}\mathbf{X}^\top\mathbf{X}\mathbf{V_i} = \sigma_\mathbf{i}\mathbf{V_i}$$

Furthermore

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

$\mathbf{U_i}$ are the eigenvectors of $\mathbf{X}\mathbf{X}^\top$

# Singular value decomposition (SVD)

Summarizing

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

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

# Singular value decomposition (SVD)

We defined

$$\mathbf{U_i} = \sigma_{\mathbf{i}}^{-1}\mathbf{X}\mathbf{V_i}$$

# Singular value decomposition (SVD)

We defined

$$\mathbf{U_i} = \sigma_{\mathbf{i}}^{-1}\mathbf{X}\mathbf{V_i}$$

In matrix from

# Singular value decomposition (SVD)

We defined

$$\mathbf{U_i} = \sigma_{\mathbf{i}}^{-1}\mathbf{X}\mathbf{V_i}$$

In matrix from

$$\mathbf{U} = \mathbf{X}\mathbf{V}\boldsymbol{\Sigma}^{-1}$$

# Singular value decomposition (SVD)

We defined

$$\mathbf{U_i} = \sigma_{\mathbf{i}}^{-1}\mathbf{X}\mathbf{V_i}$$

In matrix from

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

$$\mathbf{U}\mathbf{\Sigma} = \mathbf{X}\mathbf{V}$$

n.perra@qmul.ac.uk

# Singular value decomposition (SVD)

We defined

$$\mathbf{U_i} = \sigma_{\mathbf{i}}^{-1}\mathbf{X}\mathbf{V_i}$$

In matrix from

$$\mathbf{U} = \mathbf{X}\mathbf{V}\boldsymbol{\Sigma}^{-1}$$

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

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

# Singular value decomposition (SVD)

We defined

$$\mathbf{U_i} = \sigma_{\mathbf{i}}^{-1}\mathbf{X}\mathbf{V_i}$$

In matrix from

$$\mathbf{U} = \mathbf{X}\mathbf{V}\boldsymbol{\Sigma}^{-1}$$

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

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

# 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

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

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

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

# 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

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

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

# Calculus

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

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

# Calculus

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

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

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

# Calculus

We can extend this to functions $f : \mathbb{R}^d \to \mathbb{R}^s$ with multiple outputs via the Jacobian matrix $J_f : \mathbb{R}^d \to \mathbb{R}^{s \times d}$ defined as

$$J_f(x) := \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_d} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_s}{\partial x_1} & \cdots & \frac{\partial f_s}{\partial x_d} \end{pmatrix}$$

# Calculus

We can extend this to functions $f : \mathbb{R}^d \to \mathbb{R}^s$ with multiple outputs via the Jacobian matrix $J_f : \mathbb{R}^d \to \mathbb{R}^{s \times d}$ defined as

$$J_f(x) := \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_d} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_s}{\partial x_1} & \cdots & \frac{\partial f_s}{\partial x_d} \end{pmatrix}$$

And we can even define a second-order derivative matrix (known as Hessian) via

$$H_f(x) := J_{\nabla f}(x)$$

# 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!

n.perra@qmul.ac.uk

# 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 $\rho$. 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 $\rho(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$

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

# 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

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

# 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

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

# TUTORIAL ON FRIDAY

We will discuss the solutions of Coursework 0

To make the most of these tutorials, attempt completing the coursework before!