MACHINE LEARNING COURSE

PRESENTED BY ABDEL RAHMAN ALSABBAGH

LECTURE #1 – TUE - 10.5.2023

In the name of Allah, the most gracious, the most merciful, we start :)

Let's do some EDA
(´⊙ω⊙`)

Today's Quote

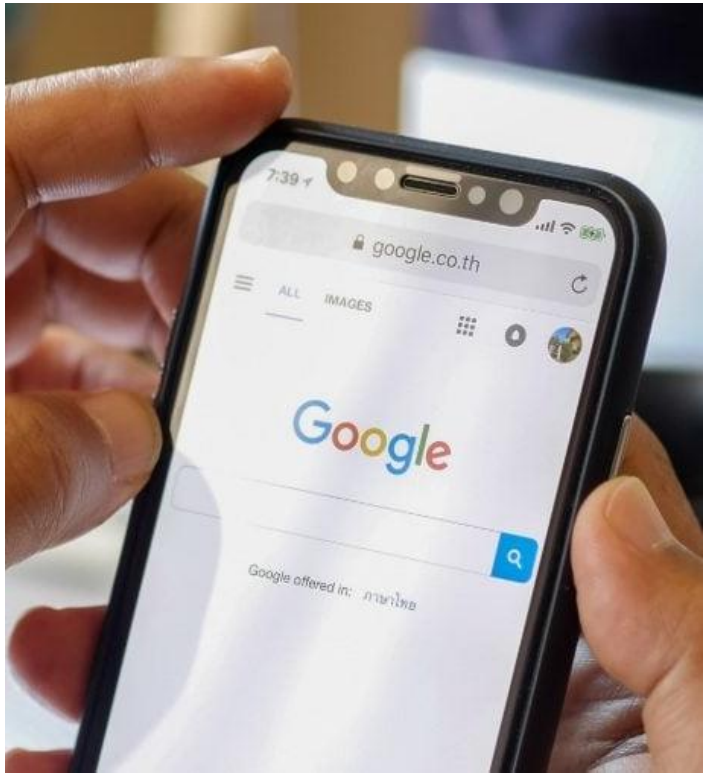
"The man who asks a question is a fool for a minute,
the man who does not ask is a fool for life."
- Confucius

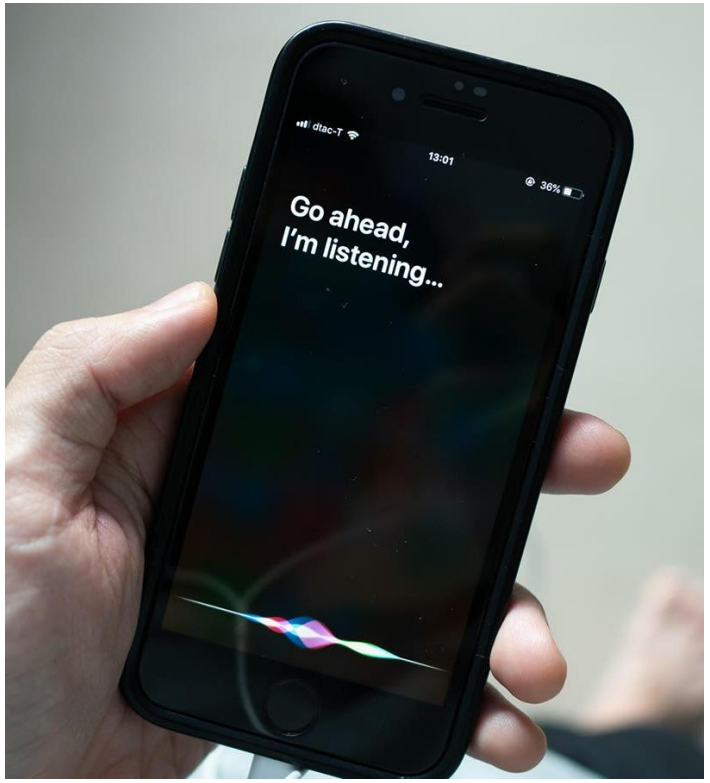
Motivation

When most people hear “Machine Learning,” they picture a robot: a dependable butler or a deadly Terminator depending on who you ask.

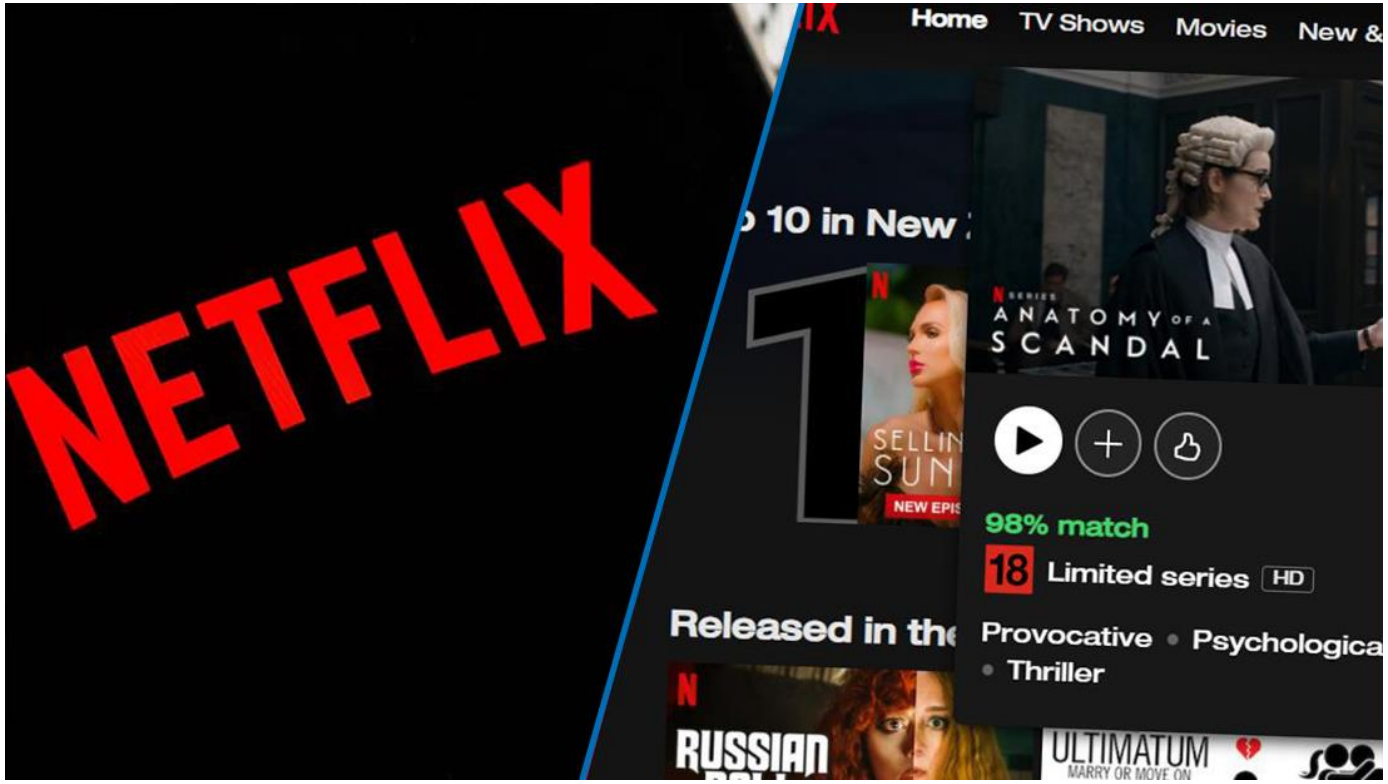
But Machine Learning is not just a futuristic fantasy,
it’s already here.

Motivation – You Use ML Everyday





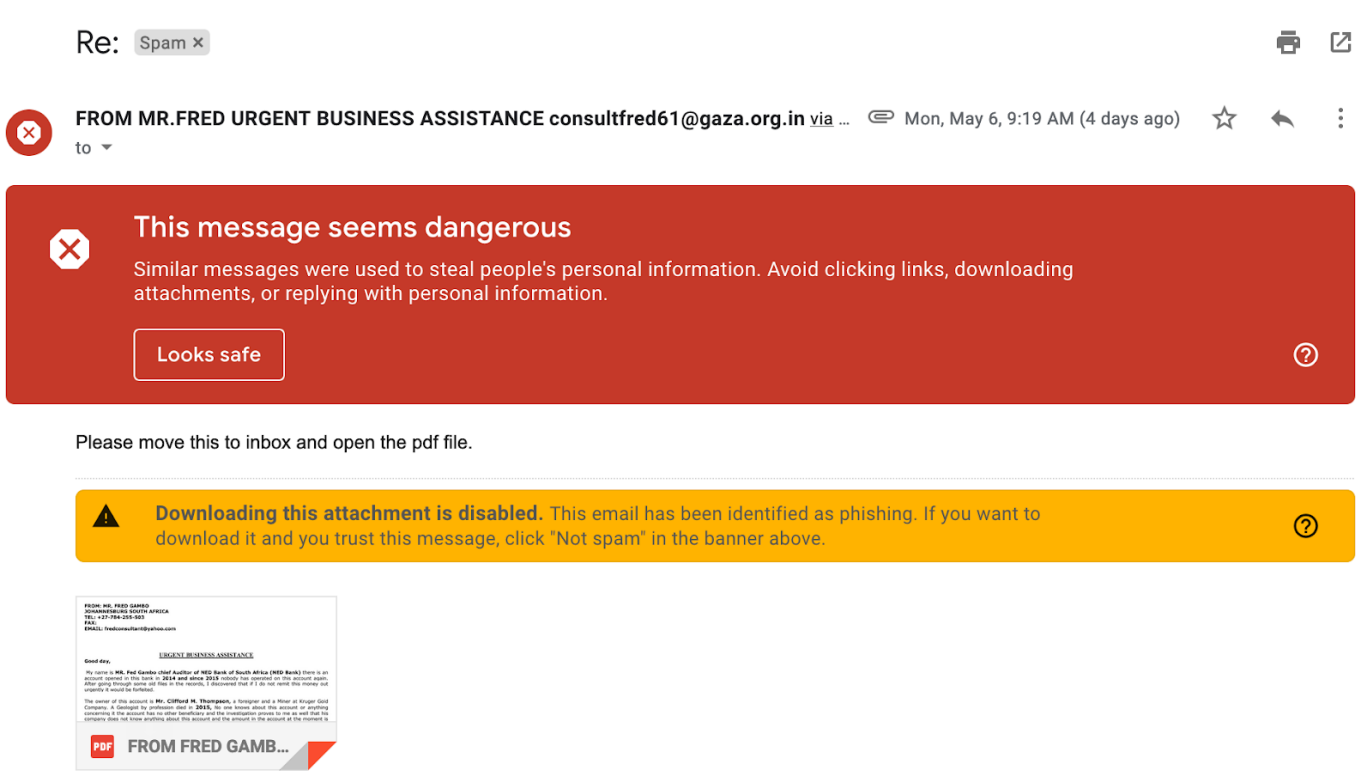
Motivation – You Use ML Everyday

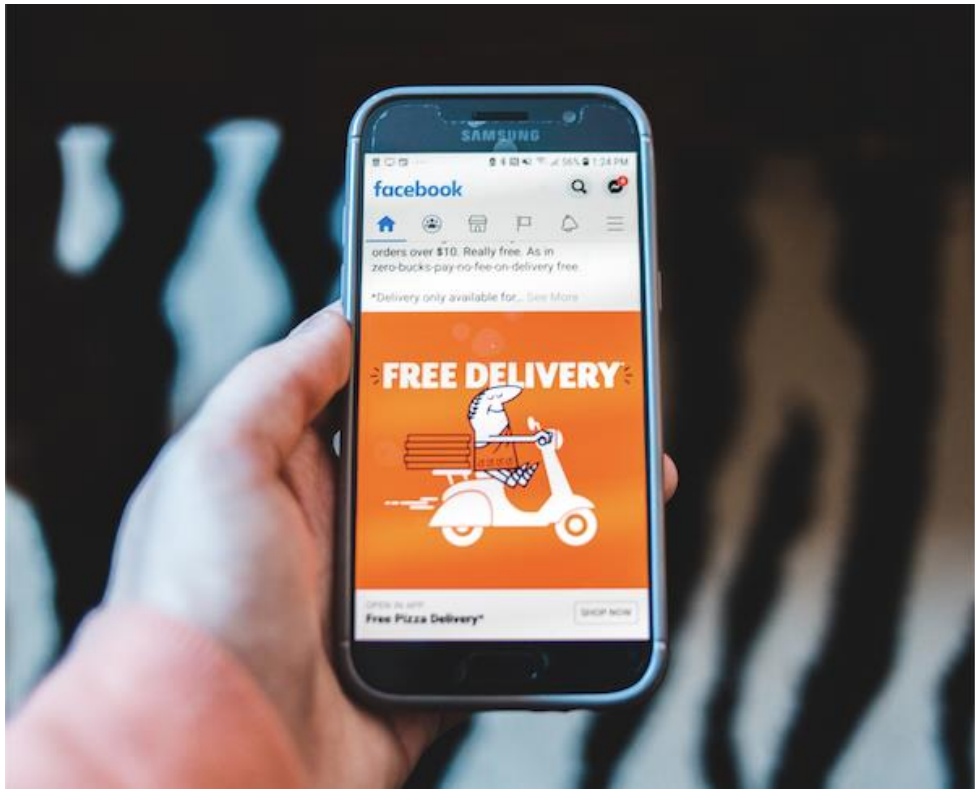


Nobody:

Youtube recommendations at 3am

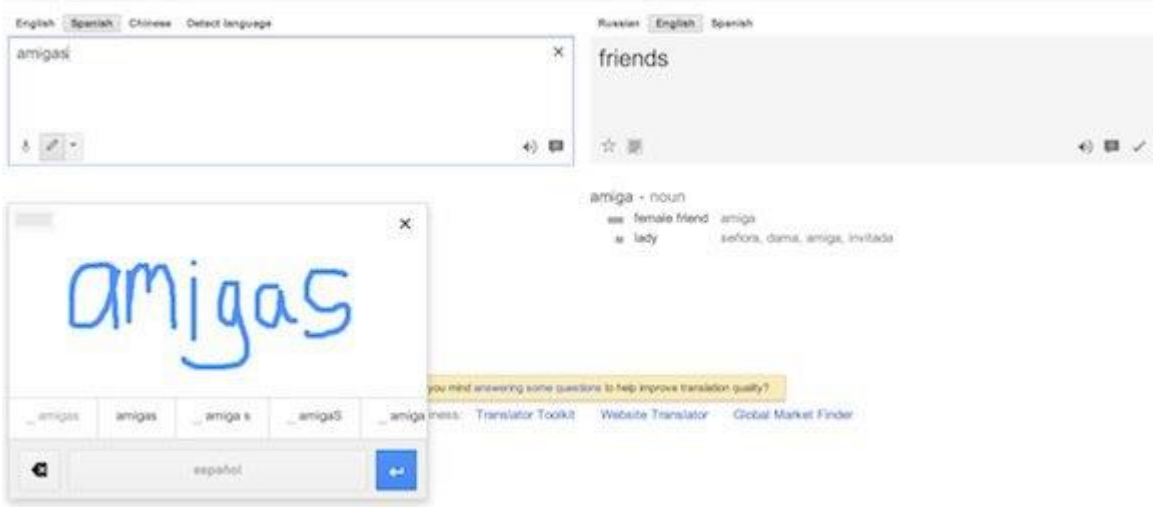
Motivation – You Use ML Everyday



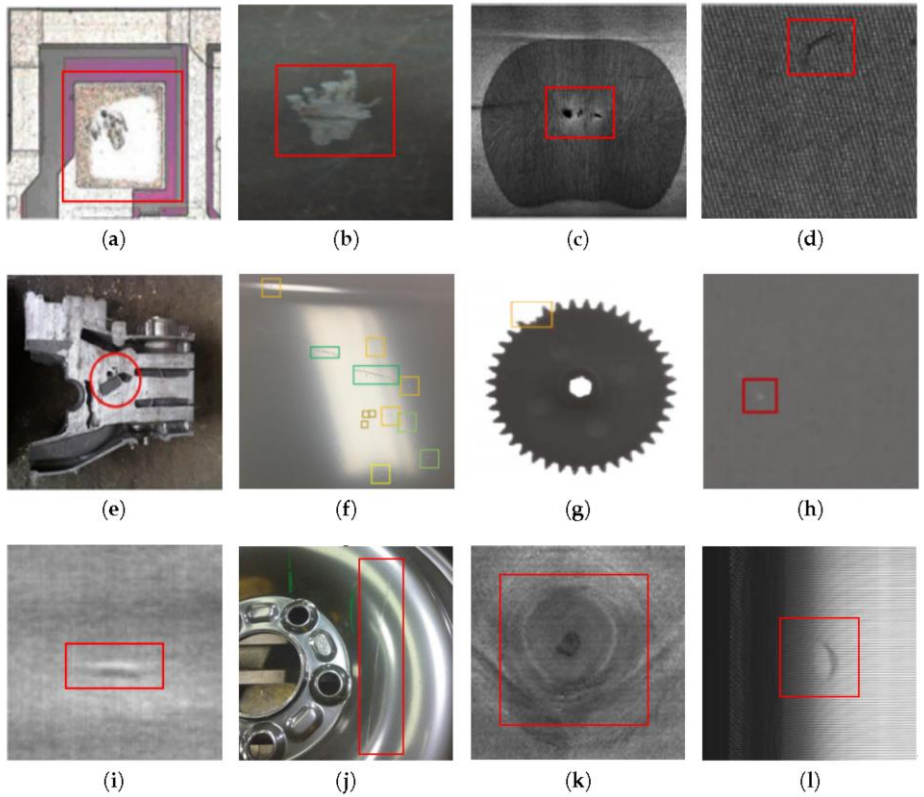


Motivation – Other ML Applications

Motivation – Other ML Applications



Motivation – Other ML Applications



Motivation – Other ML Applications

Motivation – Other ML Applications

Motivation – Other ML Applications

Can you think of one?

So, What is Machine Learning in a sentence?

It's the science of getting computers to learn without being **explicitly** programmed.

Prerequisites

We expect you to have knowledge in:

- Programming, specifically in Python programming language.
- Data preprocessing techniques.
- Principles of statistics, calculus, and linear algebra.

Grading

In order to get a certificate for this course, you should pass the grade of 85%, that is through completion of:

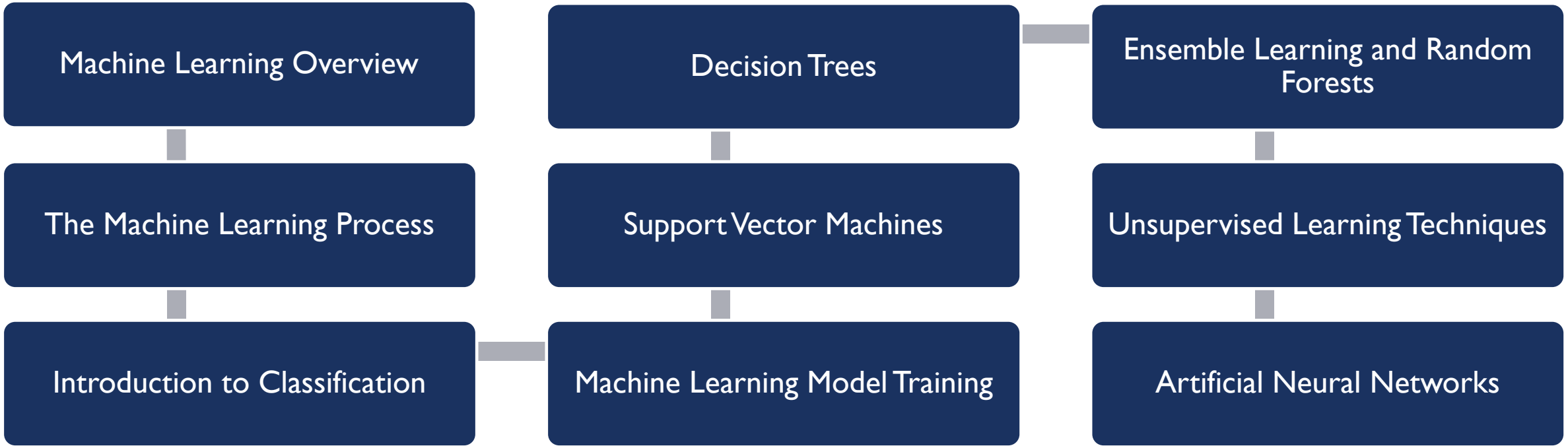
- Assignments (50%).
- Attendance (20%).
- Final Project (30%).

Expectations

You should expect us to provide the knowledge you need to start your journey in Machine Learning.

We expect from you **commitment**, and enthusiasm.

Course Outline



Resources used:

- Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow by Aurélien Géron
- Machine Learning Specialization by Andrew Ng and Stanford Online

Machine Learning Overview

Topics covered in this section include:

- What Is Machine Learning.
- Why Use Machine Learning?.
- Types of Machine Learning Systems.
- Main Challenges of Machine Learning.
- Testing and Validating.
- Exercises.

What Is Machine Learning?

Machine Learning is the science (and art) of programming computers so they can learn from *data*.

A slightly more general definition:

The field of study that gives computers the ability to learn without being explicitly programmed.

- Arthur Samuel, 1959

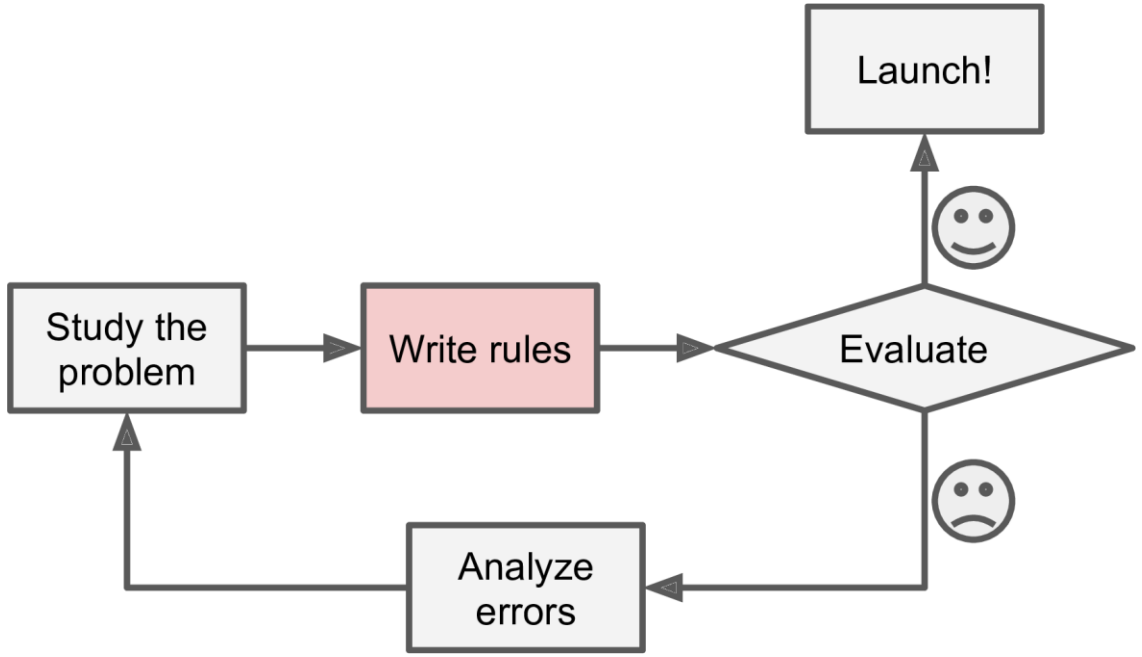
A more engineering-oriented one:

A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.

- Tom Mitchell, 1997

Why Use Machine Learning?

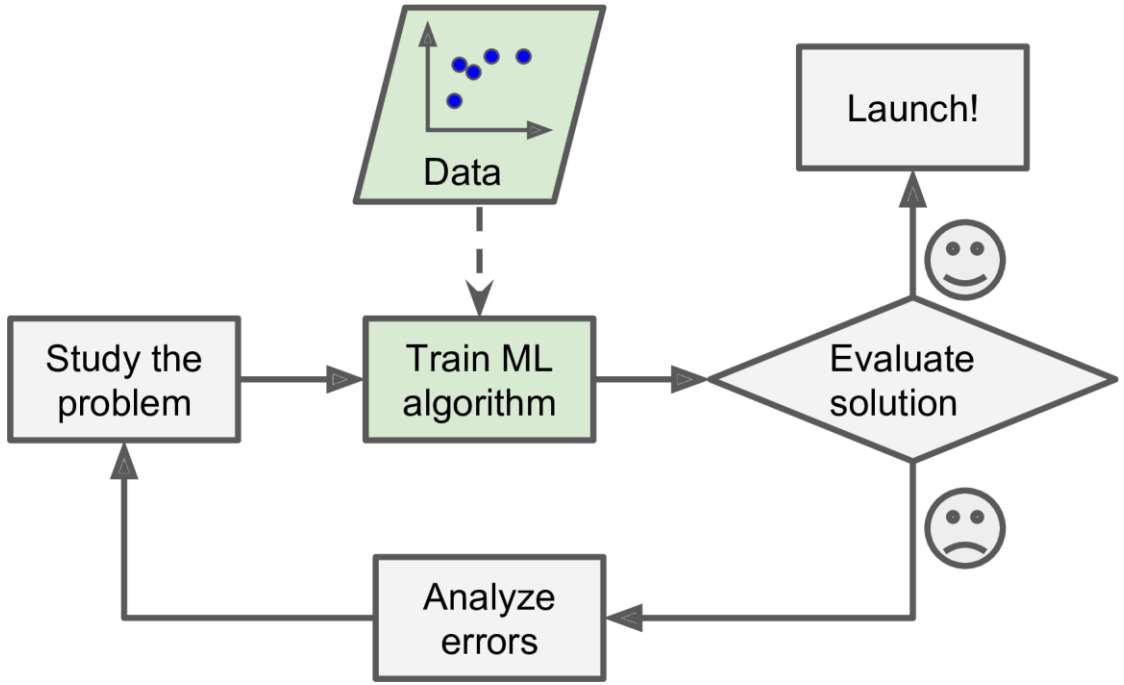
Let's take the example of spam filtering in both **traditional** approach and **Machine Learning** approach.



Traditional approach

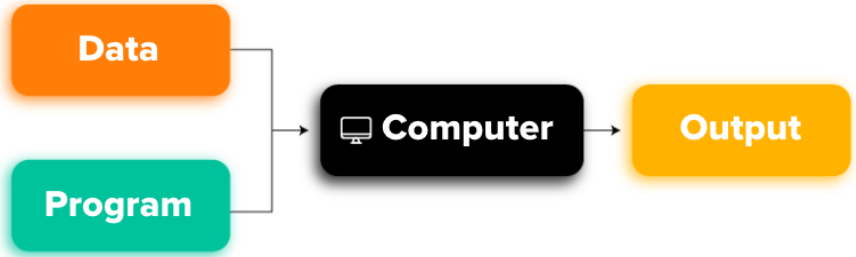
Why Use Machine Learning?

Let's take the example of spam filtering in both **traditional** approach and **Machine Learning** approach.

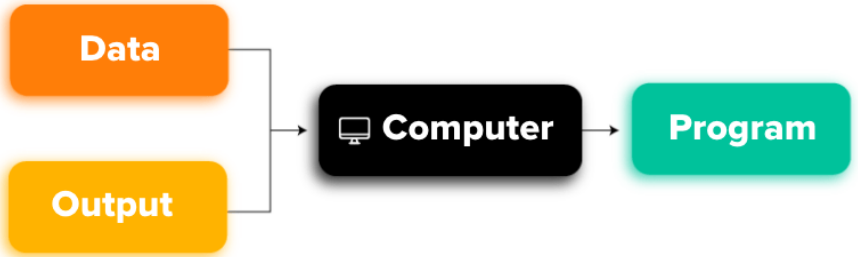


Machine Learning approach

Why Use Machine Learning?



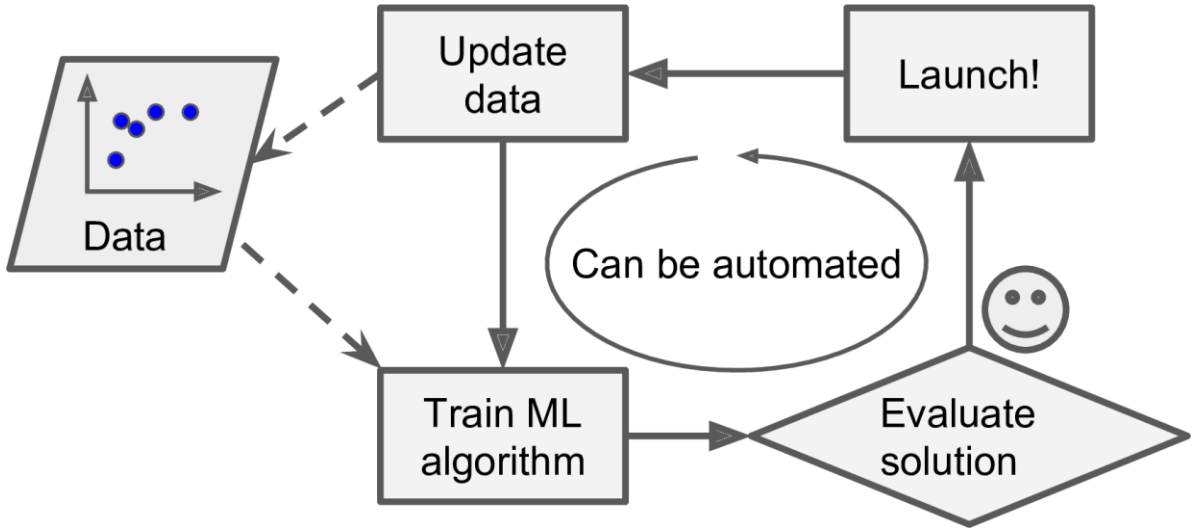
Traditional approach



Machine Learning approach

Why Use Machine Learning?

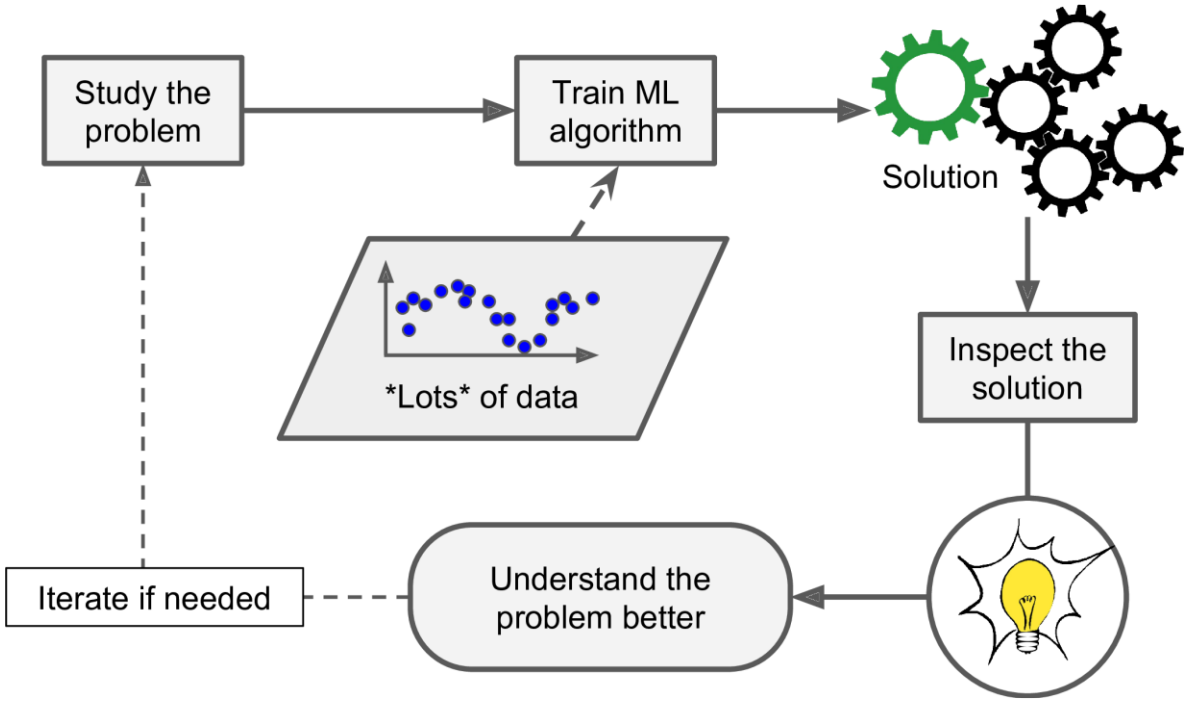
Let's take the example of spam filtering in both **traditional** approach and **Machine Learning** approach.



Automatically adapting to change

Why Use Machine Learning?

Let's take the example of spam filtering in both **traditional** approach and **Machine Learning** approach.



Machine Learning can help humans learn

Why Use Machine Learning?

To summarize, **Machine Learning is great for:**

- Problems for which existing solutions require a lot of hand-tuning or long lists of rules: one Machine Learning algorithm can often simplify code and perform better.
- Complex problems for which there is no good solution at all using a traditional approach: the best Machine Learning techniques can find a solution.
- Fluctuating environments: a Machine Learning system can adapt to new data.
- Getting insights about complex problems and large amounts of data.

Types of Machine Learning Systems

There are so many different **types of Machine Learning systems** that it is useful to classify them in broad categories based on:

- Whether or not they are trained with human supervision (supervised, unsupervised, semisupervised, and Reinforcement Learning)
- Whether or not they can learn incrementally on the fly (online versus batch learning)
- Whether they work by simply comparing new data points to known data points, or instead detect patterns in the training data and build a predictive model, much like scientists do (instance-based versus model-based learning)

Supervised/Unsupervised Learning

Machine Learning systems can be classified according to the amount and type of supervision they get during training.

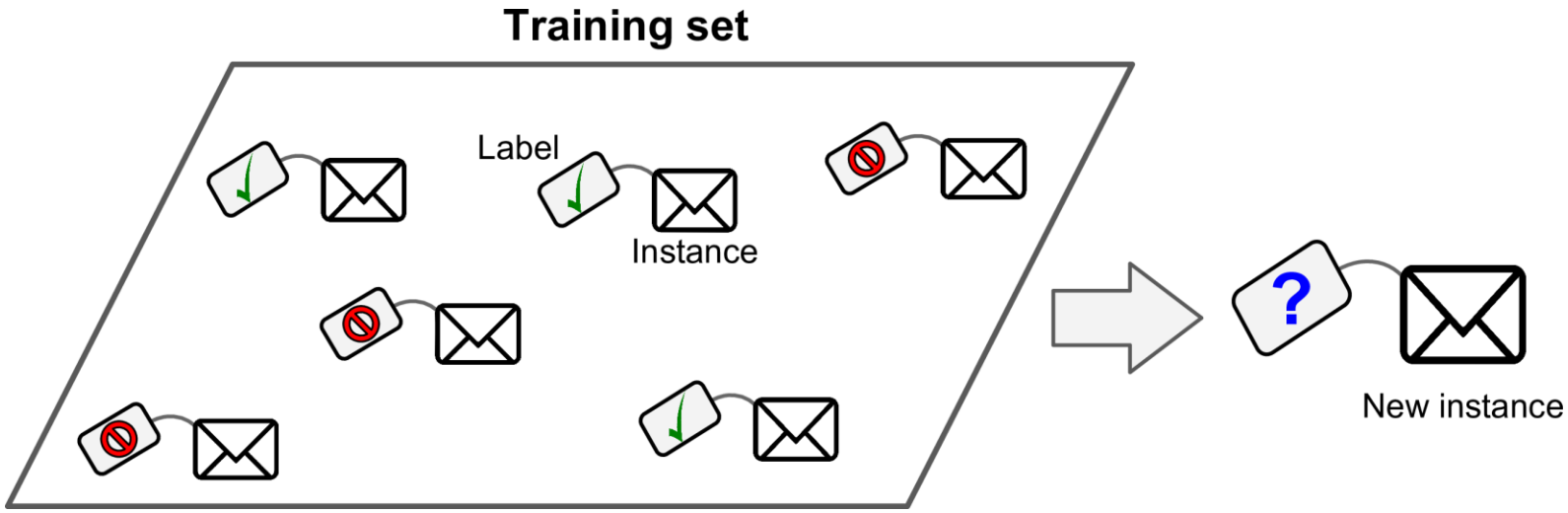
There are four major categories:

- Supervised learning
- Unsupervised learning
- Semisupervised learning
- Reinforcement learning

Supervised Learning

In **supervised learning**, the training data you feed to the algorithm includes the desired solutions, called ***labels***.

This type of supervised learning task is called **classification**. Where we train the machine to what *class* should it *classify* a new instance.

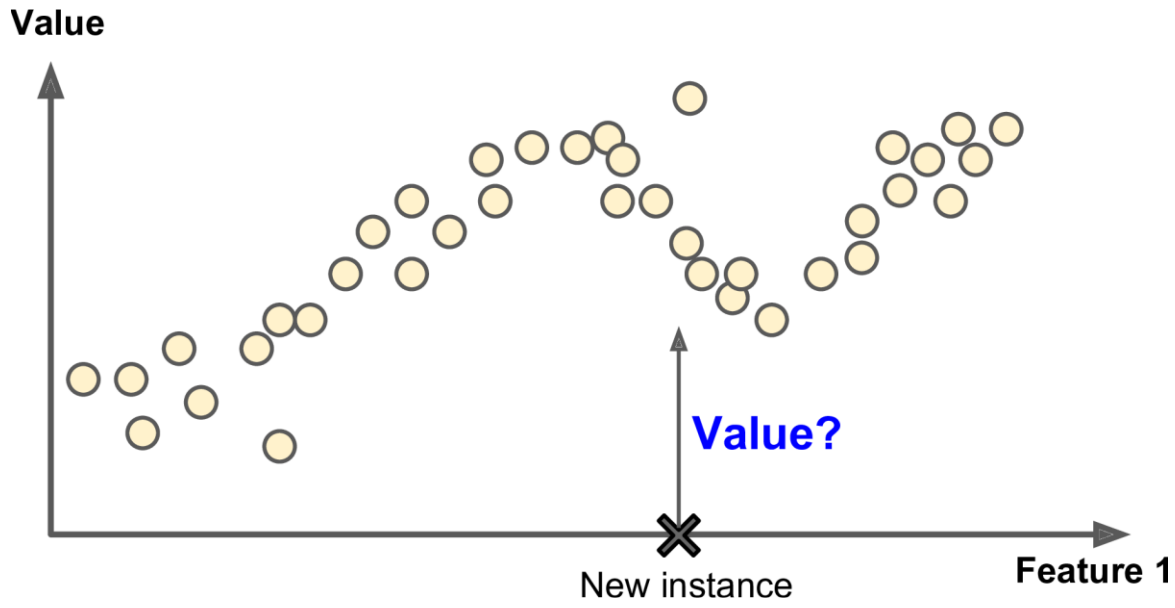


Supervised Learning

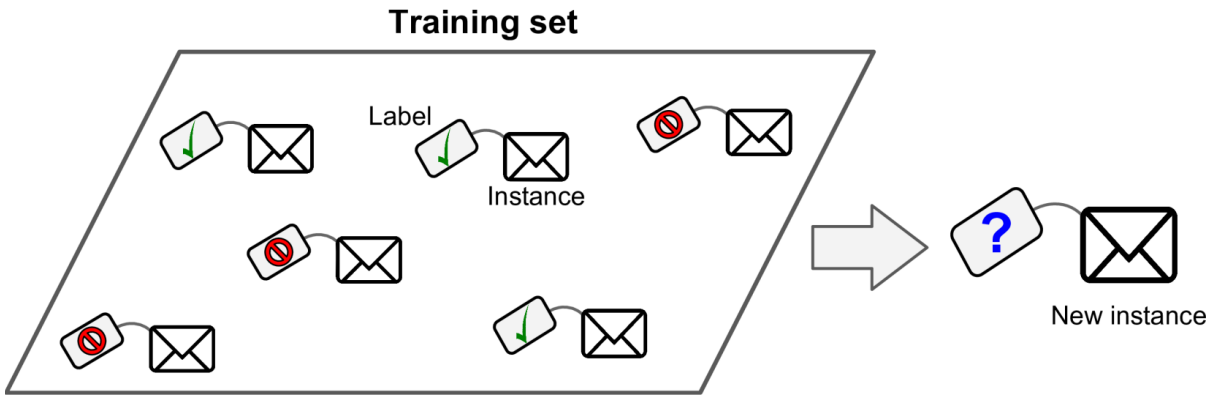
Another typical task is to predict a target numeric value, such as the price of a car, given a set of features (mileage, age, brand, etc.) called *predictors or attributes*.

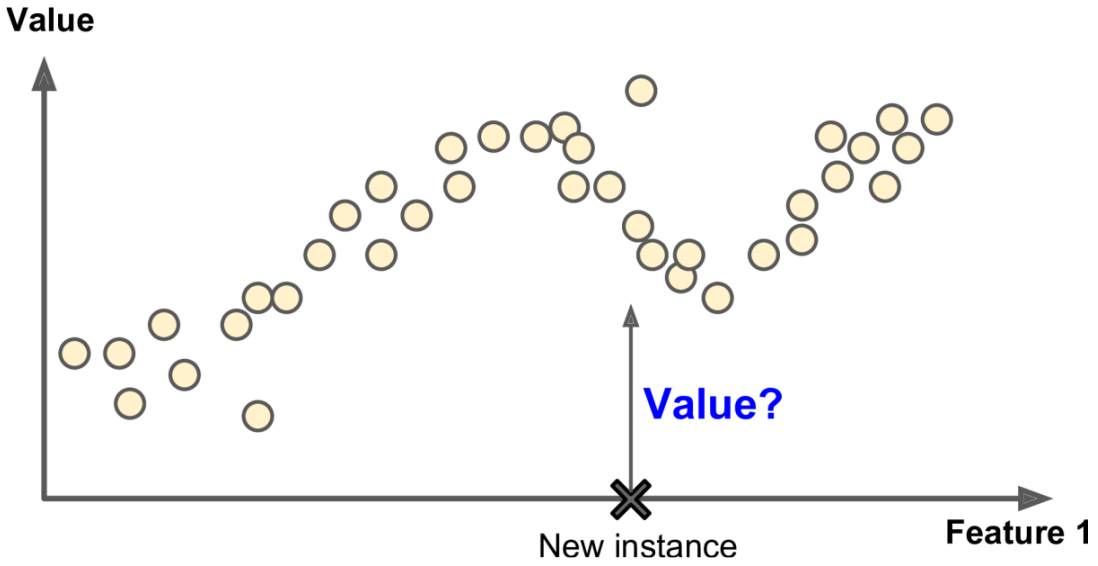
This sort of task is called **regression**.

To train the system, you need to give it many examples of cars, including both their predictors and their labels (i.e., their prices).



Supervised Learning





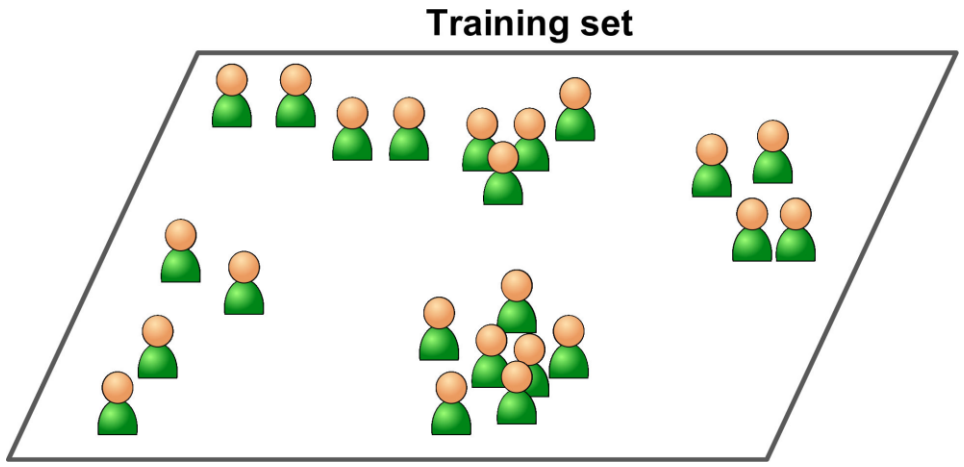
In **classification** we predict **a discrete value, a class**. In **regression** we predict a **continuous value, a number**.

Unsupervised Learning

I think it's your time to guess, what is **unsupervised learning?**
Its where the training data is **unlabeled,** and the system tries to learn without a teacher.

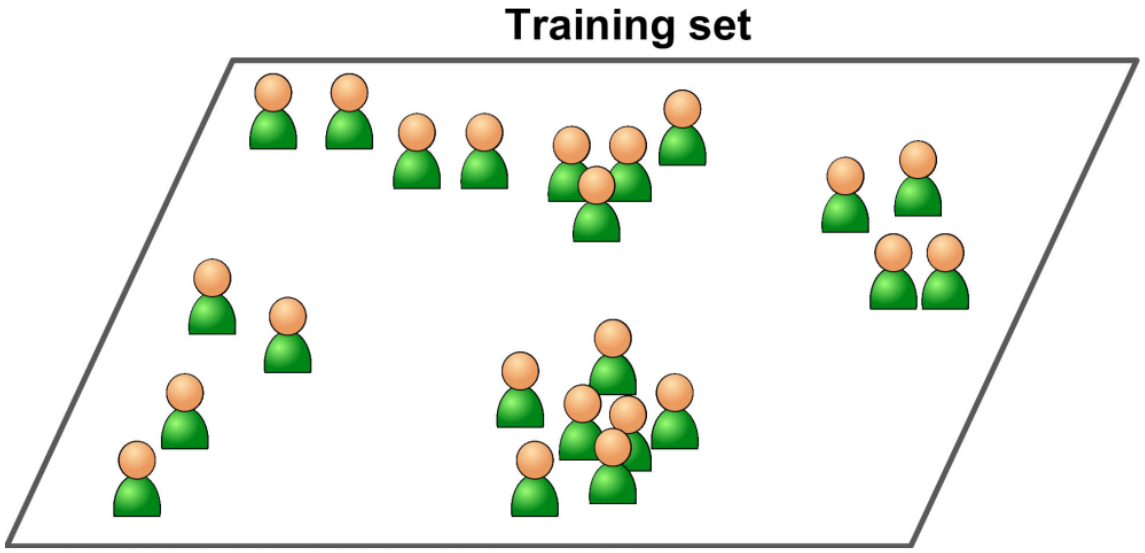
How can we benefit from unlabeled data you might ask? Through the following techniques:

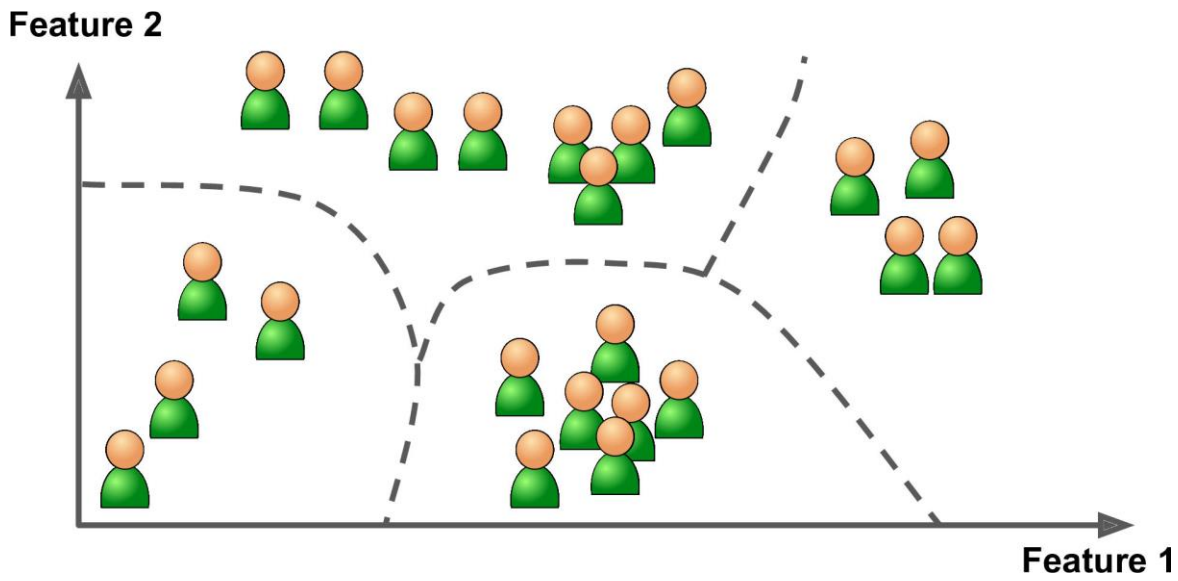
- **Clustering**



Clustering

You may want to run a clustering algorithm to try to detect groups of similar features.



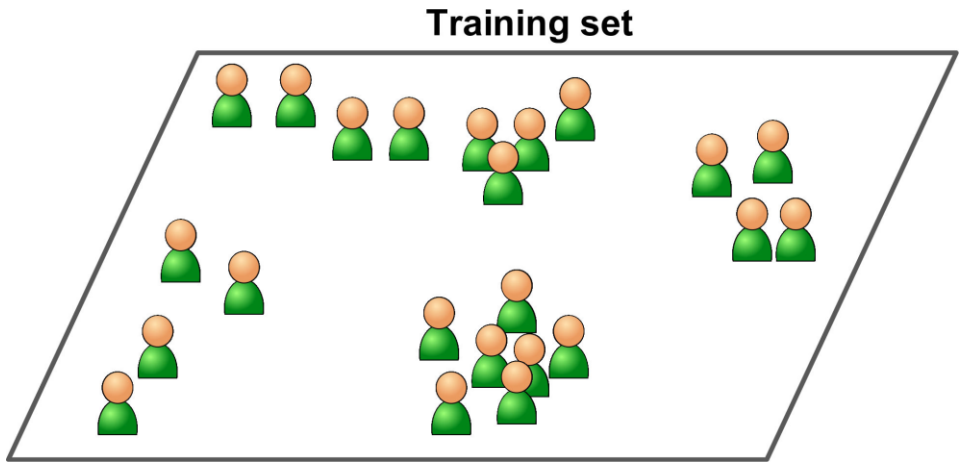


Unsupervised Learning

I think it's your time to guess, what is **unsupervised learning?**
Its where the training data is **unlabeled,** and the system tries to learn without a teacher.

How can we benefit from unlabeled data you might ask? Through the following techniques:

- Clustering
- **Anomaly detection and novelty detection**



Anomaly Detection and Novelty Detection

A system that is shown mostly normal instances during training, so it learns to recognize them and when it sees a new instance it can tell whether it looks like a normal one or whether it is likely an anomaly.

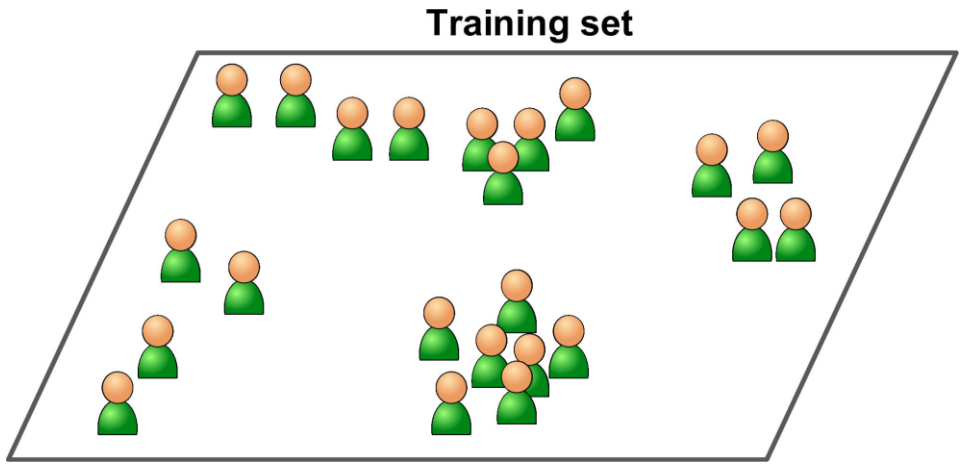
Anomaly detection vs. **novelty** detection? Novelty detection algorithms expect to see only normal data during training, while anomaly detection algorithms are usually more tolerant, they can often perform well even with a small percentage of outliers in the training set.

Unsupervised Learning

I think it's your time to guess, what is **unsupervised learning?**
Its where the training data is **unlabeled,** and the system tries to learn without a teacher.

How can we benefit from unlabeled data you might ask? Through the following techniques:

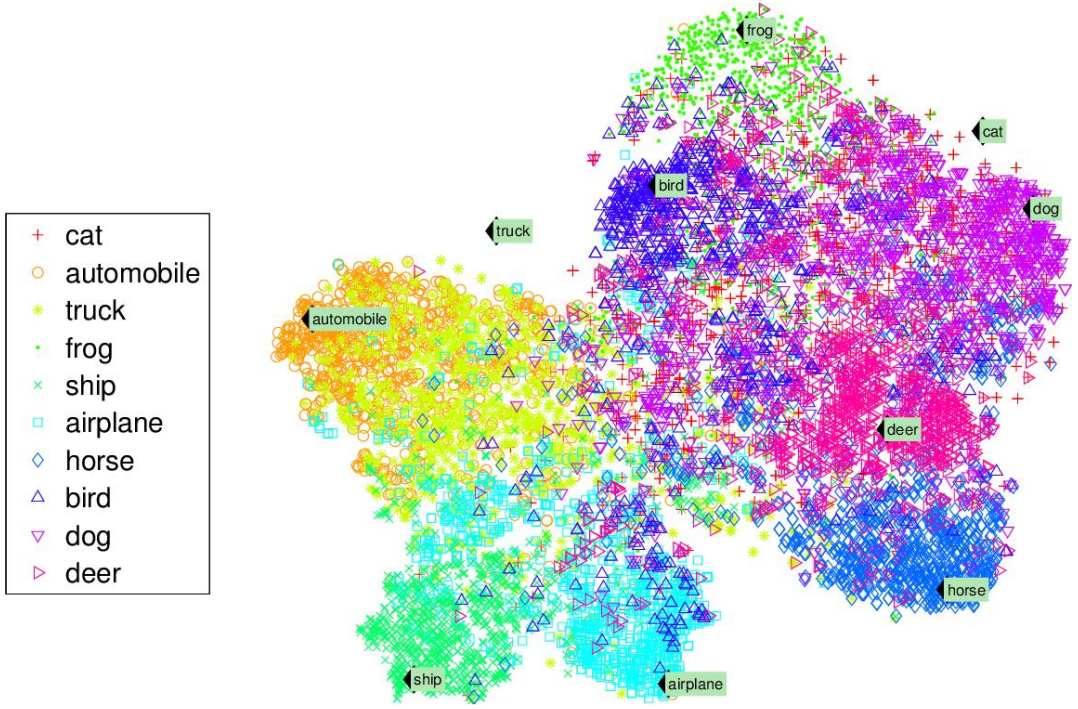
- Clustering
- Anomaly detection and novelty detection
- **Visualization and dimensionality reduction**



Visualization and dimensionality reduction

Visualization algorithms output a 2D or 3D representation of your data that can easily be plotted. These algorithms try to preserve as much structure as they can so you can understand how the data is organized and perhaps identify unsuspected patterns.

Dimensionality reduction algorithms aim to simplify the data without losing too much information. Merging multiple features to create a new one is called ***feature extraction.***



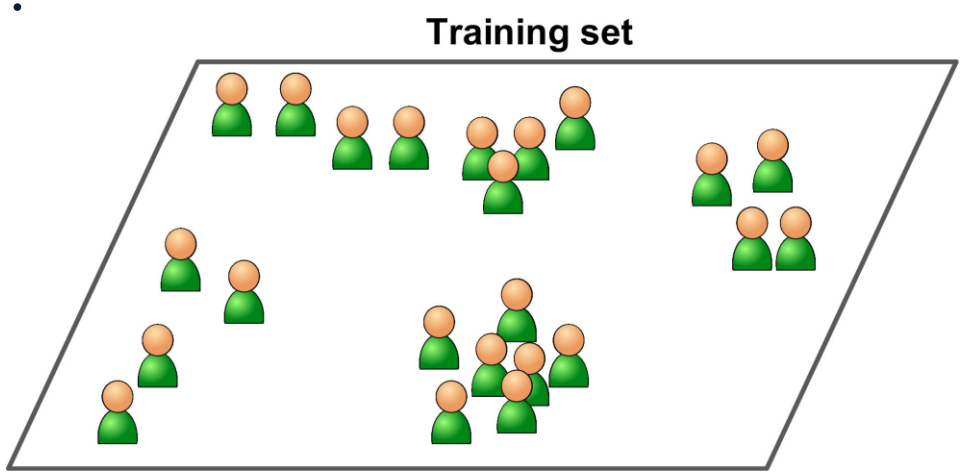
Example of a t-SNE visualization highlighting semantic clusters

Unsupervised Learning

I think it's your time to guess, what is **unsupervised learning?**
Its where the training data is **unlabeled,** and the system tries to learn without a teacher.

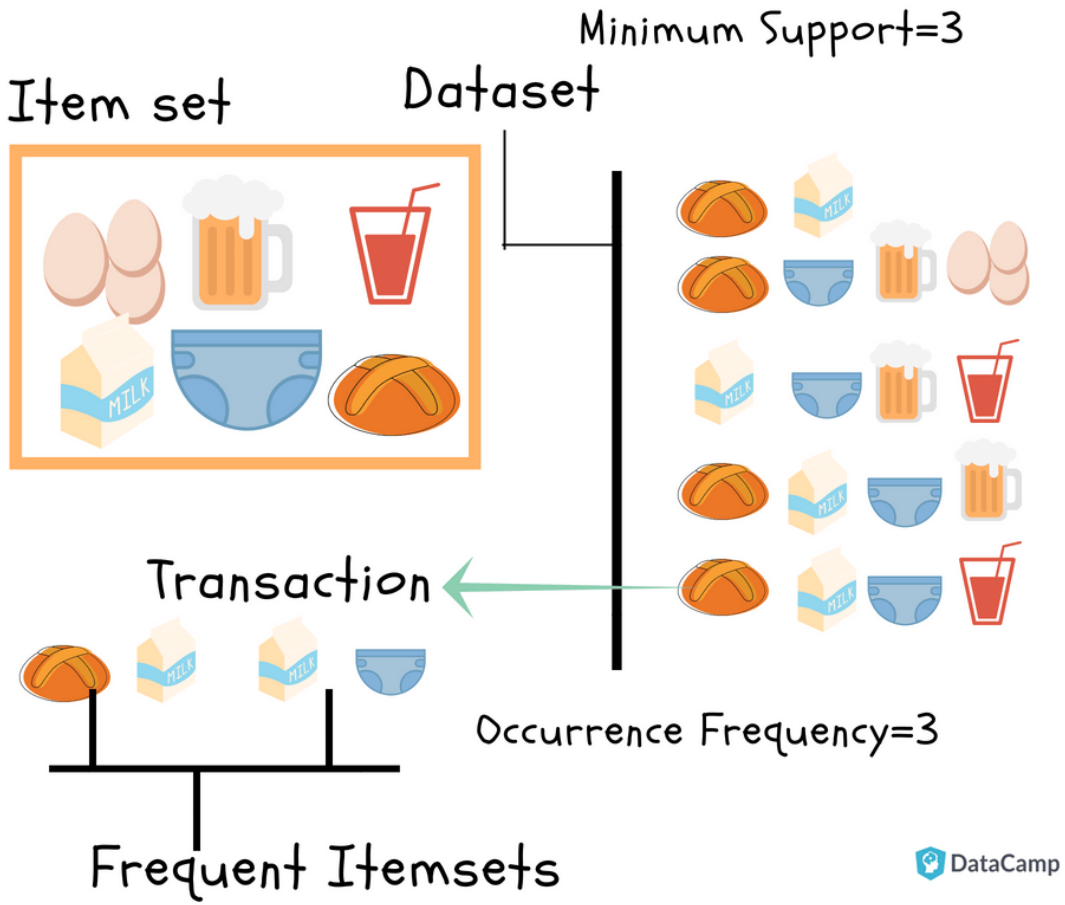
How can we benefit from unlabeled data you might say?
Through the following techniques:

- Clustering
- Anomaly detection and novelty detection
- Visualization and dimensionality reduction
- **Association rule learning**



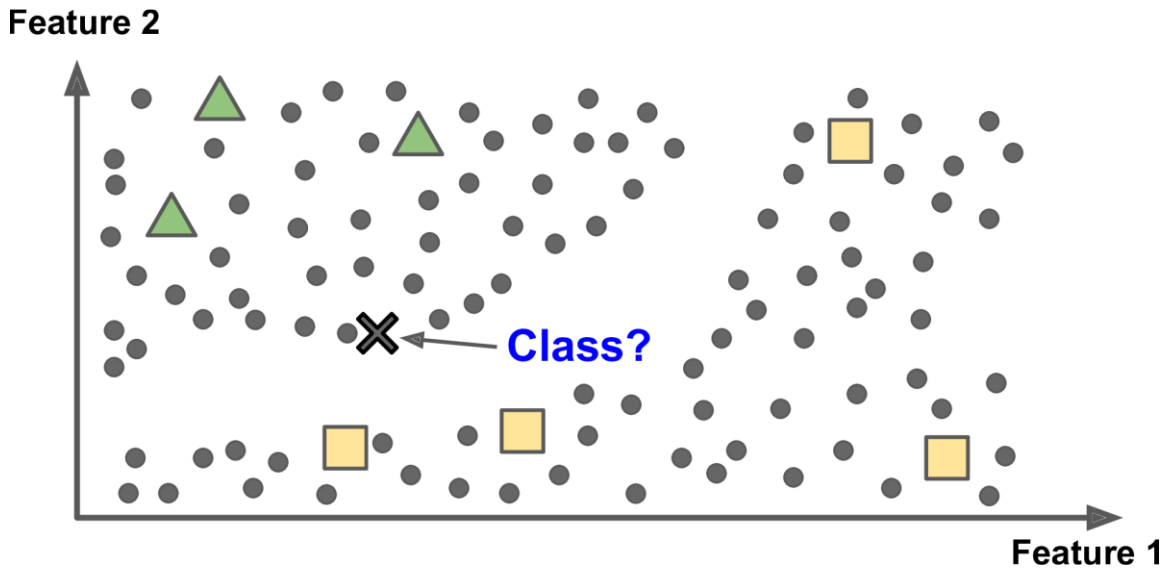
Association Rule Learning

Association rule learning algorithms' goal is to dig into large amounts of data and discover interesting relations between attributes.



Semisupervised Learning

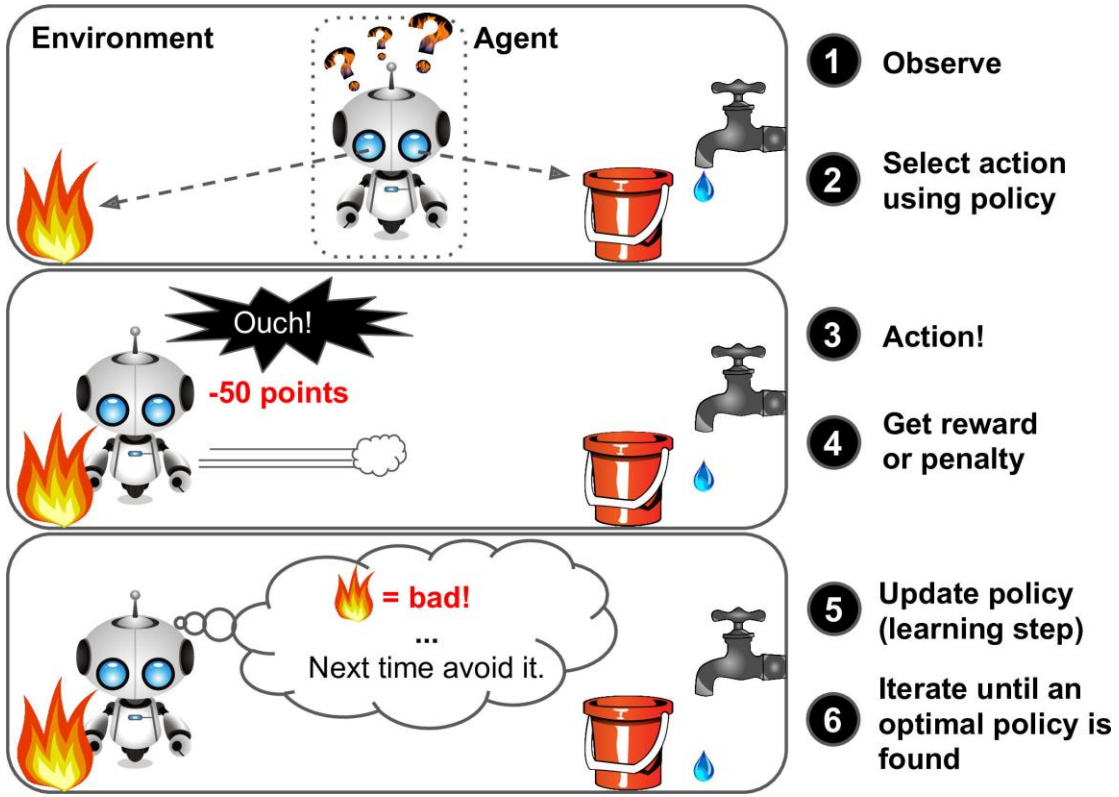
Some algorithms can deal with partially labeled training data, usually a lot of unlabeled data and a little bit of labeled data. This is called ***semisupervised learning***.



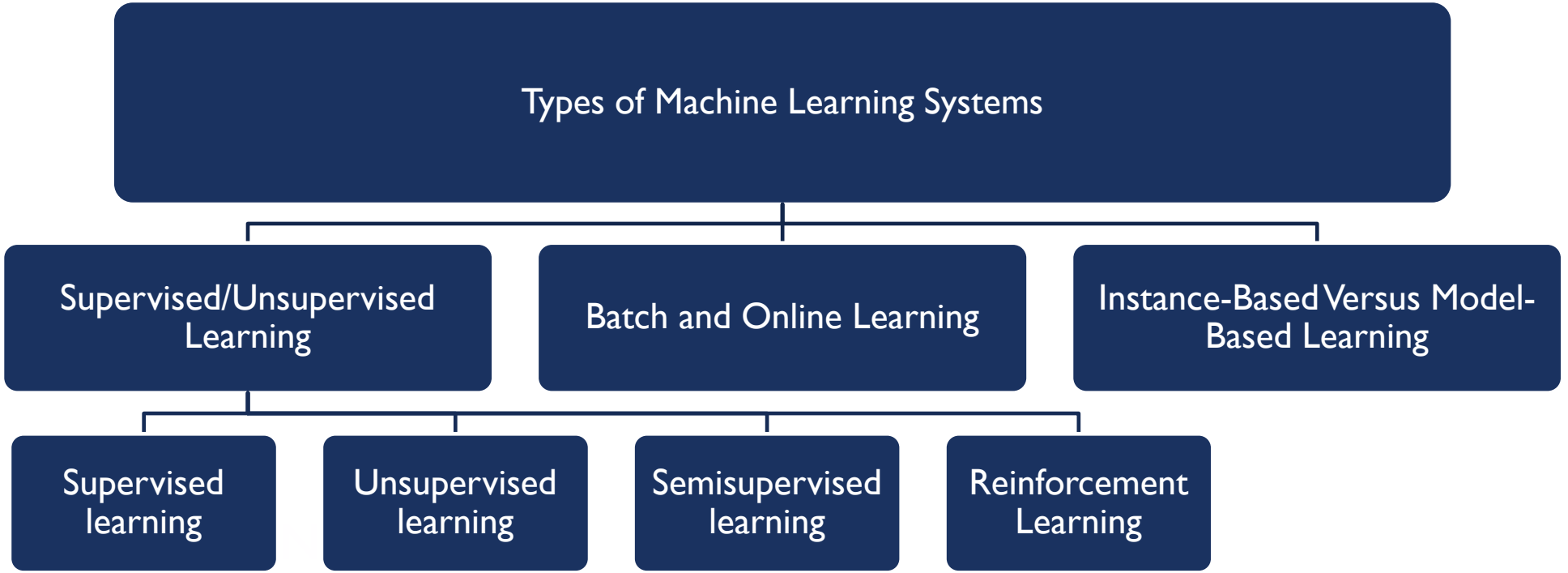
Reinforcement Learning

The learning system, called an **agent** in this context, can observe the **environment**, select and perform **actions**, and get **rewards** in return or **penalties** in the form of negative rewards,

It must then learn by itself what is the best strategy, called a **policy**, to get the most reward over time. A policy defines what action the agent should choose when it is in a given situation.



Recap


Types of Machine Learning Systems
Supervised/Unsupervised Learning
Batch and Online Learning
Instance-Based Versus Model-Based Learning
Supervised learning
Unsupervised learning
Semisupervised learning
Reinforcement Learning

Batch and Online Learning

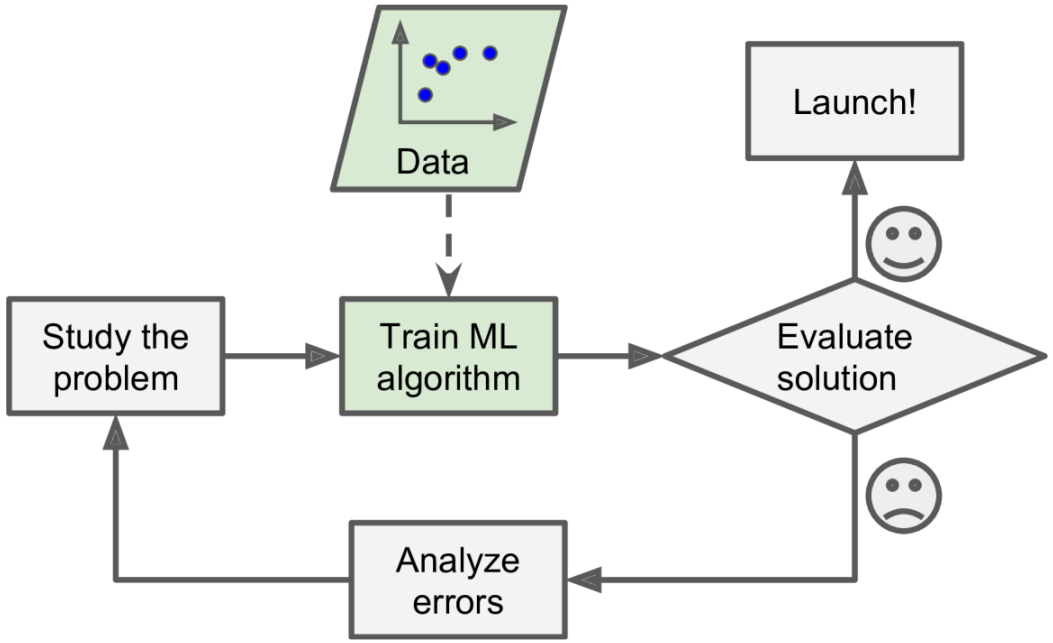
Another criterion used to classify Machine Learning systems is whether or not the system can learn incrementally from a stream of incoming data.

There are two main approaches: **batch** learning and **online** learning.

Batch Learning

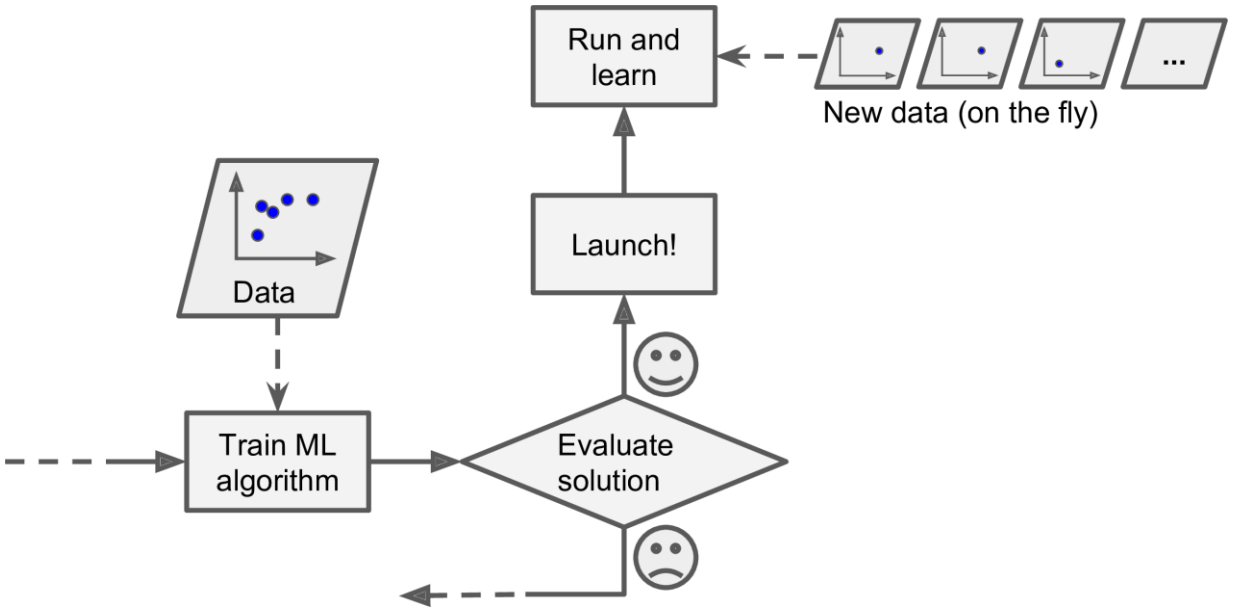
In batch learning, the system is incapable of learning incrementally: it must be trained using all the available data. This will generally take a lot of time and computing resources, so it is typically done offline.

First the system is trained, and then it is launched into production and runs without learning anymore; it just applies what it has learned. This is called ***offline learning***.



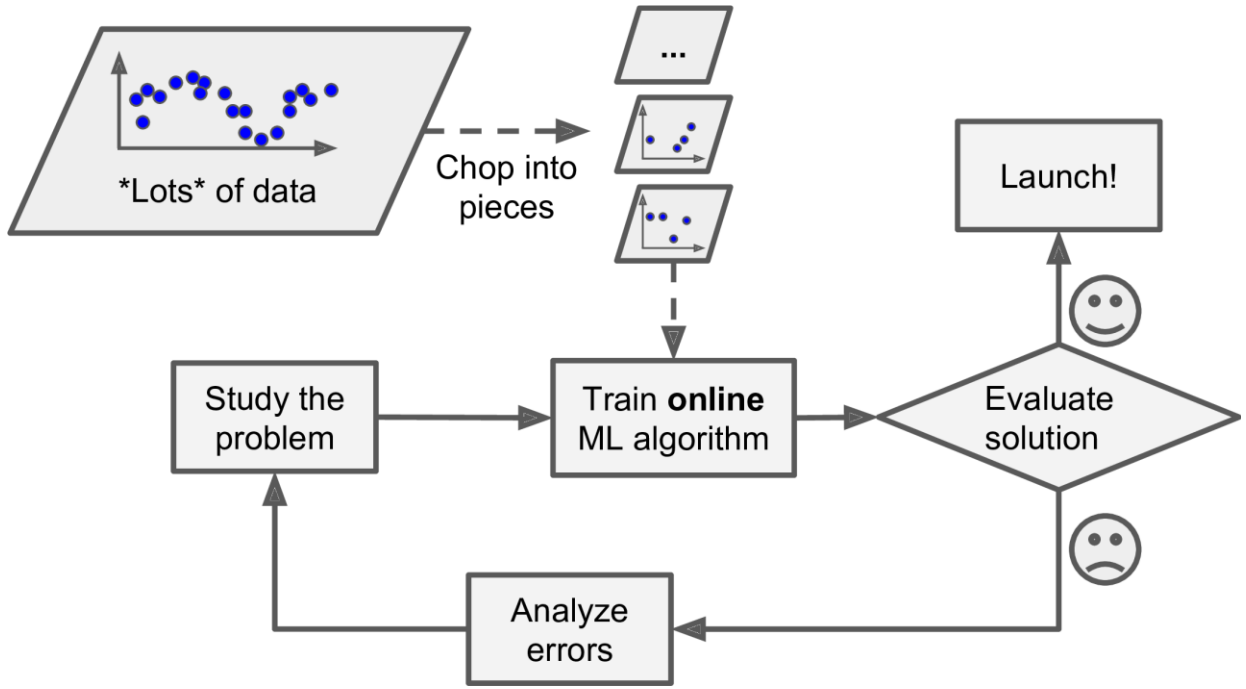
Online Learning

In online learning, you train the system incrementally by feeding it data instances sequentially, either individually or by small groups called mini-batches. Each learning step is fast and cheap, so the system can learn about new data on the fly, as it arrives.



Online Learning

Online learning algorithms can also be used to train systems on huge datasets that cannot fit in one machine's main memory (this is called out-of-core learning). The algorithm loads part of the data, runs a training step on that data, and repeats the process until it has run on all of the data.



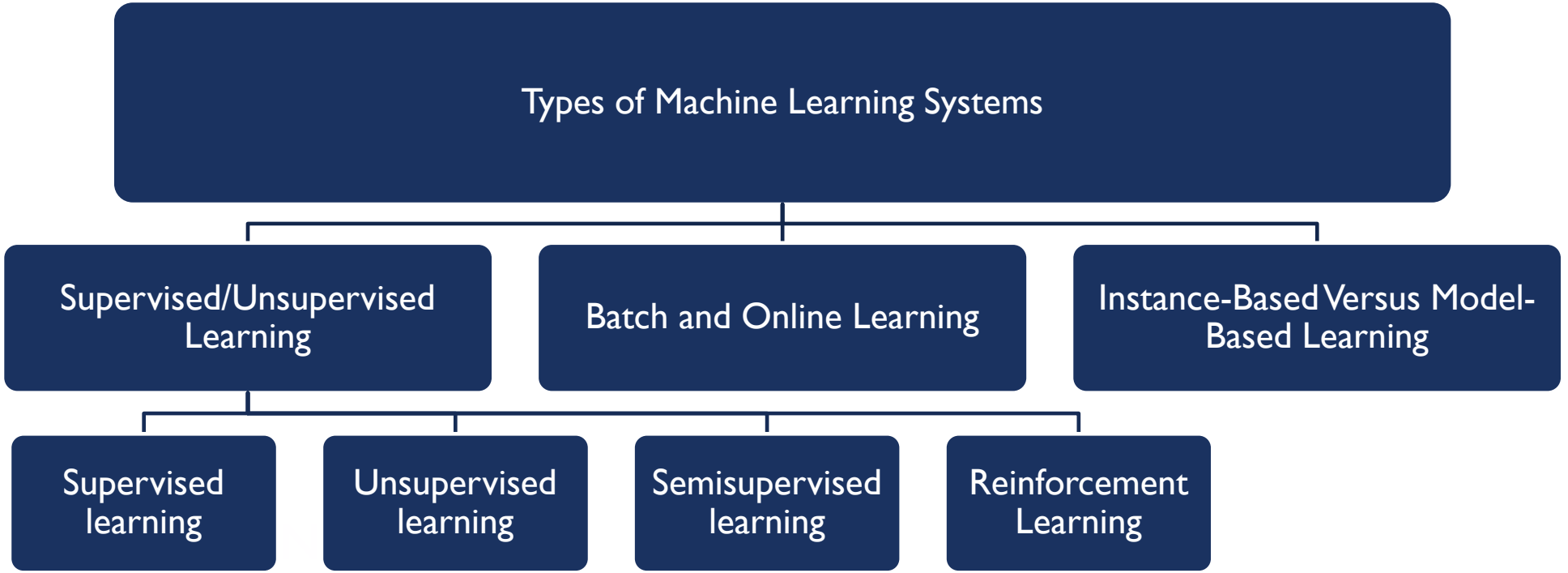
Online Learning

One important parameter of online learning systems is how fast they should adapt to changing data: this is called the learning rate.

If you set a high learning rate, then your system will rapidly adapt to new data, but it will also tend to quickly forget the old data. (you don't want a spam filter to flag only the latest kinds of spam it was shown)

Conversely, if you set a low learning rate, the system will have more inertia; that is, it will learn more slowly, but it will also be less sensitive to noise in the new data or to sequences of nonrepresentative data points (outliers).

Recap


Types of Machine Learning Systems
Supervised/Unsupervised Learning
Batch and Online Learning
Instance-Based Versus Model-Based Learning
Supervised learning
Unsupervised learning
Semisupervised learning
Reinforcement Learning

THANK YOU

NEXT LECTURE WILL BE ONLINE
ON SAT, 13.5.2023, IN SHAA ALLAH!

SABBAGH@IEEE.ORG

CONNECT ON LINKEDIN