Machine Learning Basics

Your AI Starter Kit

Lara Wehbe - TheAIEngineer - July 2024

Course Overview

This course if for:

- University Students Eager to Learn AI
- Software Engineers that want to start a career in AI
- Non-tech people looking to know more about AI
- Anyone interested in nourishing their skills
- If you dont know where to start, this course is for you
- If you want to find a job in Machine Learning this course is for you

Why do you need to learn AI?

A must-have skill in 2025 that helps you secure a job!

- #1 Job needed in 2024 and expecting high demand in the coming 5 years
- Highest paying job. Example: Netlflix wants to pay $900k per year for an AI Engineer!
- It’s the future!
- A job that won’t be replaced by AI!

What to expect?

Embark on a journey through "Machine Learning Basics" and discover how supervised learning can empower you to turn data into actionable insights, making you a valuable problem solver in the world of AI.

PS: This is written by ChatGPT!

Outcome

Boost your resume with at least 3 projects and secure a job

Outline

Course Outline

1. Introduction and Definitions
    1. AI History
    2. AI Applications
2. Machine Learning
3. Supervised Learning
4. Unsupervised Learning
5. Supervised vs Unsupervised
6. Cost function
7. Gradient Descent
8. BONUS Section

Introduction

What is Artificial Intelligence (AI)?

AI Definition

Artificial Intelligence (AI), a term coined by

emeritus Stanford Professor John McCarthy in 1955,

was defined by him as “the science and engineering of

making intelligent machines”

It gives the machines the ability to Mimic human behavior.

What is Machine Learning (ML)?

ML is a subset of AI that focuses on the development of algorithms that allow computers to learn from and make predictions based on data.

It specifically deals with data-driven learning and improving predictions.

Artificial Intelligence or Machine Learning?

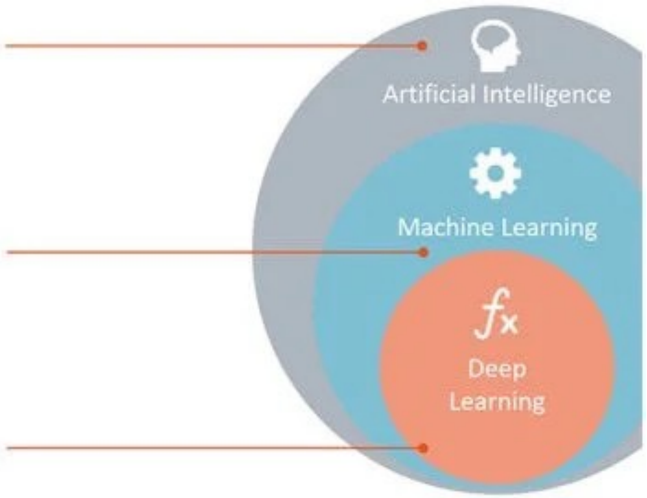
What's the difference?

The terms Artificial Intelligence (AI) and machine learning (ML) are often used interchangeably, but they are not the same.

Artificial Intelligence
Any technique which enables computers to mimic human behavior.

Machine Learning
Subset of AI techniques which use statistical methods to enable machines to improve with experiences.

Deep Learning
Subset of ML which make the computation of multi-layer neural networks feasible.



AI vs ML

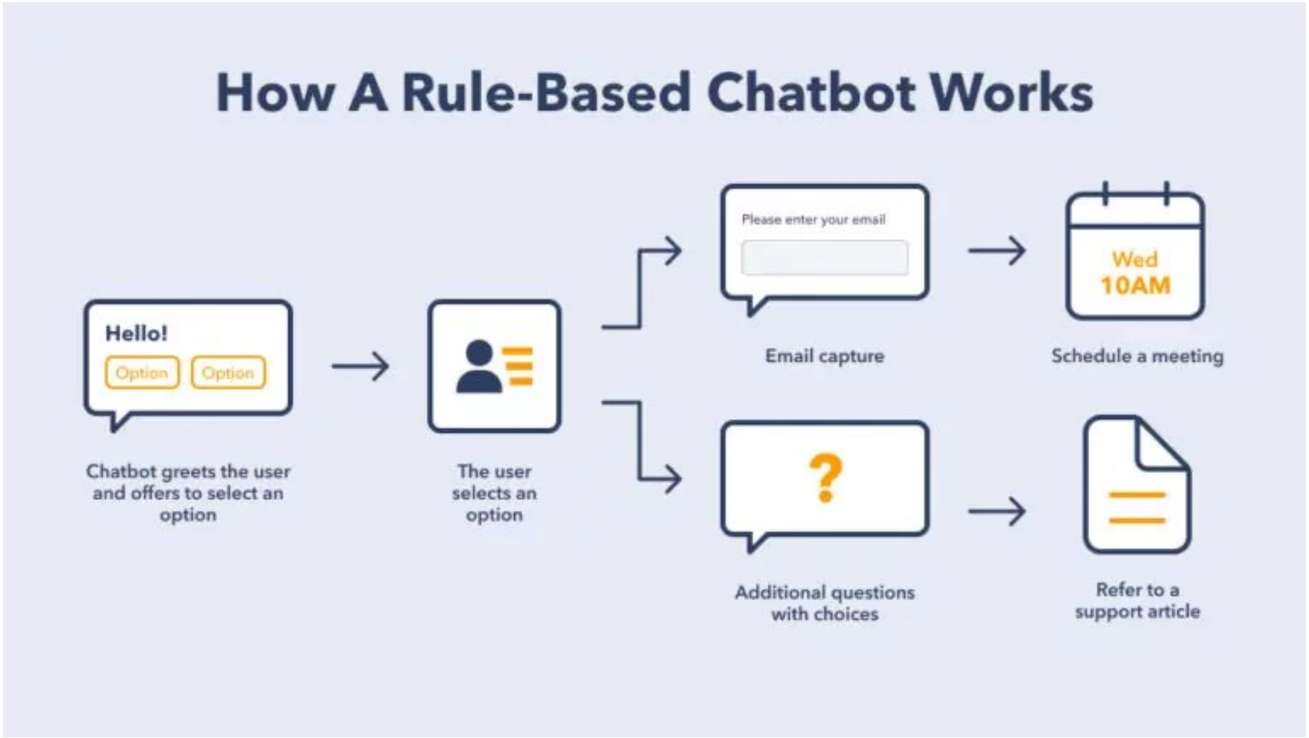
Key Similarities and Differences

Aspect	Artificial Intelligence (AI)	Machine Learning (ML)
Data-driven	Relies on data to make informed decisions	Uses data to learn and improve
Automation	Aims to automate tasks requiring human intervention	Automates data analysis
Improvement over time	Can become more effective with more data	Improves performance with more data exposure
Computational complexity	Requires significant computational power	Requires significant computational power
Interdisciplinary fields	Draws from computer science, statistics, mathematics, engineering	Draws from computer science, statistics, mathematics, engineering
Scope	Broader scope: mimics human intelligence in various domains	Focused scope: development of learning algorithms
Goal	Create systems performing tasks requiring human intelligence	Enable machines to learn from data for accurate predictions/decisions
Learning	Does not necessarily learn from data (e.g., rule-based systems)	Specifically involves learning from data
Dependency	Encompasses machine learning	Subset of AI
Types of learning	Can use rule-based, heuristic, or machine learning techniques	Supervised, unsupervised, or reinforcement learning
Human intervention	Varies: manual tuning or autonomous systems	Aims to minimize human intervention, automating learning

Example of AI vs ML

Can you think of an example where AI is used without ML?

Rule-Based Systems: simple chatbot that responds with pre-defined answers based on keywords you use



AI Umbrella

AI << ML << DL << GenAI

Artificial Intelligence

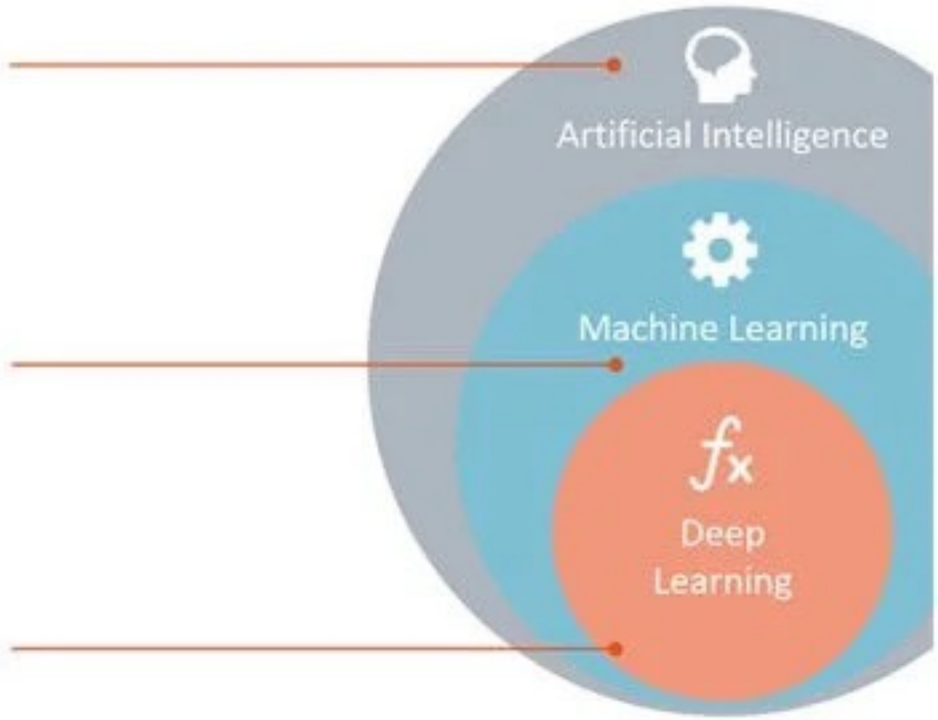
Any technique which enables computers to mimic human behavior.

Machine Learning

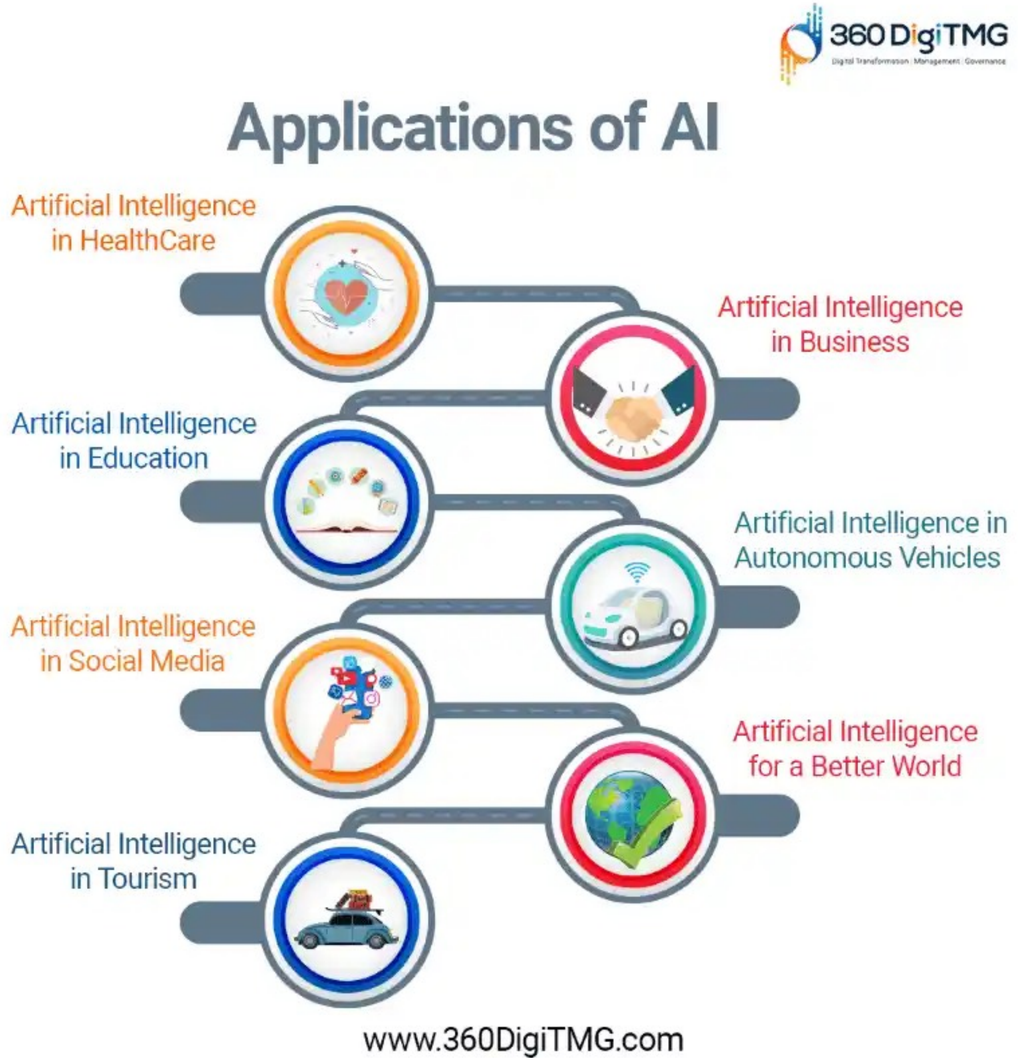
Subset of AI techniques which use statistical methods to enable machines to improve with experiences.

Deep Learning

Subset of ML which make the computation of multi-layer neural networks feasible.

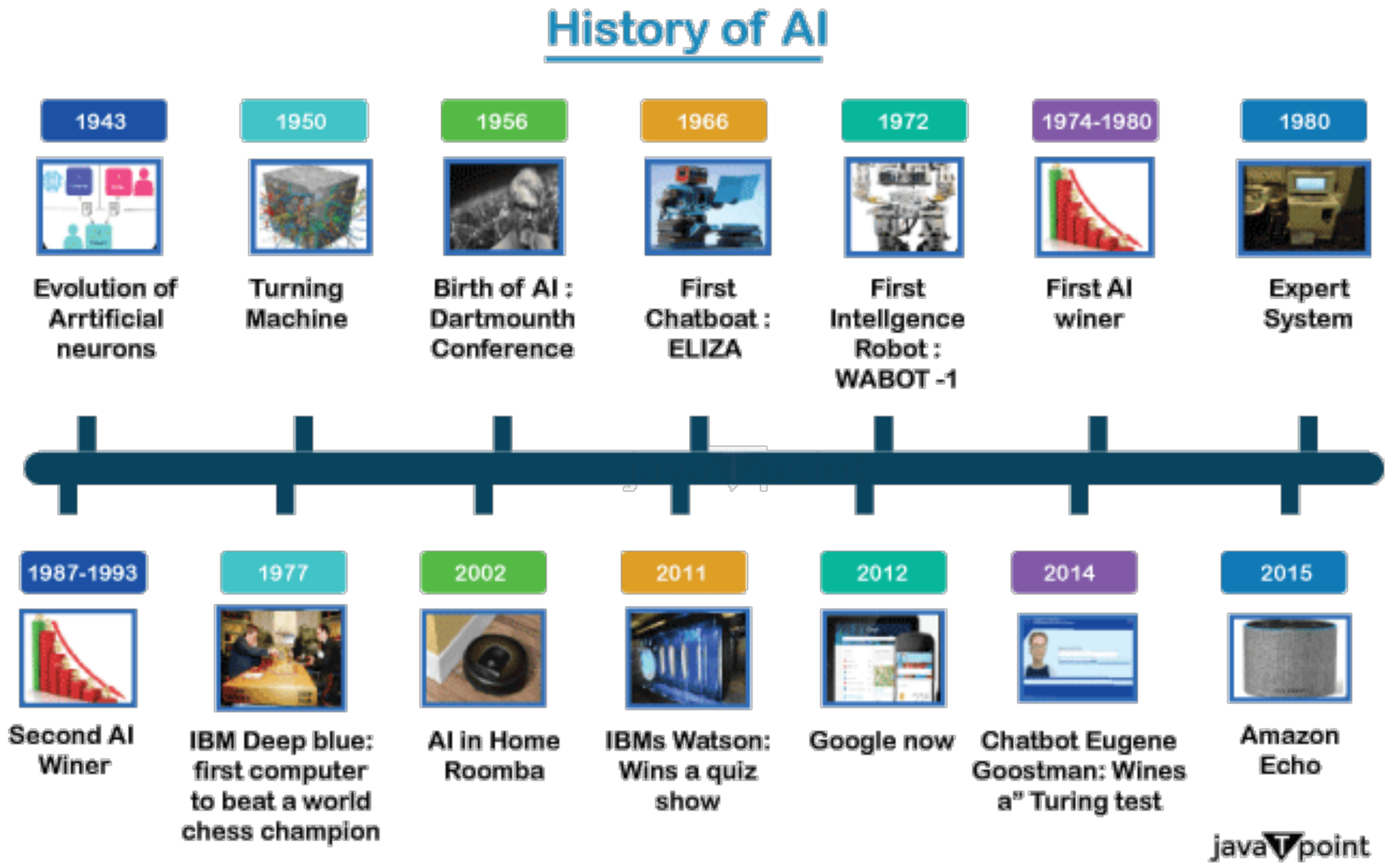


AI Applications



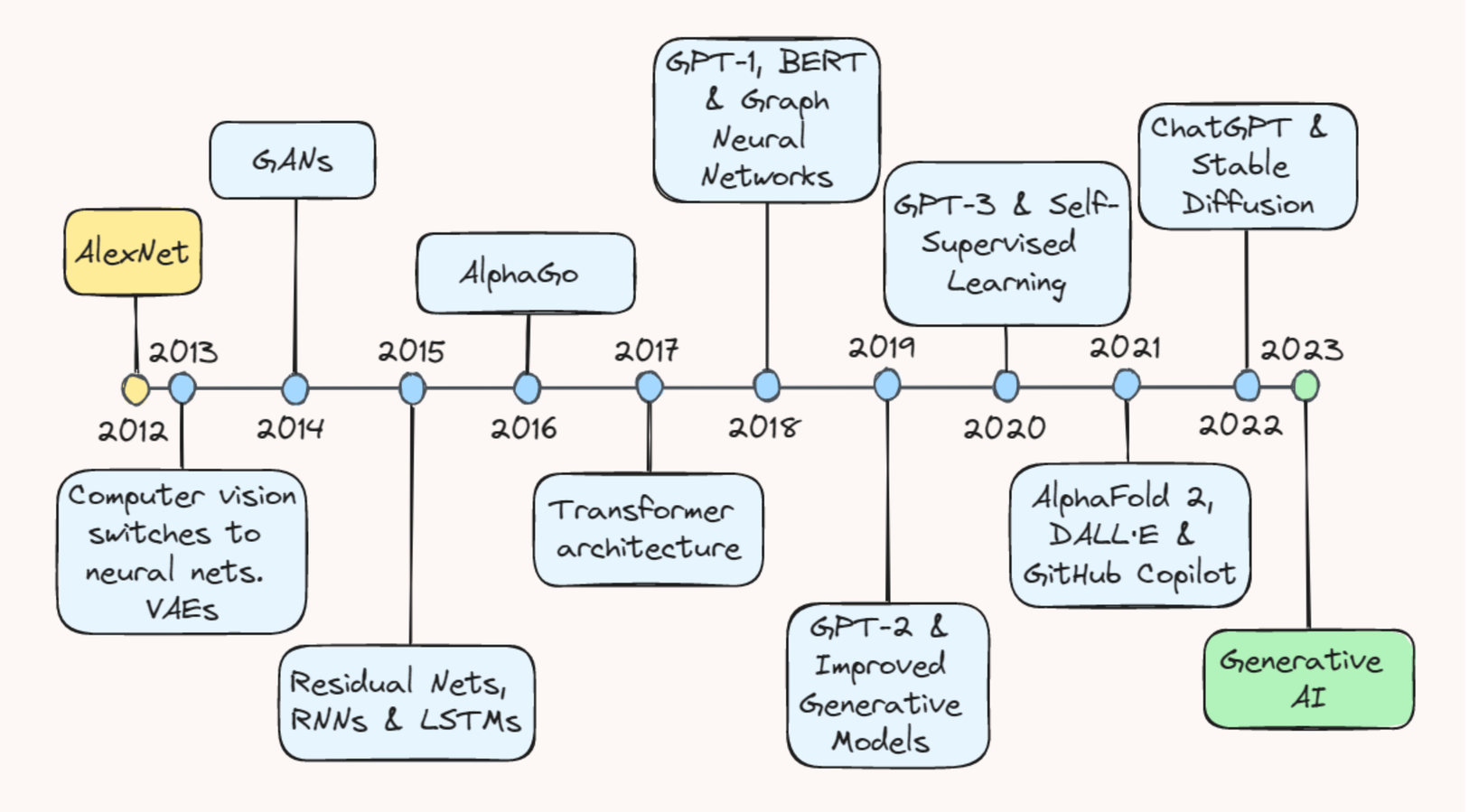
AI History

Did it really start smart?



AI History

Did it really start smart?



Machine Learning

Definition and Introduction

Machine Learning (ML) is a subset of artificial intelligence (AI) that focuses on developing algorithms that allow computers to learn from and make decisions based on data.

Definition and Introduction

Instead of being explicitly programmed, ML algorithms improve and adapt through experience, using data to find patterns and make predictions.

Famous Definitions

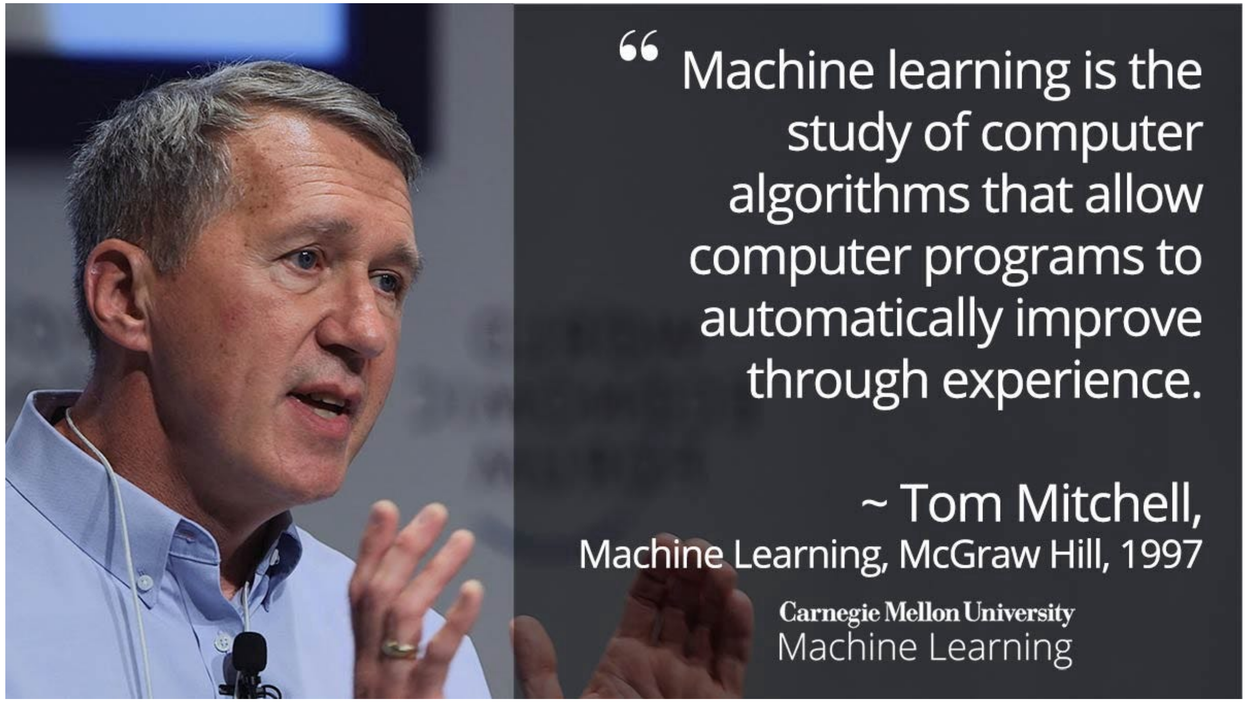
Arthur Samuel (1959)

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

Famous Definitions

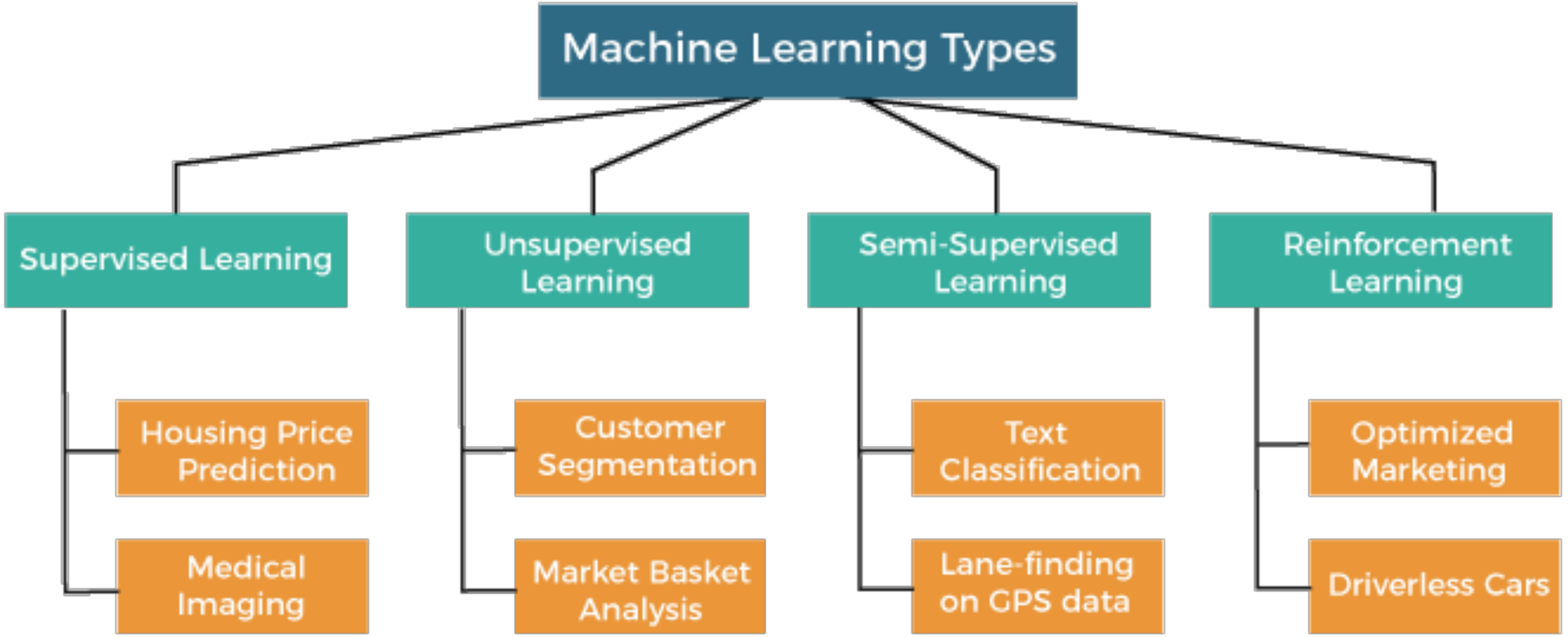
Tom Mitchell, Founders University Professor at Carnegie Mellon University

“A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.”



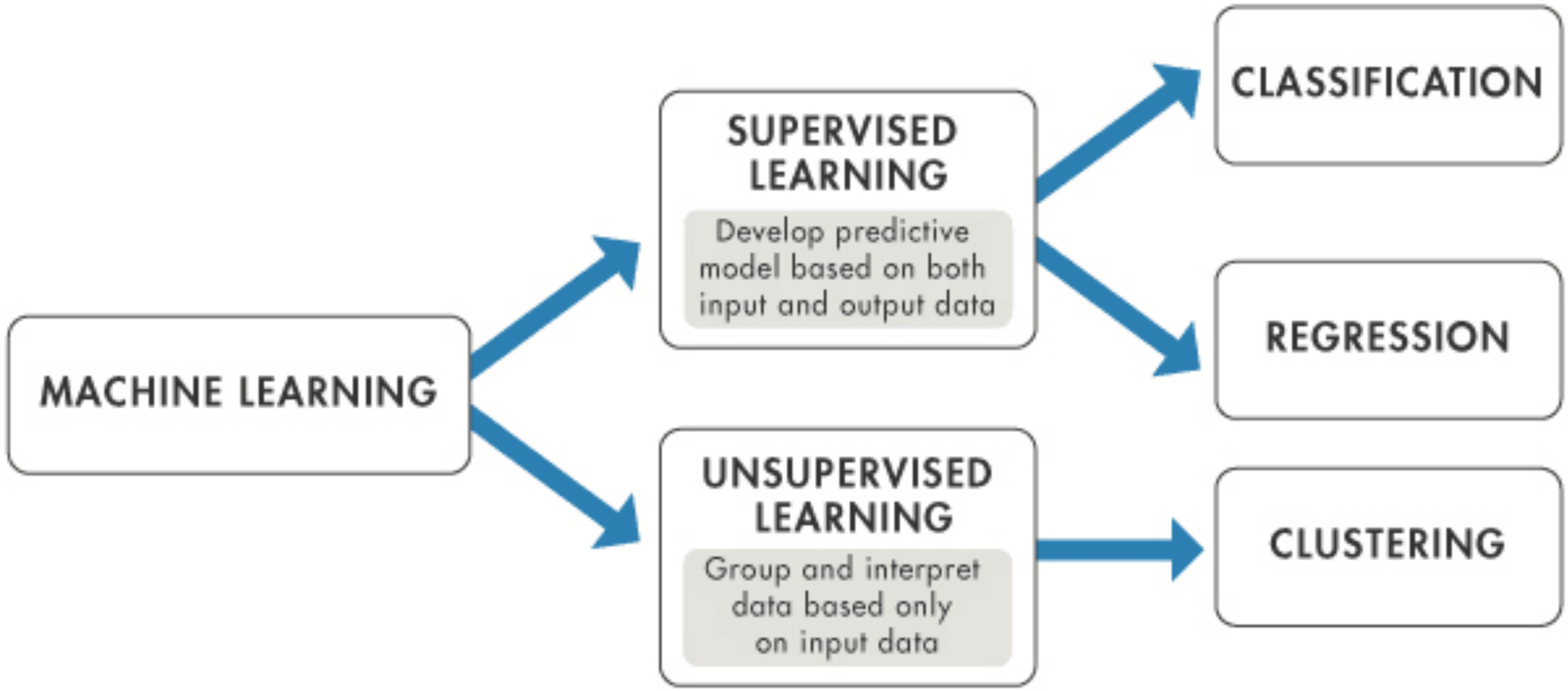
Machine Learning: Types of Learning

Supervised, unsupervised, semisupervised, self supervised



Machine Learning: Types of Learning

Supervised, unsupervised, semisupervised, self supervised



Terminologies

Terminologies

We will use a lot in this course

Model: Blackbox that runs the algorithms

Algorithm: Set of rules we give to the machine to abide

Data: The entry points that we need the model to learn from

Dataset: A group of data that we need to use in our machine learning model

Training Set: The set used for training the model

Testing Set: The set used for testing the model

X, y: X is the input, and y is the truth value (ground truth)

Cost Function:function that measures the error or difference between the predicted output and the actual output

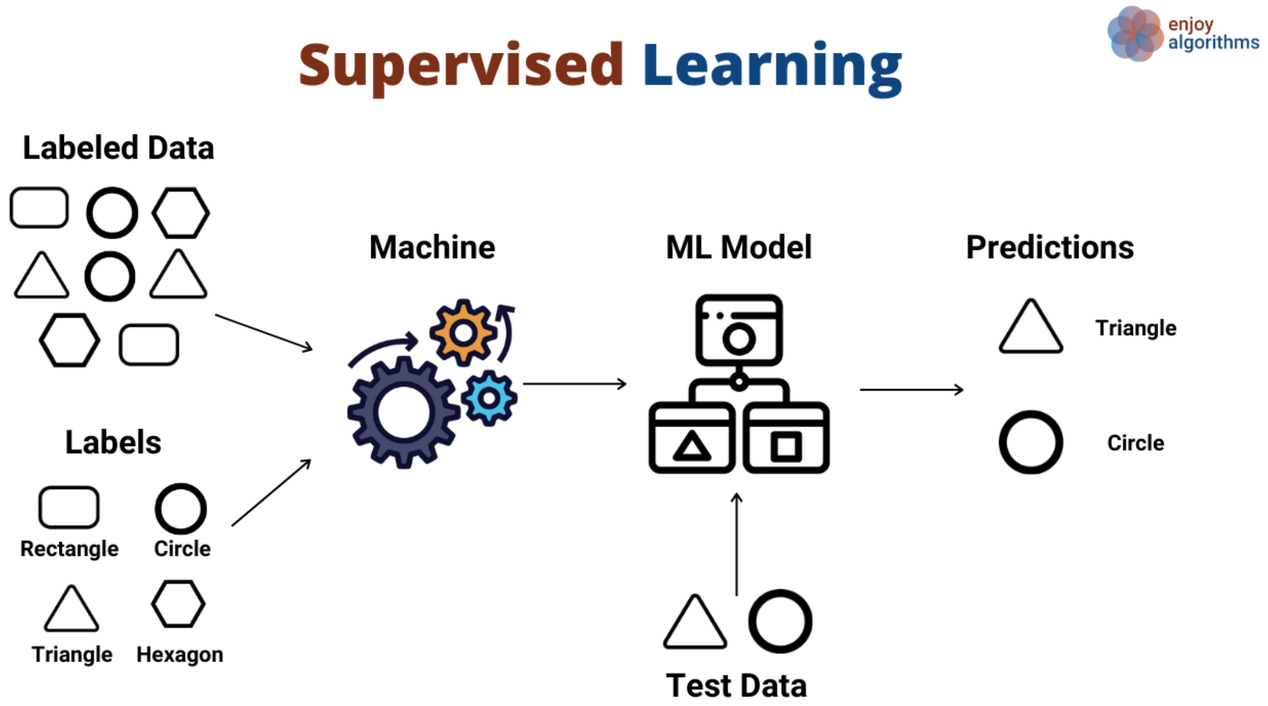
Supervised Learning

Definition

What is Supervised Learning?

Supervised learning is a category of *Machine Learning* that uses labeled datasets to train algorithms to predict outcomes and recognize patterns.

How it works?


Supervised Learning
enjoy algorithms
Labeled Data
Machine
ML Model
Predictions
Triangle
Circle
Labels
Rectangle
Circle
Triangle
Hexagon
Test Data

Supervised Learning: Algorithms

Two types of problems:

Regression and Classification

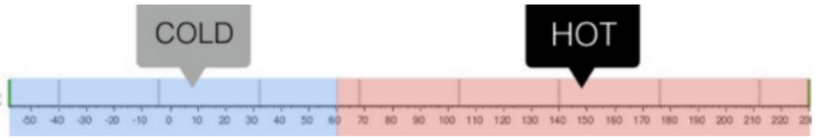
Regression

What will be the temperature tomorrow?

Fahrenheit

Classification

Will it be hot or cold tomorrow?



Fahrenheit

Supervised Learning: Algorithms

Classification

Classification: Assigns test data to specific categories.

Input: Discrete

Output: Discrete

Supervised Learning: Algorithms

Regression

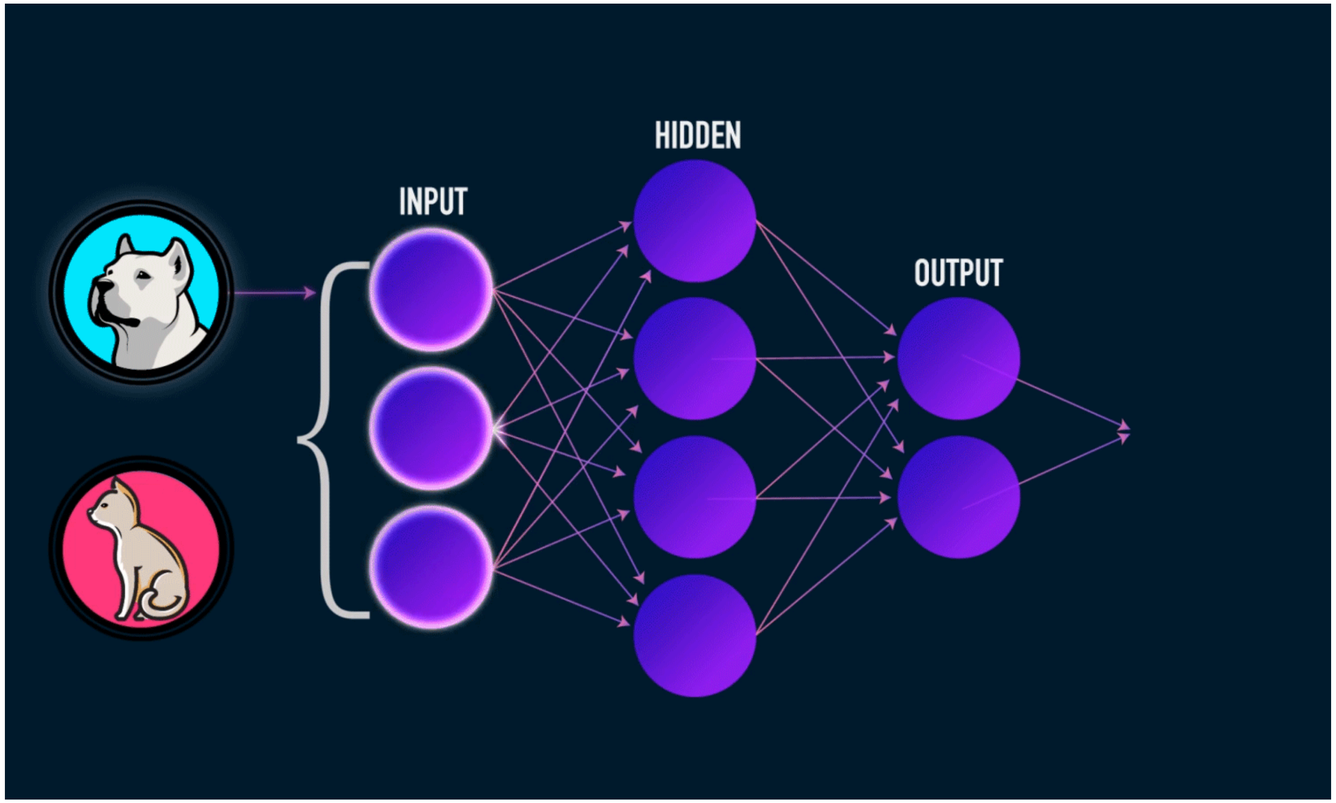
Regression: Understand relationship between dependent and independent variables.

Input: Continuous

Output: Continuous

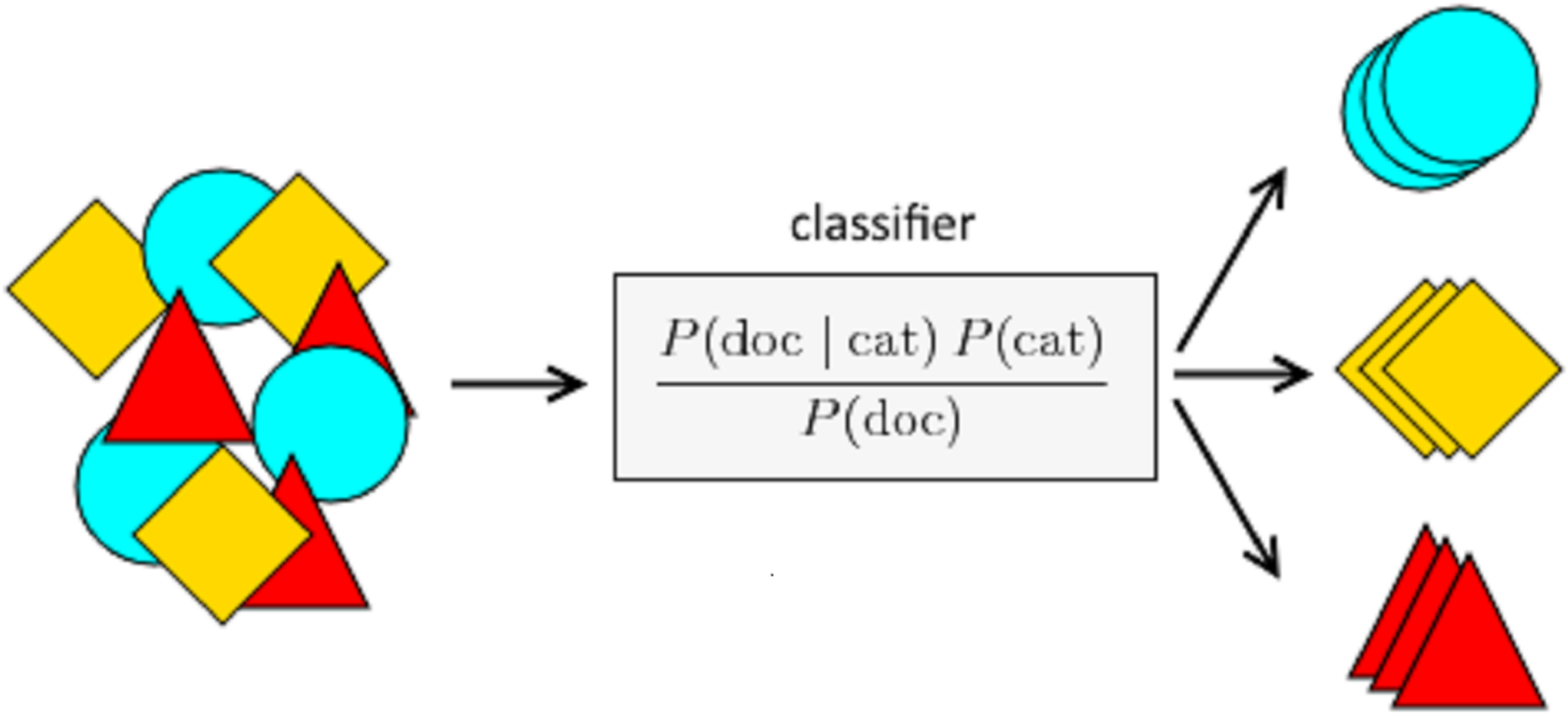
Supervised Learning: Algorithms

Famous Algorithms: Neural Networks



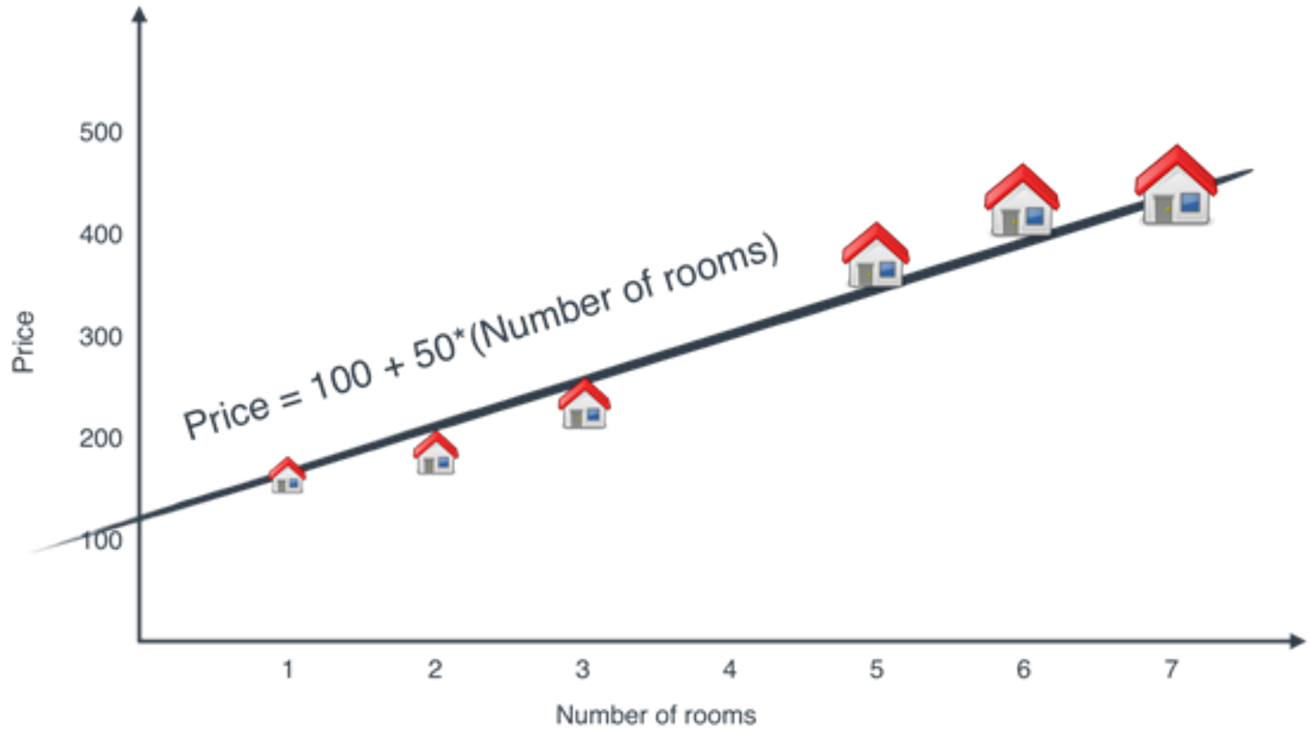
Supervised Learning: Algorithms

Famous Algorithms: Naive Bayes



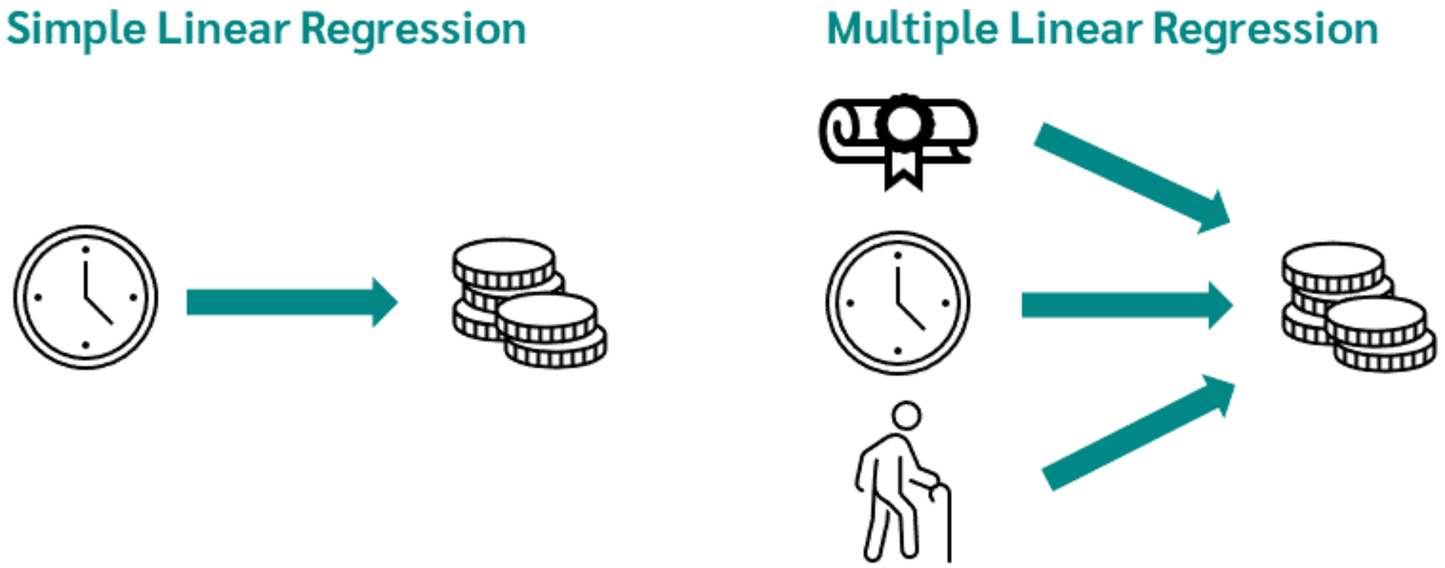
Supervised Learning: Algorithms

Famous Algorithms: Linear Regression



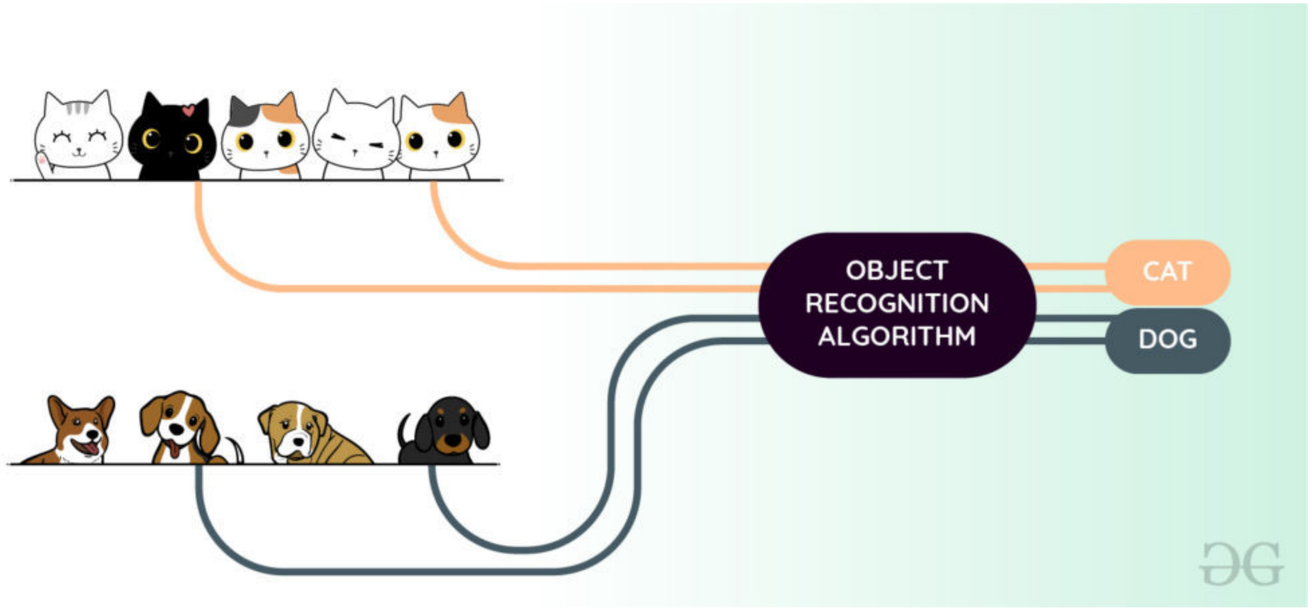
Supervised Learning: Algorithms

Famous Algorithms: Multiple Linear Regression



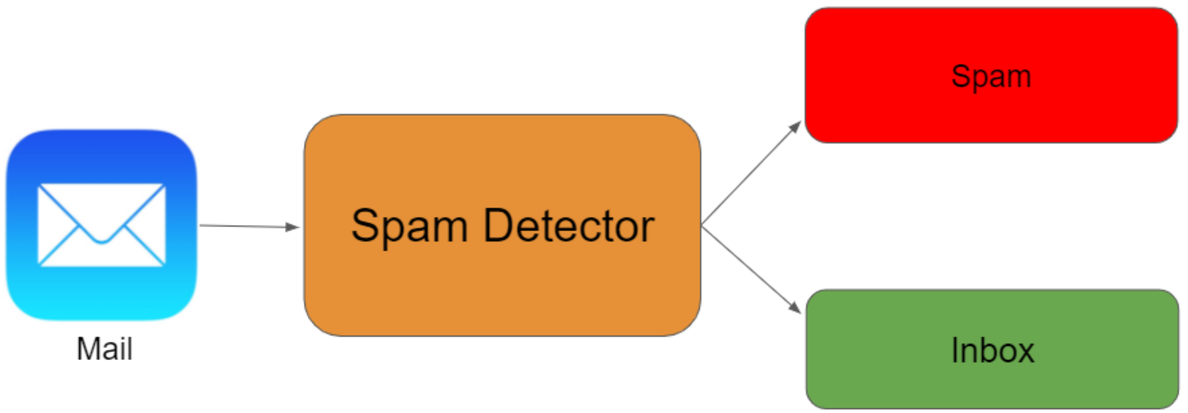
Supervised Learning - Business Applications

Object Recognition



Supervised Learning - Business Applications

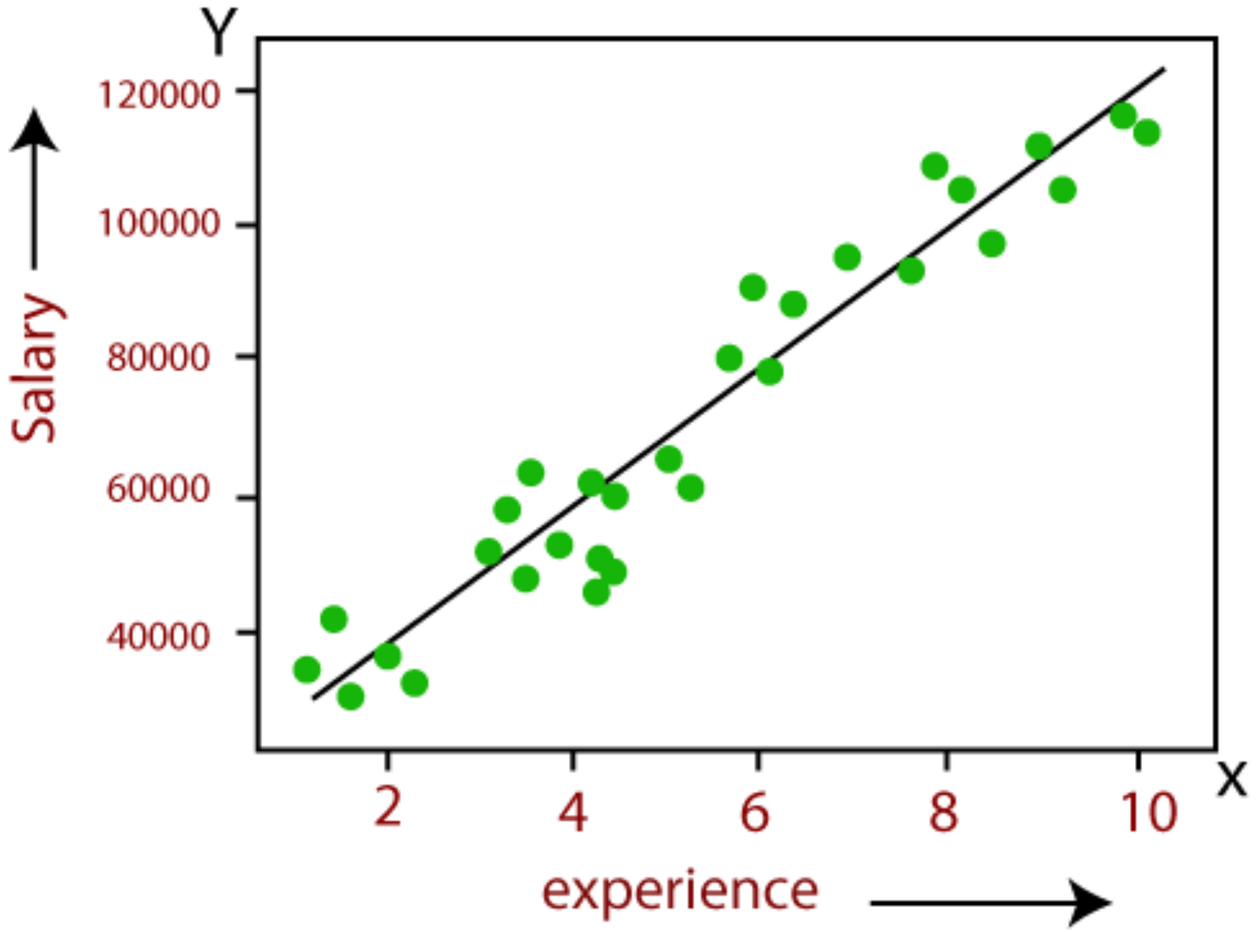
Spam Detection



Supervised Learning - Regression

Regression

Regresses to the MEAN

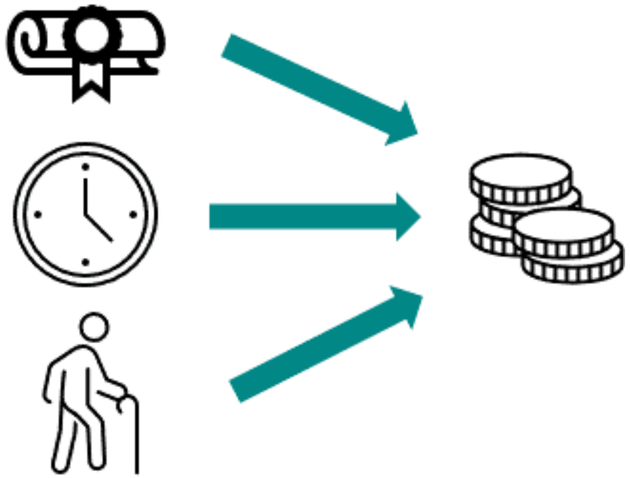


Regression

Two types

Simple Linear Regression

Multiple Linear Regression



Linear Regression

Math behind it

Linear Regression

Math behind it

Equation: $y = ax + b$

Y: Dependent Variable (response)

X: Independent variable (predictor)

B: intercept (constant term)

A: slope (coefficient)

Goal: Find the best-fitting line that minimizes the difference between actual and predicted values of *y*

Linear Regression

How do we find the best-fit line?

We need to minimize the ***Cost Function***

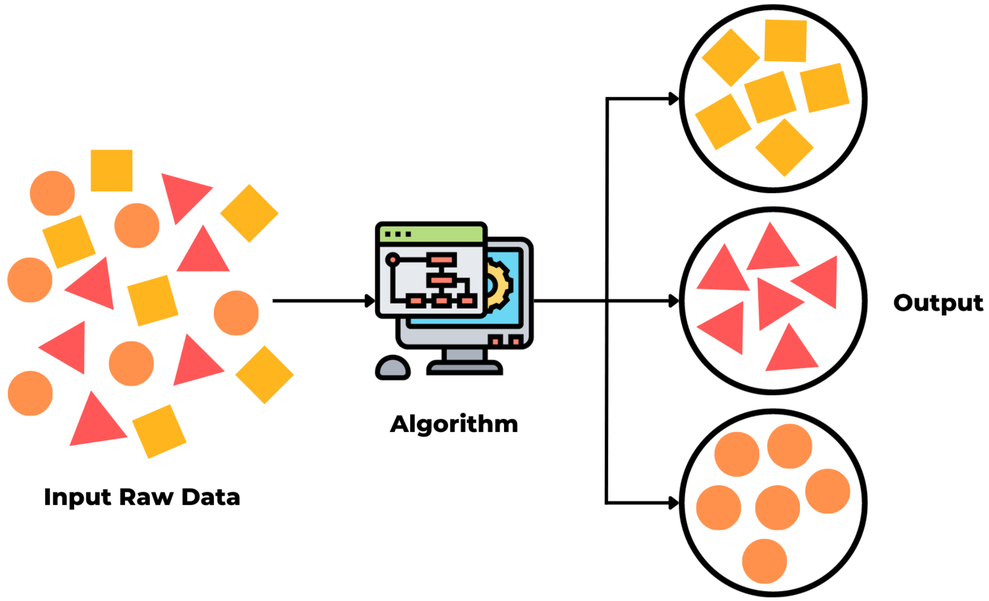
We will discuss this point later in this course

Unsupervised Learning

Definition

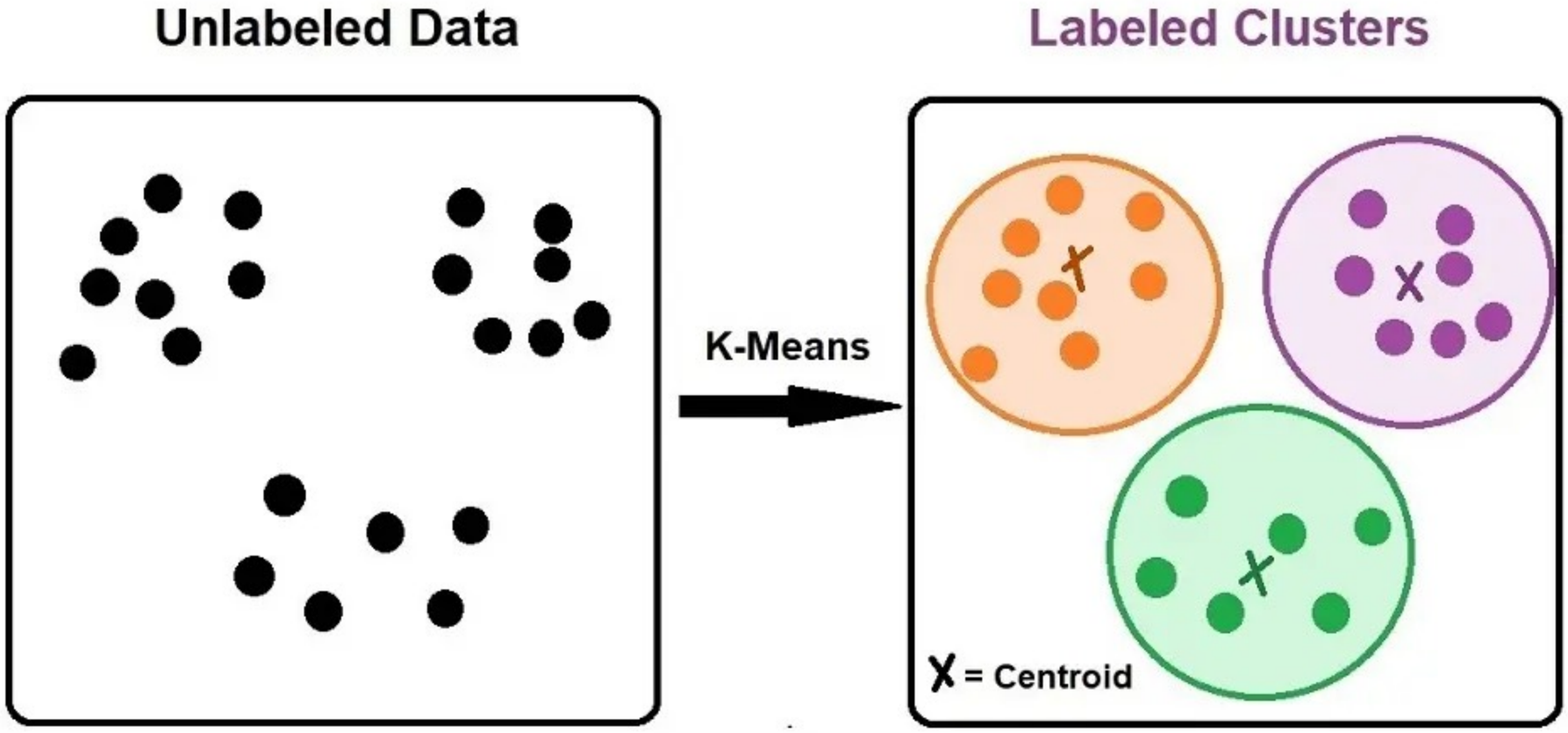
What is Unsupervised Learning?

Unsupervised learning uses machine learning algorithms to analyze and cluster unlabeled data sets. These algorithms discover hidden patterns in data without the need for human intervention (hence, they are “unsupervised”).



How it works?

Example on KNN Algorithm



Unsupervised Learning: Main Algorithms

Common unsupervised learning approaches

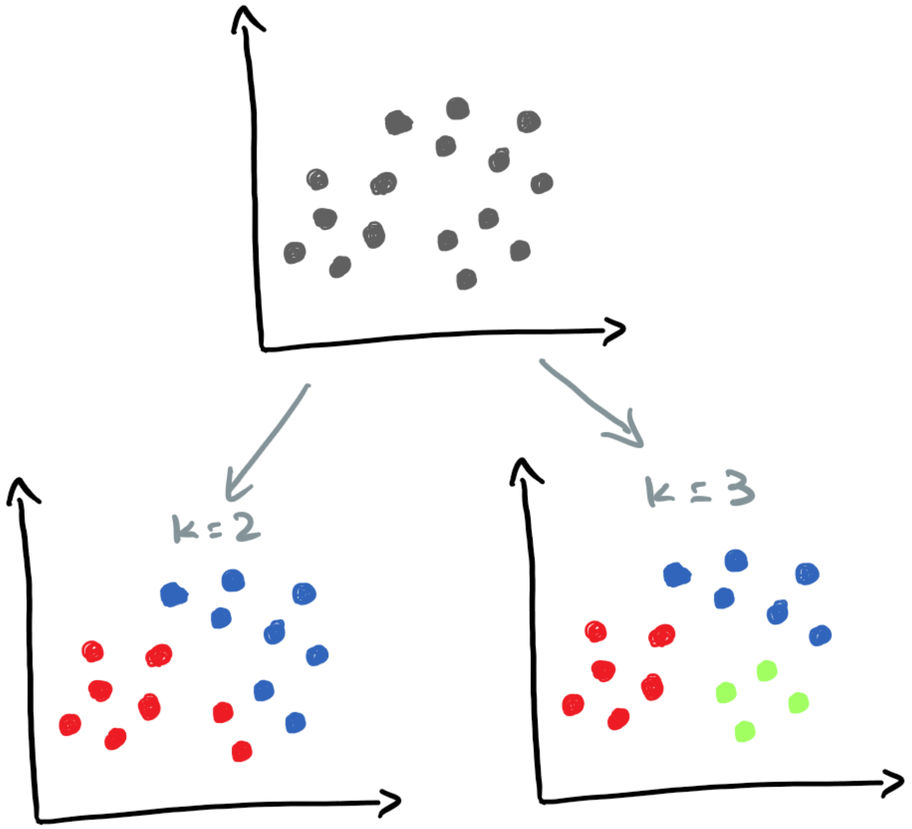
Clustering

Association

Dimensionality Reduction

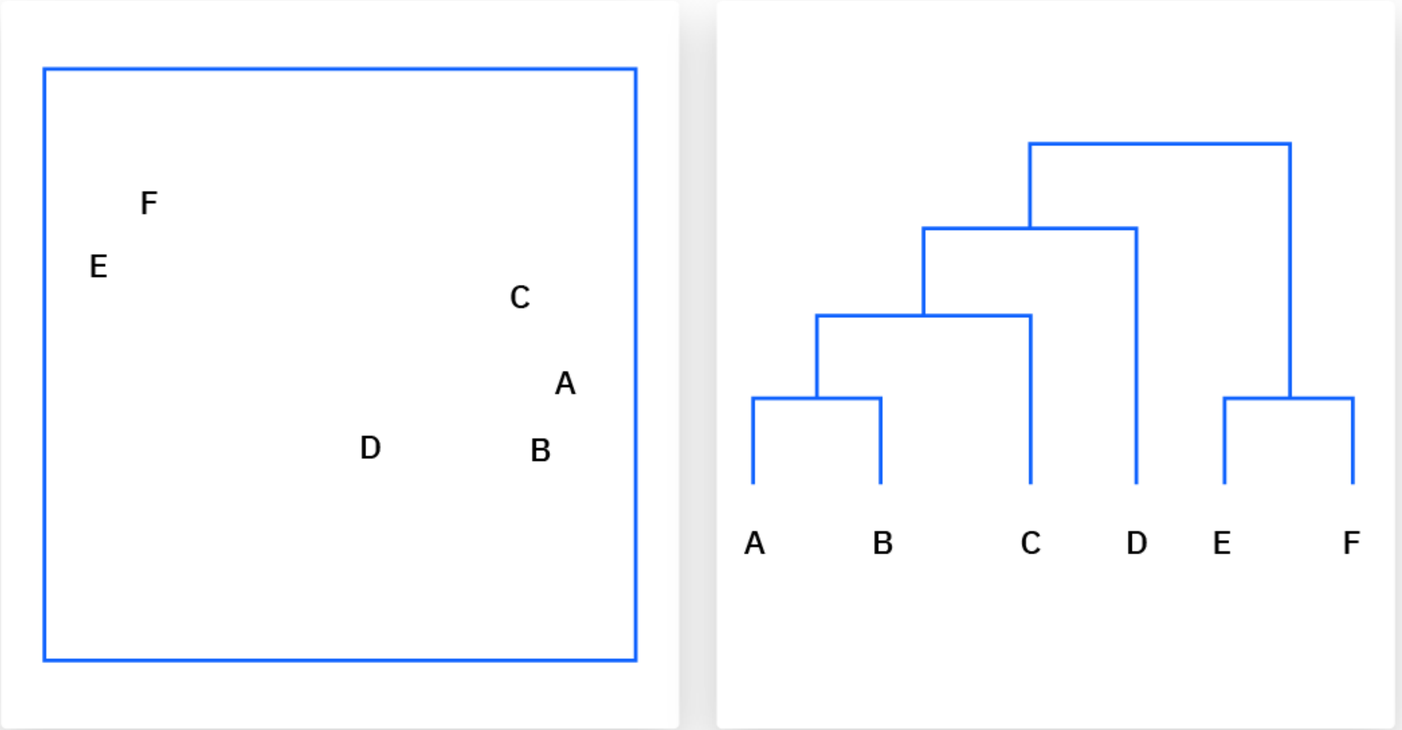
Unsupervised Learning: Clustering

K-Nearest Neighbor



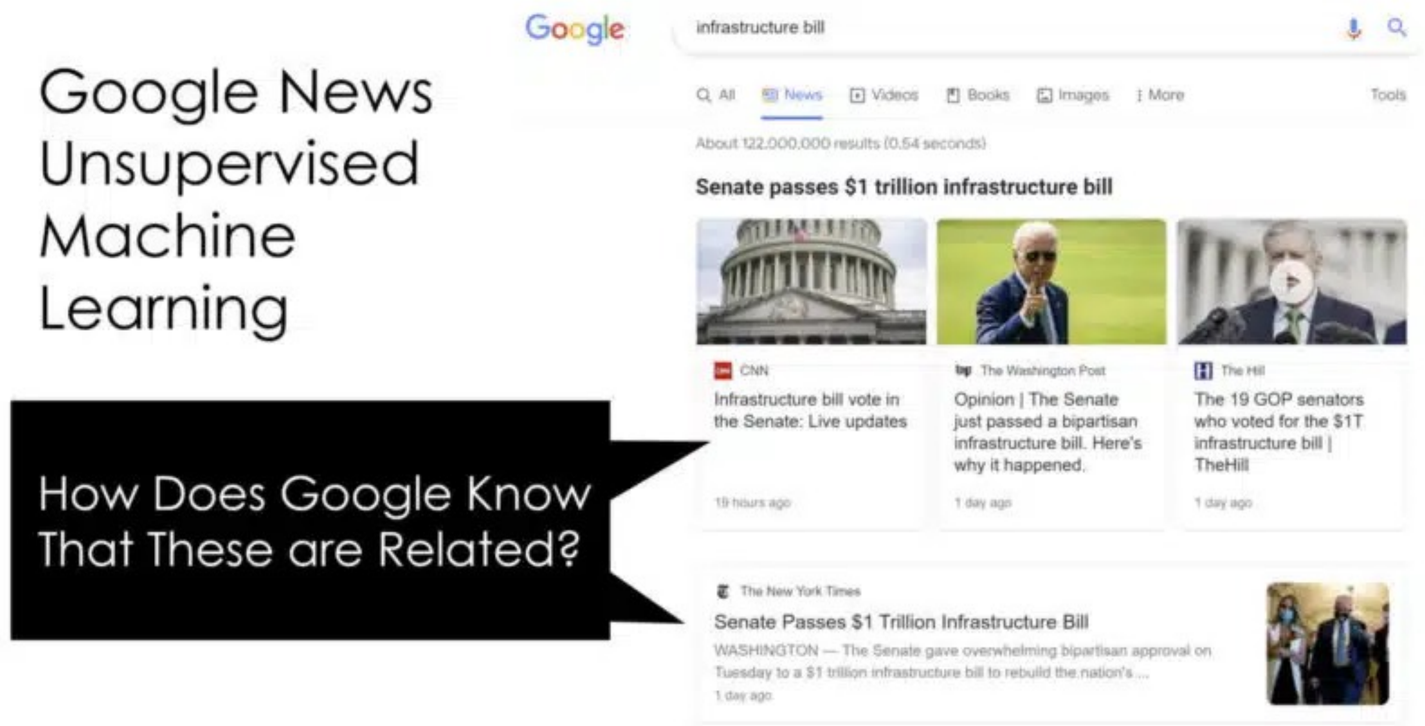
Unsupervised Learning: Clustering

Hierarchical clustering



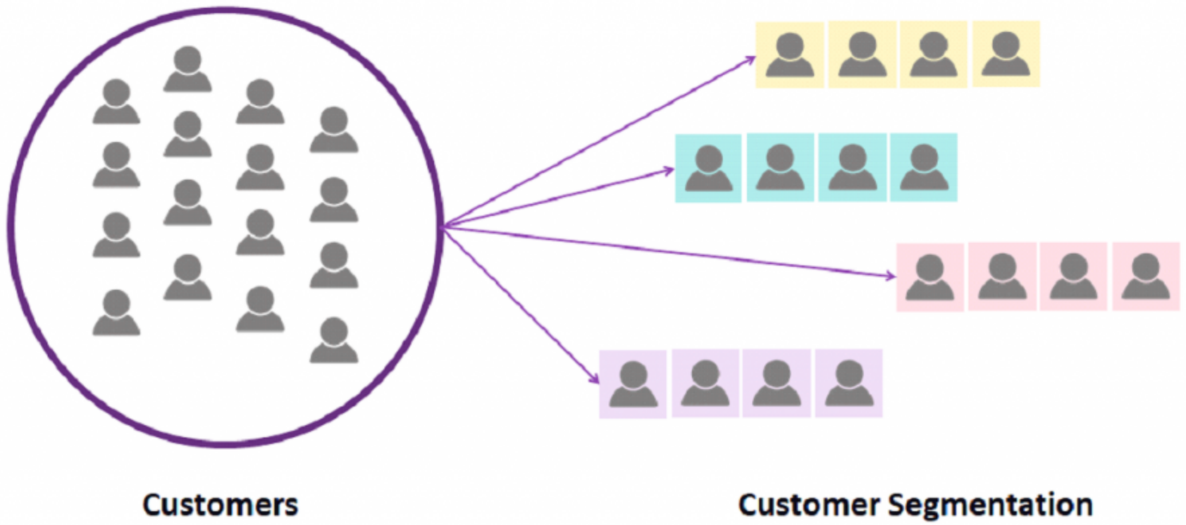
Unsupervised Learning - Business Applications

Google News



Unsupervised Learning - Business Applications

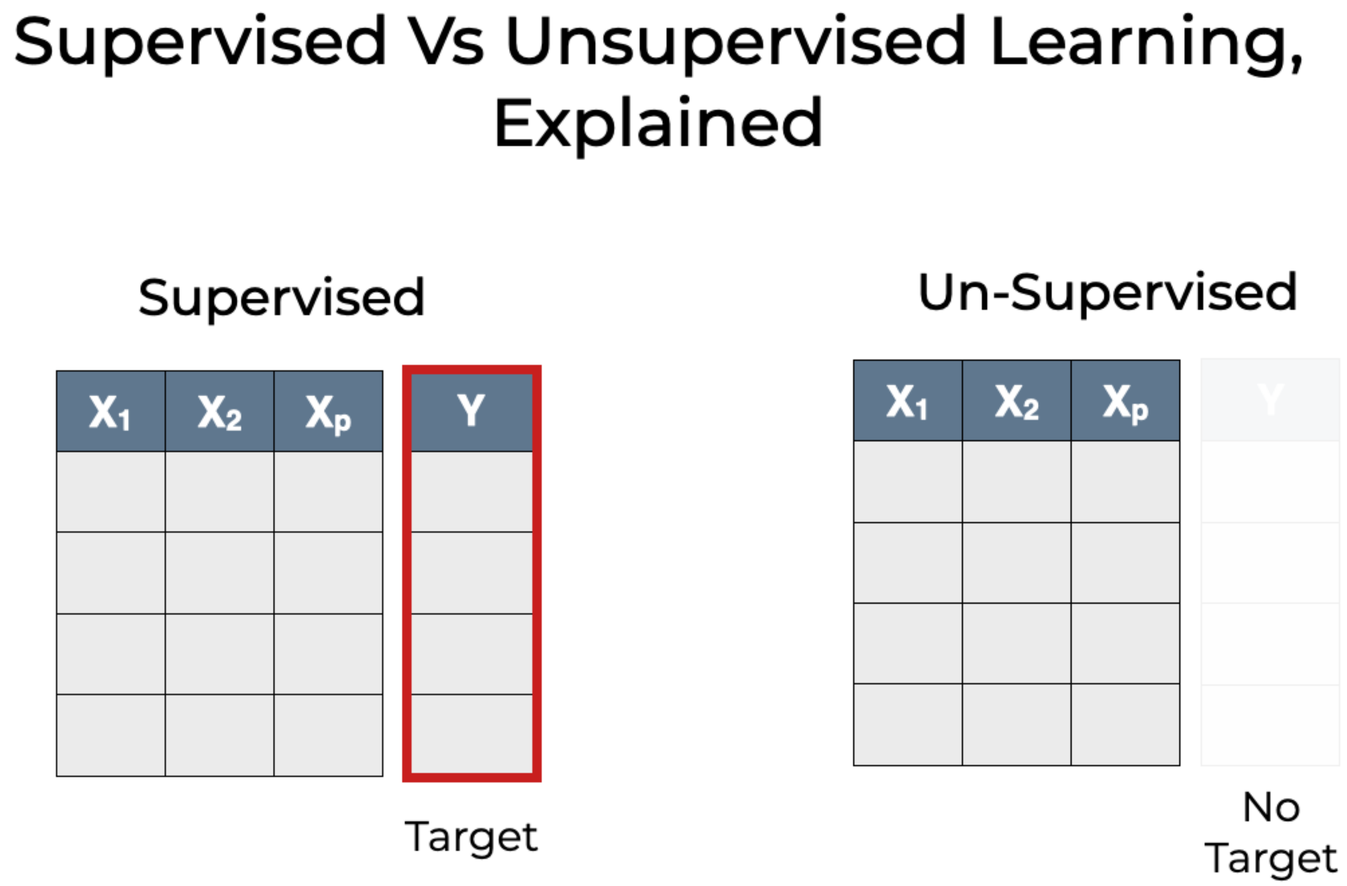
Customer Segmentation



Supervised vs Unsupervised learning: Which is best for you?

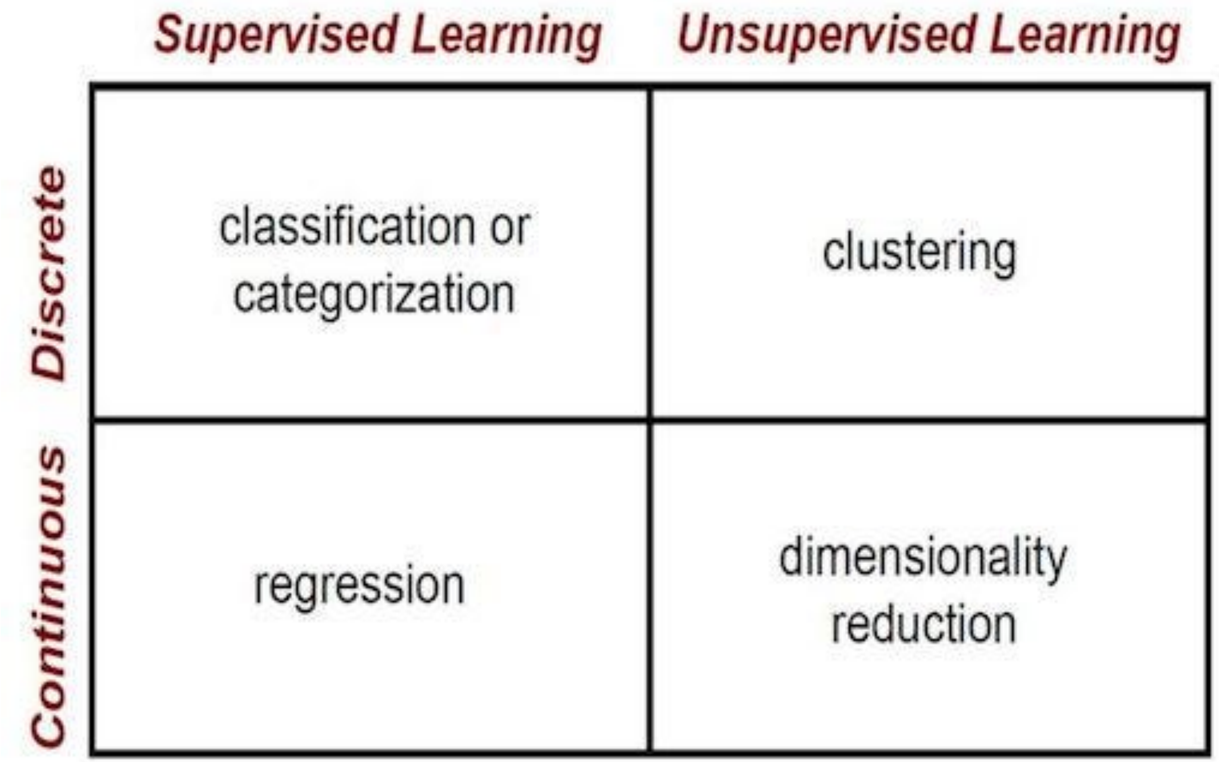
Which one should you use?

The main difference between supervised and unsupervised learning: Labeled data



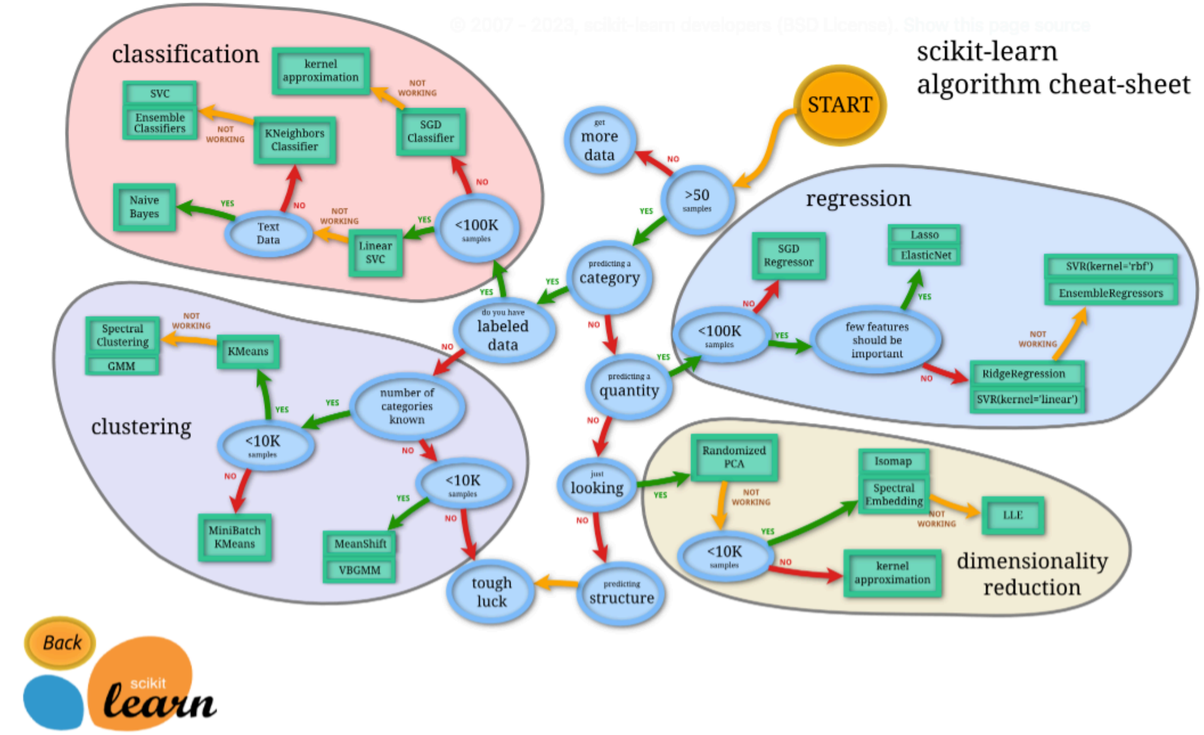
Which one should you use?

The main difference between supervised and unsupervised learning: Labeled data



Which one should you use?

The main difference between supervised and unsupervised learning: Labeled data



Cost Function

Cost Function

Cost Function is used to test if the model is at its best-fit, or still needs improvement.

The smaller the cost function values, the better

Cost Function

Objective

The primary set-up for learning neural networks is to define a cost function (also known as a loss function) that measures how well the network predicts outputs on the test set.

The goal is to then find a set of weights and biases that minimizes the cost.

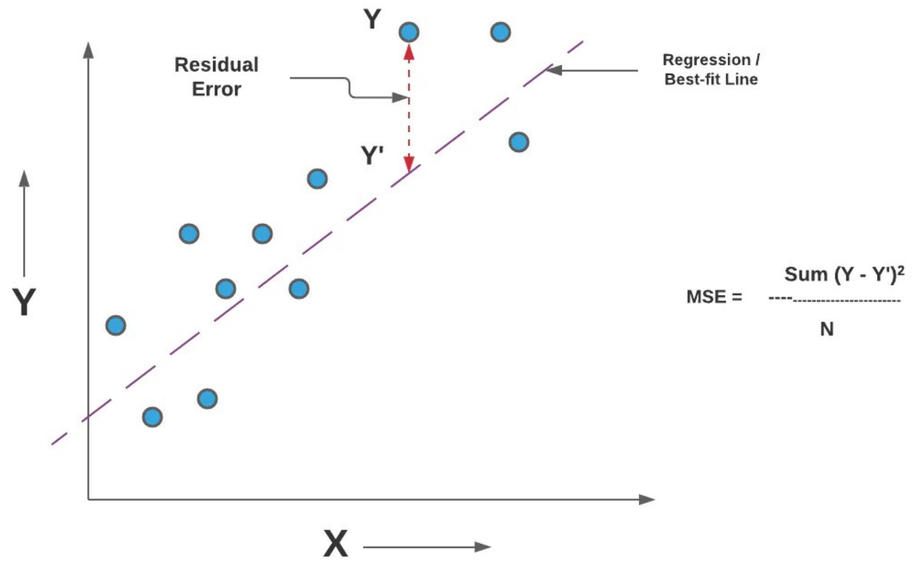
Cost Function

Common functions: Mean Squared Error (MSE)

One common function that is often used is the mean squared error, which measures the difference between the actual value of y and the estimated value of y (the prediction).

Cost Function

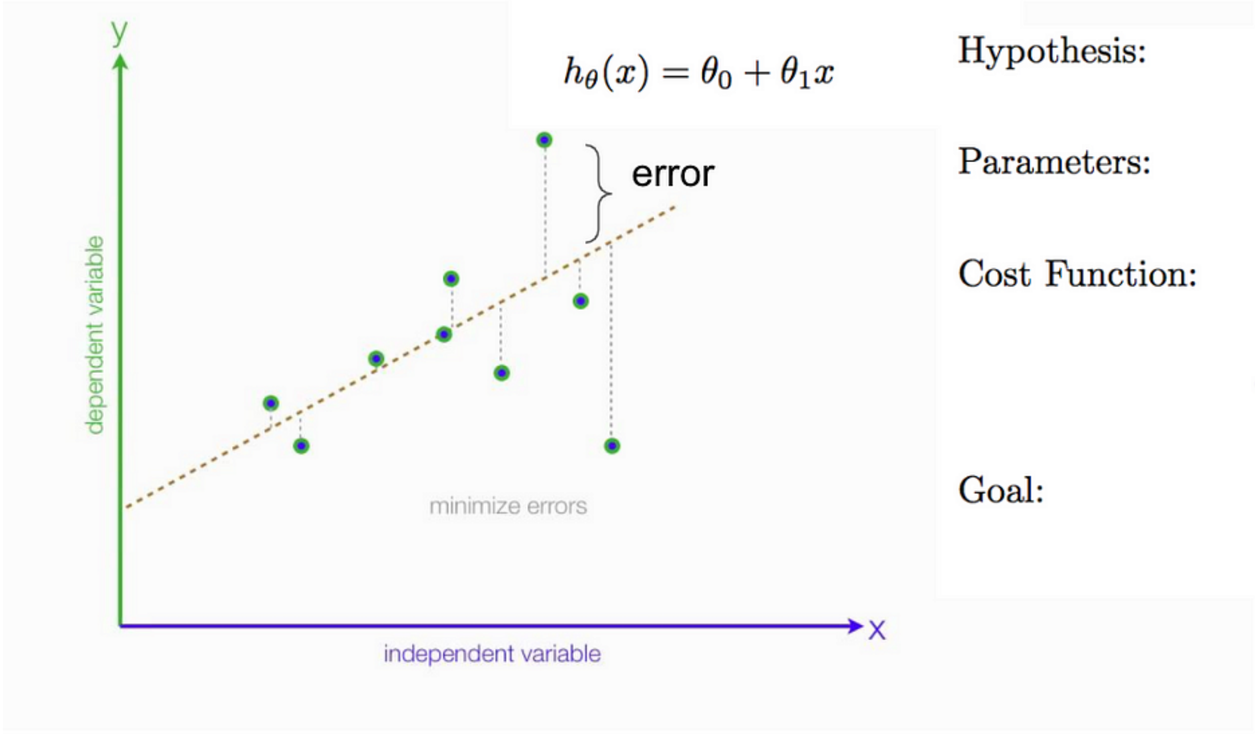
Common functions: Mean Squared Error (MSE)



$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n} (y_i - \tilde{y}_i)^2$$

Cost Function

Common functions: Mean Squared Error (MSE)



$$h_\theta(x) = \theta_0 + \theta_1 x$$

Hypothesis:

$$h_\theta(x) = \theta_0 + \theta_1 x$$

Parameters:

$$\theta_0, \theta_1$$

Cost Function:

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$$

Goal:

$$\underset{\theta_0, \theta_1}{minimize} J(\theta_0, \theta_1)$$

Gradient Descent

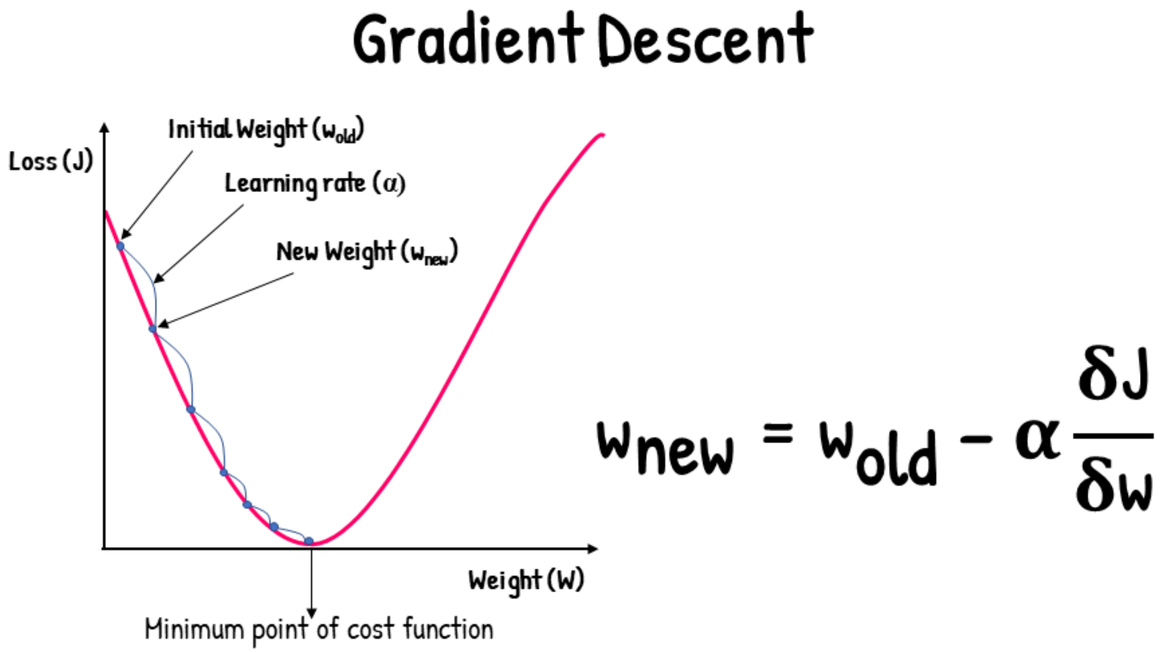
Gradient Descent

Definition

Gradient descent is an iterative optimization algorithm for finding the local minimum of a function. To find the local minimum of a function using gradient descent, we must take steps proportional to the negative of the gradient (move away from the gradient) of the function at the current point.

Gradient Descent

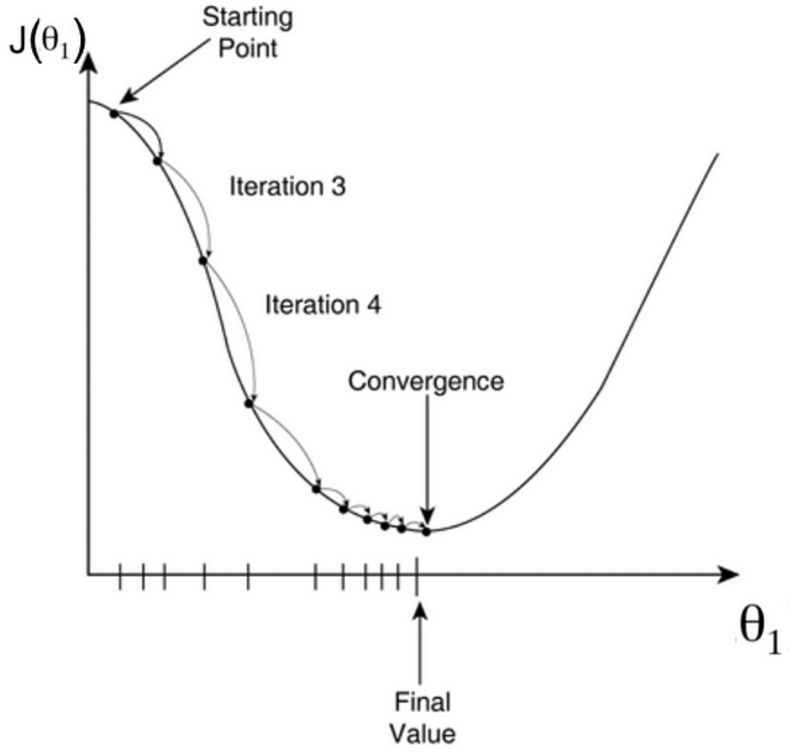
How to minimize cost function?



$$w_{new} = w_{old} - \alpha \frac{\delta J}{\delta w}$$

Gradient Descent

How to minimize cost function?



Cost Function – "One Half Mean Squared Error":

$$J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

Objective:

$$\min_{\theta_0, \theta_1} J(\theta_0, \theta_1)$$

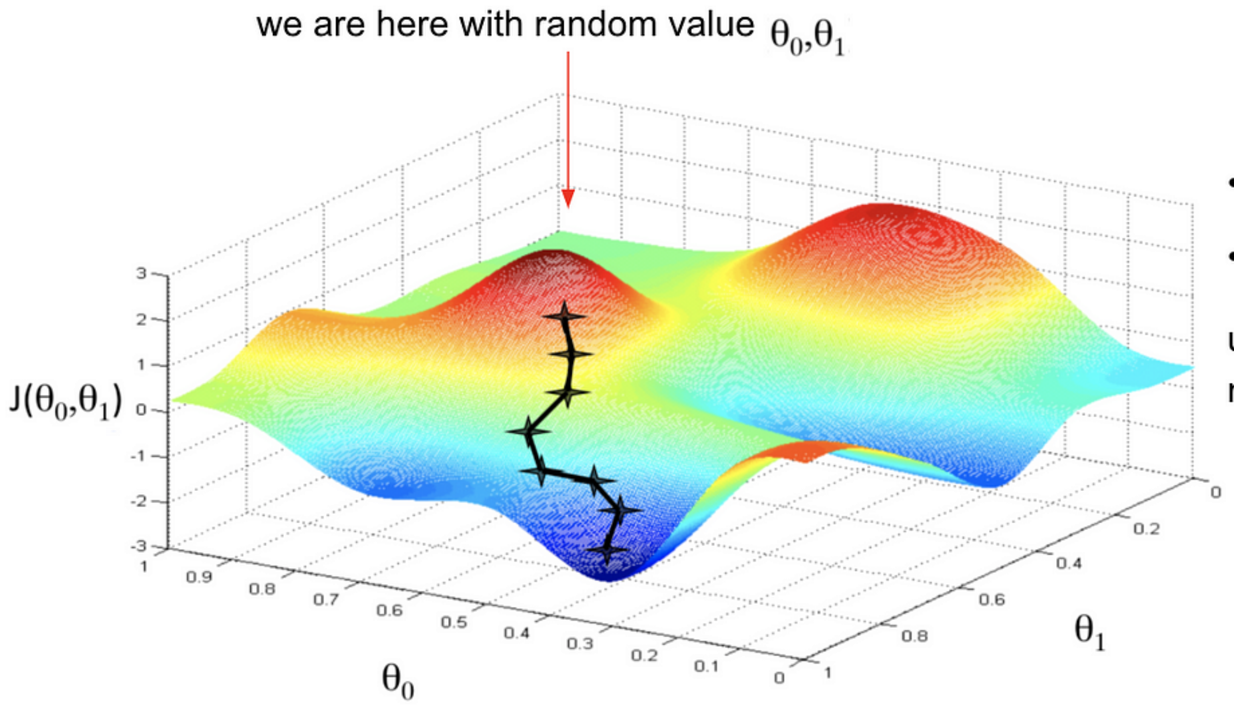
Derivatives:

$$\frac{\partial}{\partial\theta_0}J(\theta_0, \theta_1) = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)$$

$$\frac{\partial}{\partial\theta_1}J(\theta_0, \theta_1) = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)\cdot x^{(i)}$$

Gradient Descent

How to minimize cost function?



- Start with some θ_0,θ_1
- Keep changing θ_0,θ_1 to reduce $J(\theta_0,\theta_1)$

until we hopefully end up at a minimum

Gradient Descent

How to minimize cost function?

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2 \qquad [1.0]$$

$$\frac{\partial}{\partial\theta_0}J(\theta_0,\theta_1) = \frac{\partial}{\partial\theta_0}\left(\frac{1}{m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2\right) \qquad [1.1]$$

$$= \frac{1}{m}\sum_{i=1}^{m}\frac{\partial}{\partial\theta_0}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2 \qquad [1.2]$$

$$= \frac{1}{m}\sum_{i=1}^{m}2\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)\frac{\partial}{\partial\theta_0}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) \qquad [1.3]$$

$$= \frac{2}{m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) \qquad [1.4]$$

Gradient Descent

What is the size of the steps we take to reach the minimum?

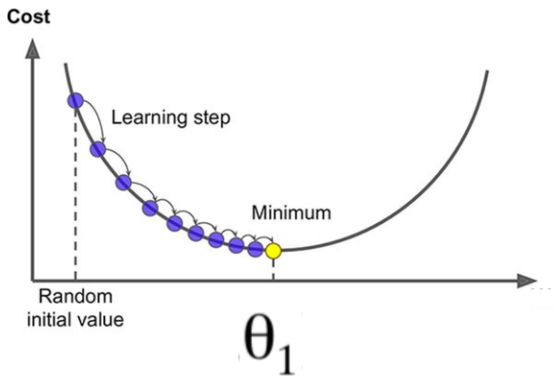
This is called ***"Learning Rate"***

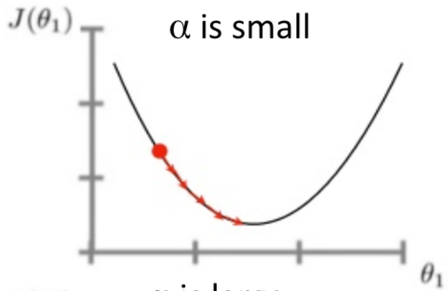
repeat until convergence {

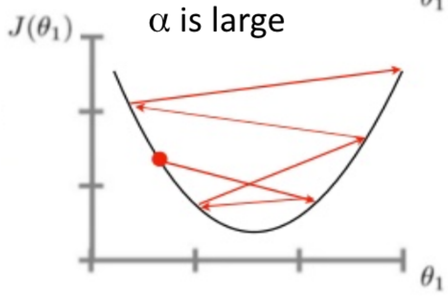
$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

(for $j = 1$ and $j = 0$)

}



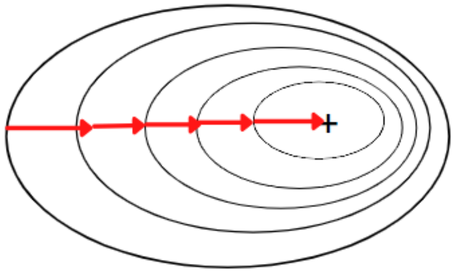




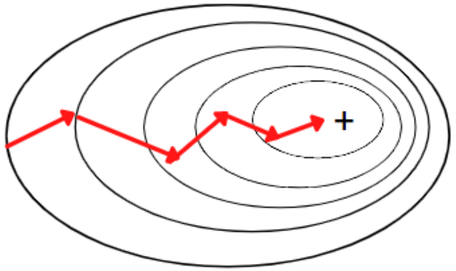
Gradient Descent

3 Gradient Descent Methods

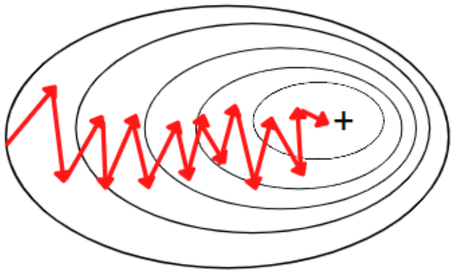
Batch Gradient Descent



Mini-Batch Gradient Descent

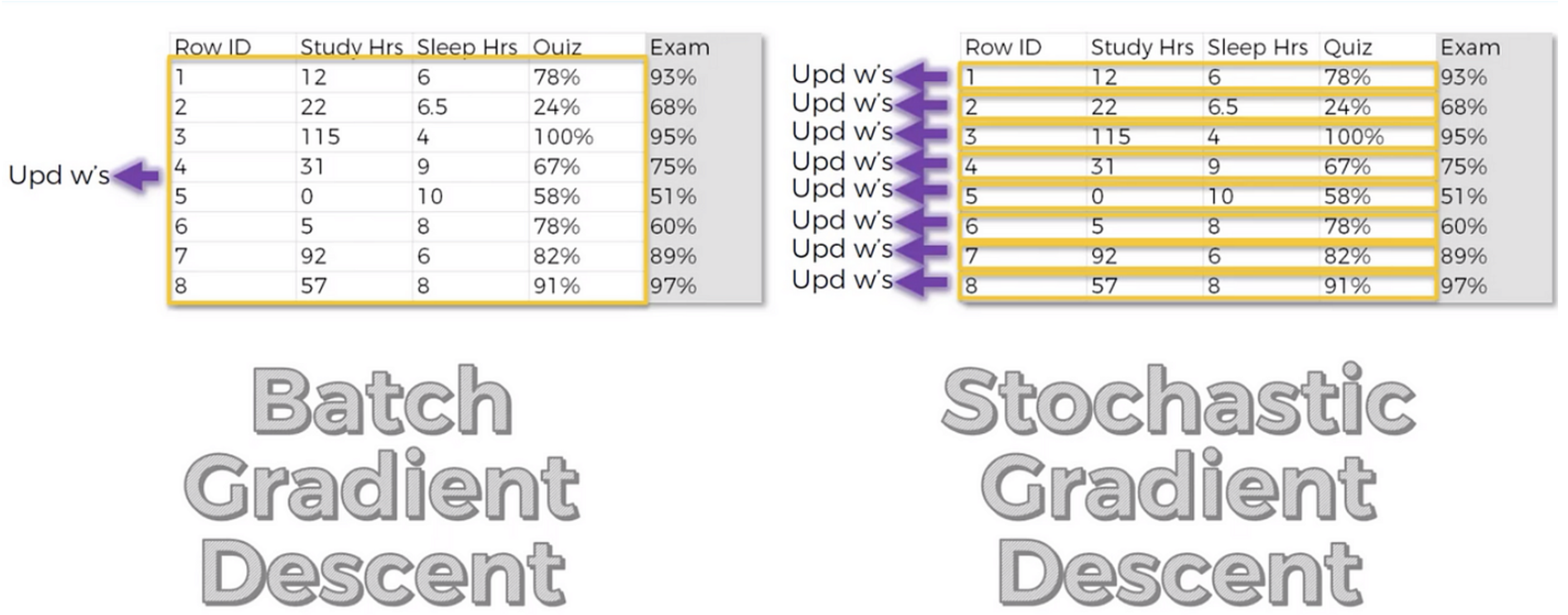


Stochastic Gradient Descent



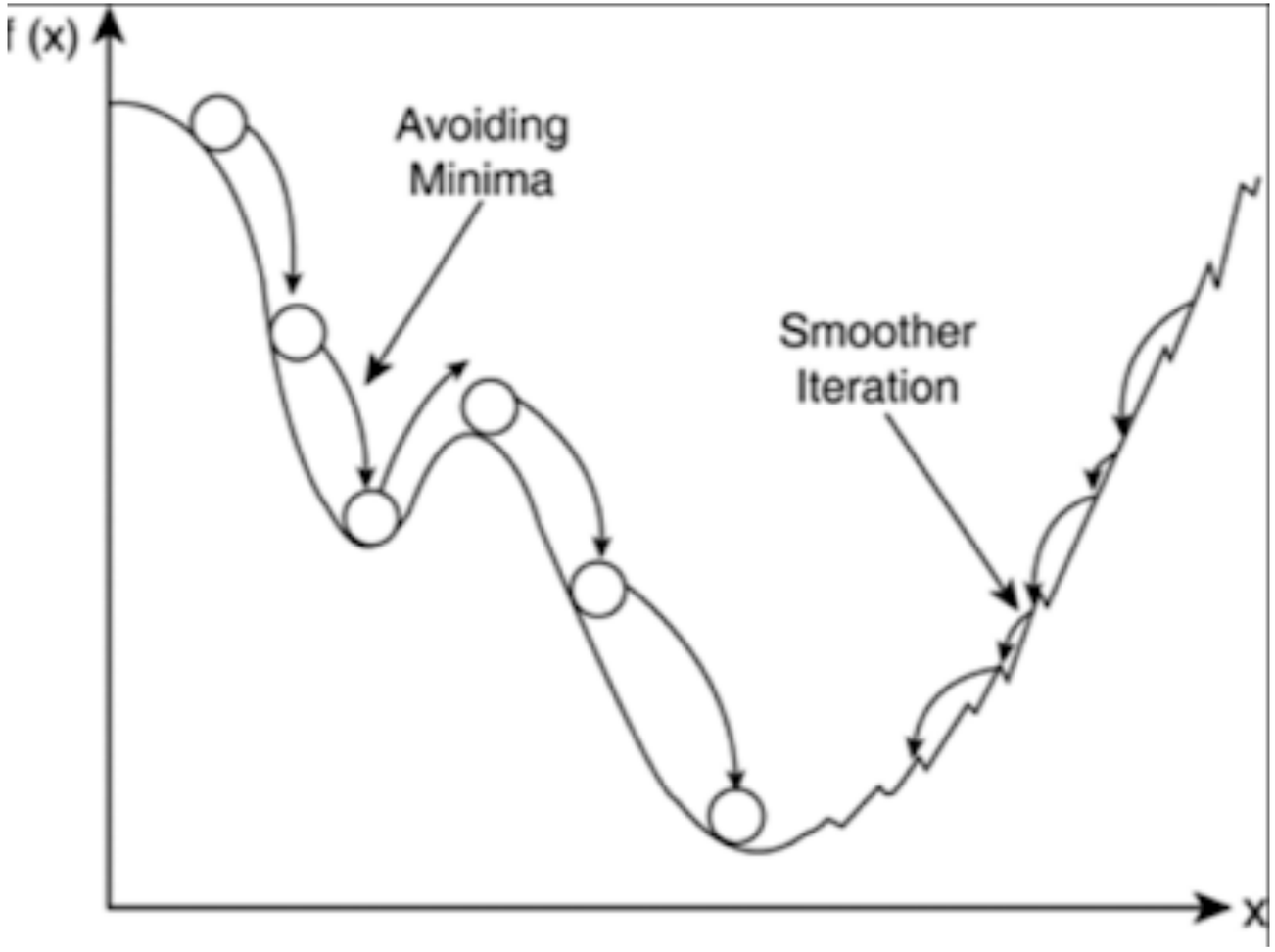
Gradient Descent

Different Methods Example



Gradient Descent - Challenges

Local Minima Problem



Gradient Descent - Challenges

Local Minima Problem

Batch Gradient Descent (BGD): Calculates gradients on the entire dataset.

Challenges: Too Slow, and will never reach the global minima

Gradient Descent - Challenges

Local Minima Problem - Solution

Stochastic Gradient Descent (SGD): Calculates gradients on a single sample of data every training step.

Challenges: network takes longer to converge because it calculates gradients on each single point.

Main Advantage: Converges to ***global minima.***

Gradient Descent - Mini-Batch

Optimal Solution

Mini-Batch Gradient Descent (MBGD): Best of SGD and BGD.

It takes a few random points and calculate the gradients based on them.

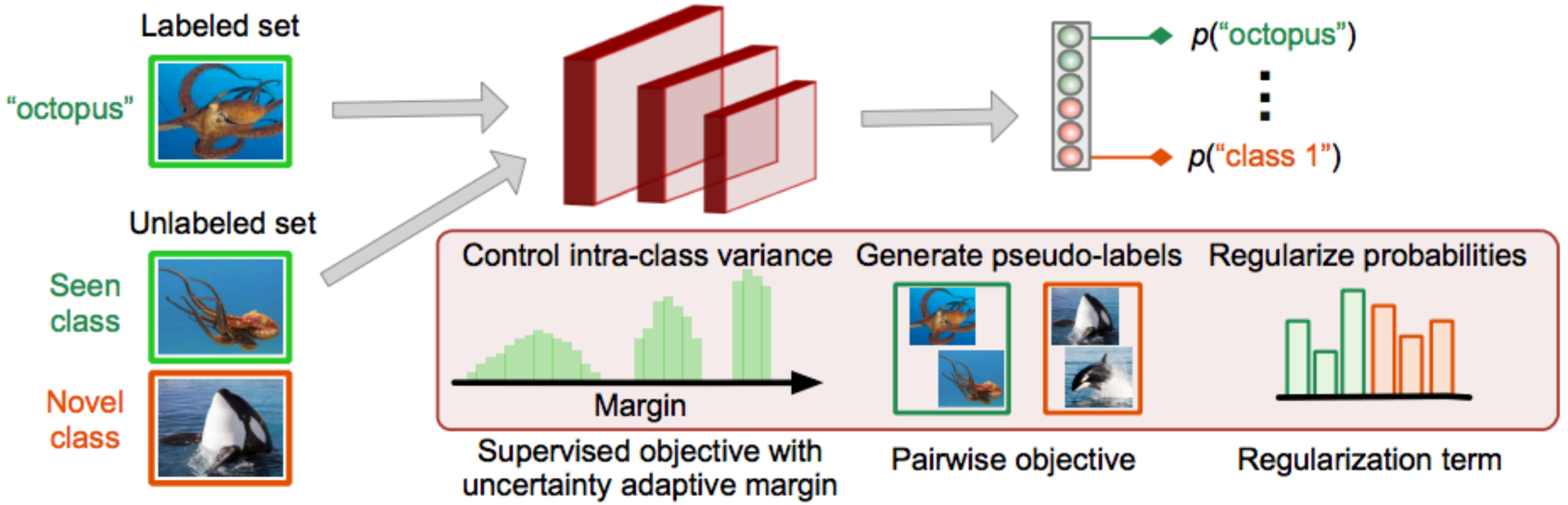
Main Advantages: Fast, and bounces out of local minima.

BONUS: Self and Semi Supervised

Semi-Supervised Learning

In a nutshell, semi-supervised learning (SSL) is a machine learning technique that uses a small portion of labeled data and lots of unlabeled data to train a predictive model.

Semi-Supervised Learning



Self-supervised Learning

Self-supervised learning is based on an artificial neural network and can be considered halfway between supervised and unsupervised learning. I

Key points:

1. **No Manual Label**
2. **Data Efficiency**

Self-supervised Learning

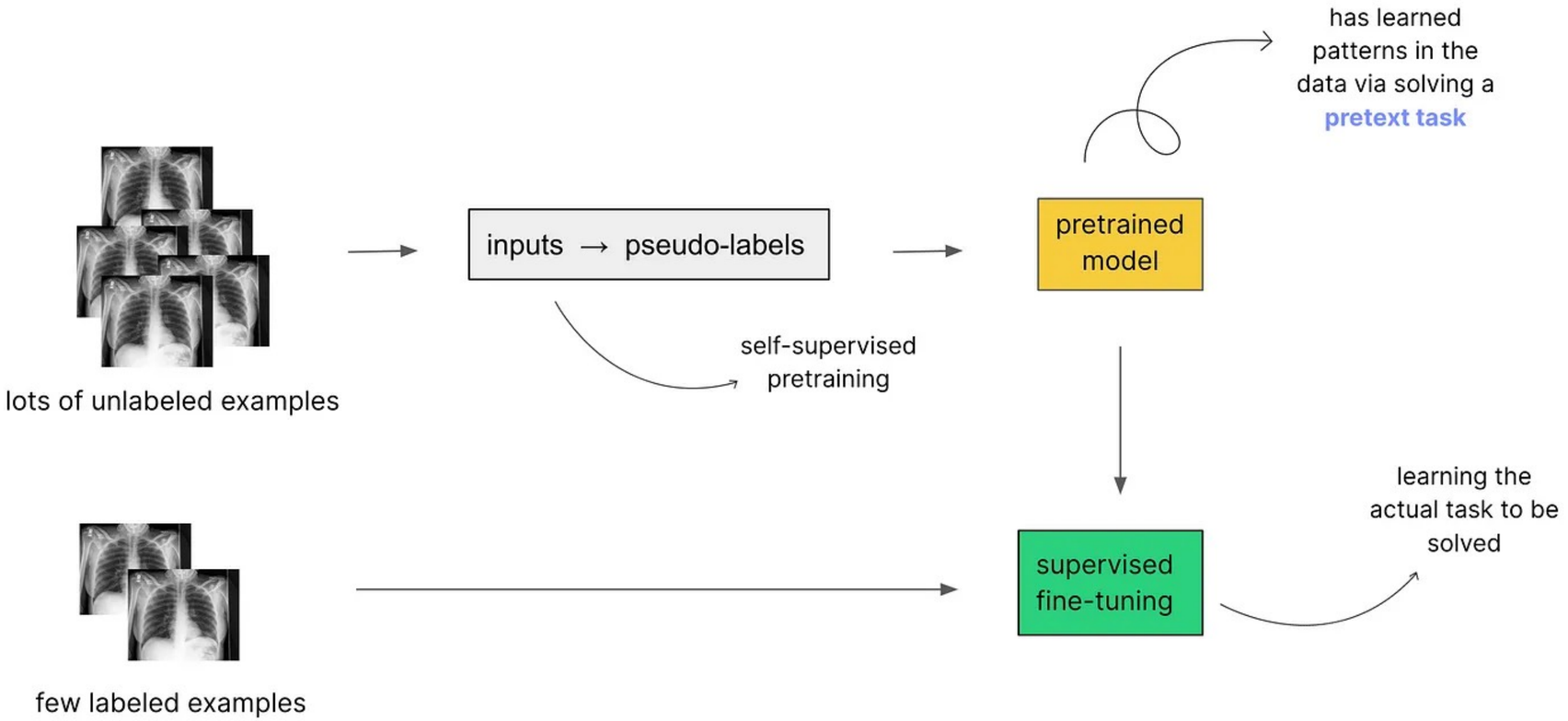
Yann LeCun - Chief AI Scientist at Meta

Self-supervised learning is one of the most promising ways to build background knowledge and approximate a form of common sense in AI systems.

~ Yann LeCun

Self-Supervised Learning

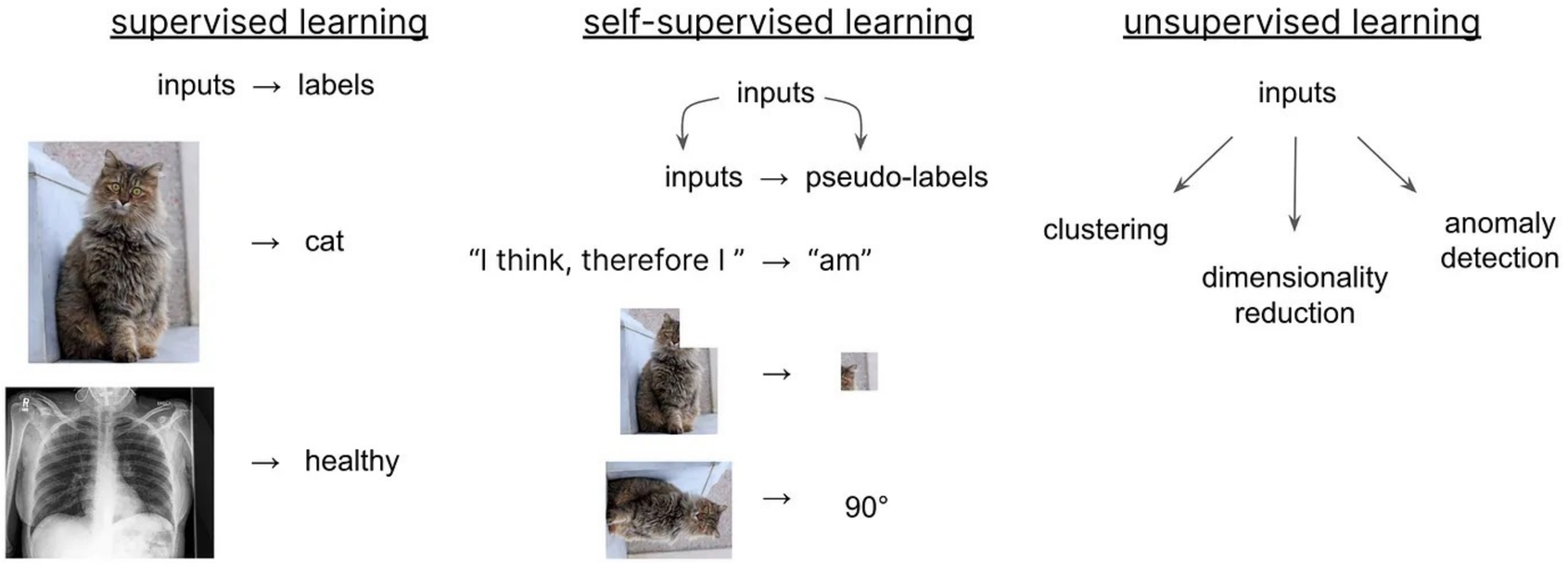
Example



Self-Supervised Learning

Example

Models trained in a self-supervised way create their own pseudo-labels from unlabeled data.



Let’s skip to the good part!!

End of the course
Let’s apply everything we learned in the Practice Course!