

Now for some machine learning...

What is Machine Learning?

Talk to your neighbor. (Say hello first.)

Give a short answer to: “What is Machine Learning?”

What is Machine Learning?

Talk to your neighbor. (Say hello first.)

Give a short answer to: “What is Machine Learning?”

(I mean it 😊)

What is Machine Learning?

Talk to your neighbor. (Say hello first.)

Give a short answer to: “What is Machine Learning?”

What is Machine Learning?

Machine learning

Article [Talk](#)

From Wikipedia, the free encyclopedia

For the journal, see [Machine Learning \(journal\)](#).

"Statistical learning" redirects here. For statistical learning in linguistics, see [stat](#)

Machine learning (ML) is a [field of study](#) in [artificial intelligence](#) concerned with the development and study of [statistical algorithms](#) that can learn from [data](#) and [generalize](#) to unseen data, and thus perform [tasks](#) without explicit [instructions](#).^[1]

Recently, [artificial neural networks](#) have been able to surpass many previous approaches in performance.^[2]

What is Machine Learning?

Machine learning

Article [Talk](#)

From Wikipedia, the free encyclopedia

For the journal, see [Machine Learning \(journal\)](#).

"Statistical learning" redirects here. For statistical learning in linguistics, see [stat](#)

Machine learning (ML) is a [field of study](#) in [artificial intelligence](#) concerned with the development and study of [statistical algorithms](#) that can learn from [data](#) and [generalize](#) to unseen data, and thus perform [tasks](#) without explicit [instructions](#).^[1]

Recently, [artificial neural networks](#) have been able to surpass many previous approaches in performance.^[2]

Model

What is Machine Learning?

Machine learning

Article [Talk](#)

From Wikipedia, the free encyclopedia

For the journal, see [Machine Learning \(journal\)](#).

"Statistical learning" redirects here. For statistical learning in linguistics, see [stat](#)

Machine learning (ML) is a [field of study](#) in [artificial intelligence](#) concerned with the development and study of [statistical algorithms](#) that can learn from [data](#) and [generalize](#) to unseen data, and thus perform [tasks](#) without explicit [instructions](#).^[1]

Recently, [artificial neural networks](#) have been able to surpass many previous approaches in performance.^[2]

Data → Trained Model

What is Machine Learning?

Machine learning

Article [Talk](#)

From Wikipedia, the free encyclopedia

For the journal, see [Machine Learning \(journal\)](#).

"Statistical learning" redirects here. For statistical learning in linguistics, see [stat](#)

Machine learning (ML) is a [field of study](#) in [artificial intelligence](#) concerned with the development and study of [statistical algorithms](#) that can learn from [data](#) and [generalize](#) to unseen data, and thus perform [tasks](#) without explicit [instructions](#).^[1]

Recently, [artificial neural networks](#) have been able to surpass many previous approaches in performance.^[2]



What is Machine Learning?

Machine learning

Article [Talk](#)

From Wikipedia, the free encyclopedia

For the journal, see [Machine Learning \(journal\)](#).

"Statistical learning" redirects here. For statistical learning in linguistics, see [stat](#)

Machine learning (ML) is a [field of study](#) in [artificial intelligence](#) concerned with the development and study of [statistical algorithms](#) that can learn from [data](#) and [generalize](#) to unseen data, and thus perform [tasks](#) without explicit [instructions](#).^[1]

Recently, [artificial neural networks](#) have been able to surpass many previous approaches in performance.^[2]

Machine Learning is fancy statistics!

What is Machine Learning?

Machine learning

Article [Talk](#)

From Wikipedia, the free encyclopedia

For the journal, see [Machine Learning \(journal\)](#).

"Statistical learning" redirects here. For statistical learning in linguistics, see [stat](#)

Machine learning (ML) is a [field of study](#) in [artificial intelligence](#) concerned with the development and study of [statistical algorithms](#) that can learn from [data](#) and [generalize](#) to unseen data, and thus perform [tasks](#) without explicit [instructions](#).^[1]

Recently, [artificial neural networks](#) have been able to surpass many previous approaches in performance.^[2]

Fitting a straight line

Principal component analysis

K nearest neighbors

Machine Learning is fancy statistics!

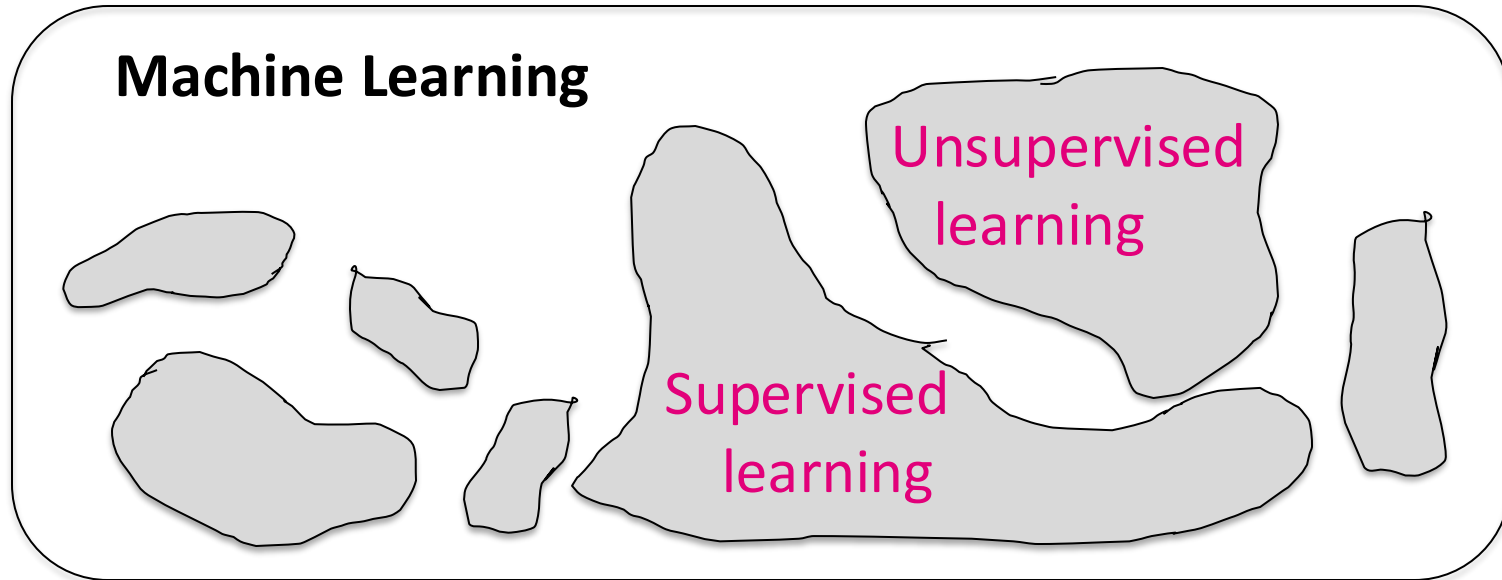
Naïve Bayes

Neural networks

Support vector machines

Decision trees

Supervised vs Unsupervised learning



Supervised vs Unsupervised learning

In **supervised** methods we know the outcome Y for each input X .

- Often the task is to predict Y based on X .

In **unsupervised** methods we have only the inputs X .

- Typical tasks are about finding structure in data.

Supervised vs Unsupervised learning

Given 50 images labelled as cats or dogs, build a predictor that can label a new image.

Example: supervised setting

Supervised vs Unsupervised learning

We have the blood pressure of 100 patients along with information about height, weight, and sex.

- What is the predicted blood pressure of a new patient?
- What information is important for the prediction?
- How does the blood pressure vary with weight?

Example of supervised setting

Supervised vs Unsupervised learning

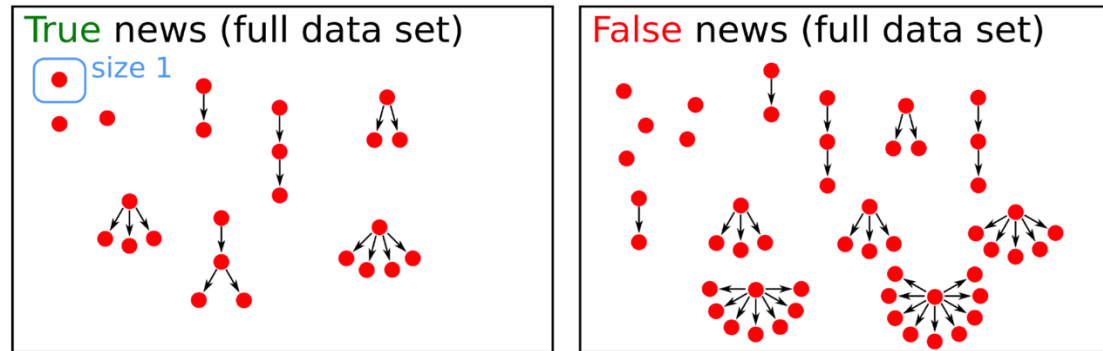
Exploring 500 images of roses, can we identify some natural clusters of similar roses?

Example of **unsupervised** setting

In my research...

I use both supervised and unsupervised methods to approach research questions.

Can an algorithm recognize what spreading Twitter stories are true or false if I train it on past examples of true/false stories spreading?



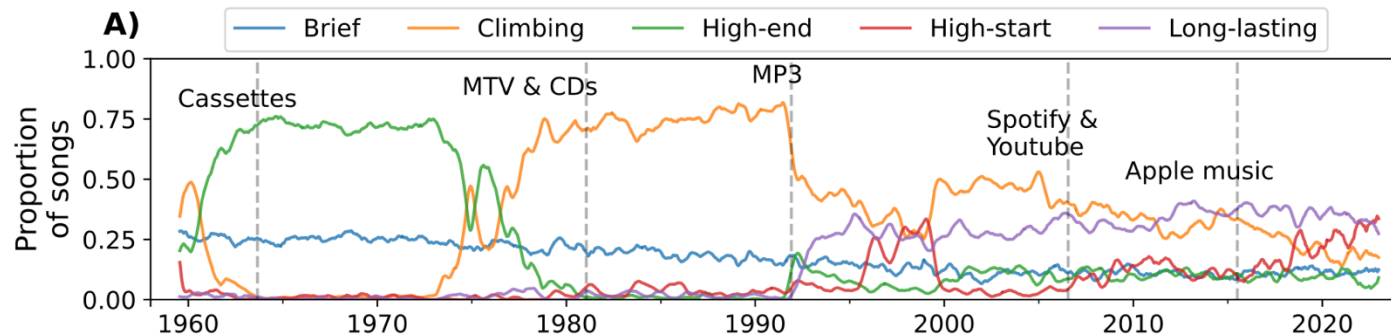
Example of **supervised** setting

(Juul & Ugander, 2021)

In my research...

I use both supervised and unsupervised methods to approach research questions.

Showing how pop songs have moved up and down the Billboard Hot 100 chart, can we find out whether “archetypes” of pop hits exist?



Example of **unsupervised** setting

(Lech et al., 2024)

Terminology

Note that a lot of different terminology is used for the same concepts. Among others,

Variable(s) X:

Input

Feature

Predictor

Covariate

Independent variable

Variable Y:

Output

Outcome

Response

Target variable

Dependent variable

Quantitative vs Qualitative

Quantitative

- BMI
- Annual wage

Qualitative (Categorical)

- Yes/no
- Cat/dog/fish
- Housing type
- Severeness of illness

For categorical variables, the labels have no numerical interpretation even if numerical. Sometimes labels are naturally ordered (e.g. low/medium/high).

Regression vs Classification

Supervised problems are often referred to as either

Regression: The outcome Y is quantitative (typically \mathbb{R})

Classification: The outcome Y is categorical

The important thing is to choose an ML method suitable for your specific outcome.

Why Machine Learning?

Why Machine Learning?

- A **revolution** is taking place (maybe).
- Machine learning algorithms (will) affect almost **every part** of our lives.
- Lots of **open questions**, new ones appearing all the time.
- Lots of **job opportunities**, opportunities to do good.

Why Machine Learning?

- A **revolution** is taking place (maybe).
- Machine learning algorithms (will) affect almost **every part** of our lives.
- Lots of **open questions**, new ones appearing all the time.
- Lots of **job opportunities**, opportunities to do good.

Ready?