INTRODUCTION TO MACHINE LEARNING

BY

SAIFUL BARI IFTU
LECTURER, DEPT. OF CSE, BRAC UNIVERSITY

CONTENTS

What Is Machine Learning (ML)?

Why ML?

ML in Practice

ML Notations & Terminologies

Types of ML

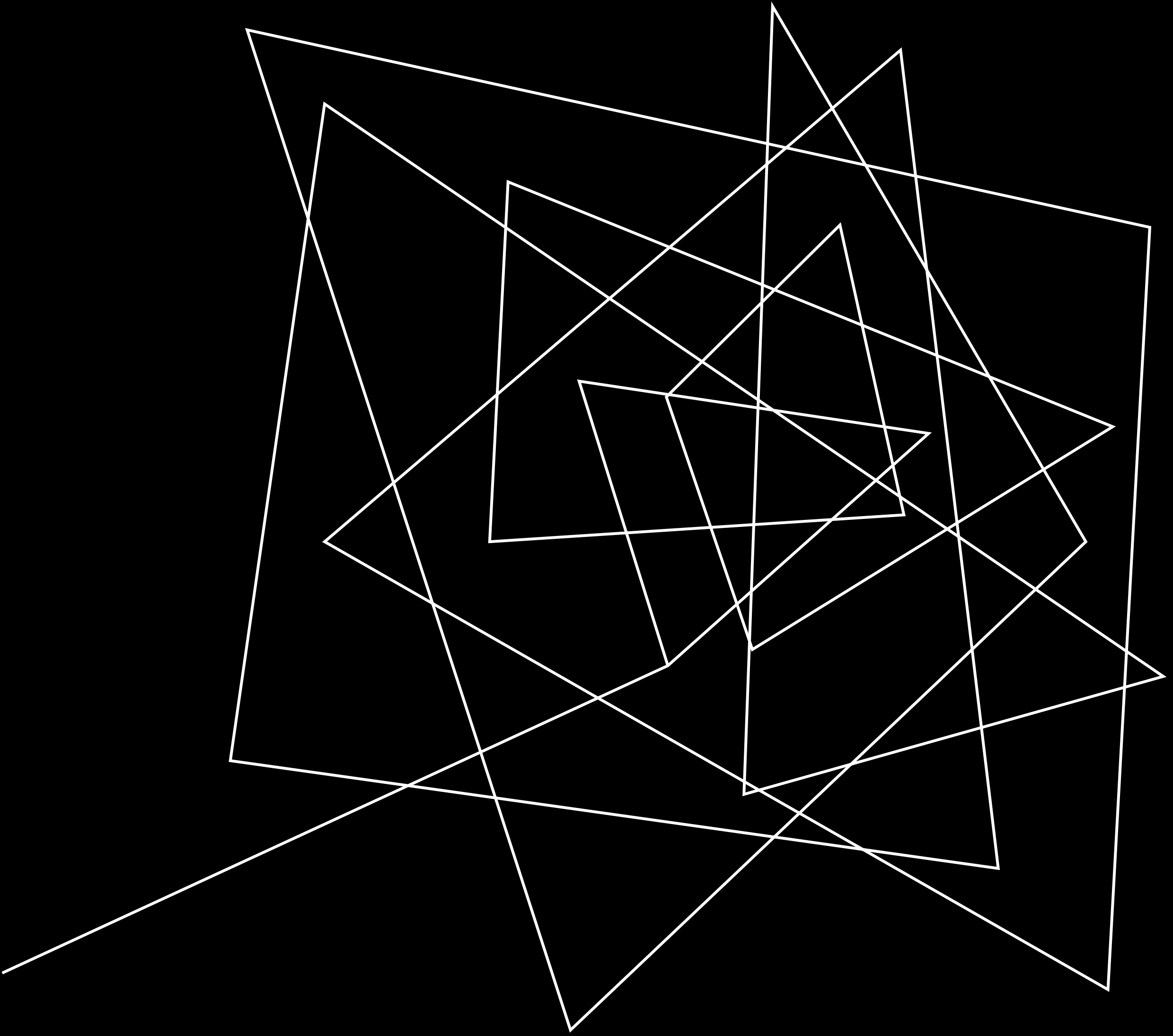
Overview of an ML Algorithm

Popular ML Algorithms

When to use ML?

Challenges with ML

ML Ethics



WHAT IS MACHINE LEARNING?

WHAT IS MACHINE LEARNING?

Formal Definition

- Machine Learning (ML) is a subfield of artificial intelligence (AI) that focuses on developing algorithms and statistical models that enable computers to perform tasks without explicit instructions. Instead of being programmed with specific rules to solve a problem, a machine learning model is trained on data to learn patterns, relationships, and structures, allowing it to make predictions or decisions based on new data.

In Short

- Machine learning is a field of study that gives computers the ability to learn without being explicitly programmed. *(Arthur Samuel, 1959.)*

DS vs AI vs ML vs DL

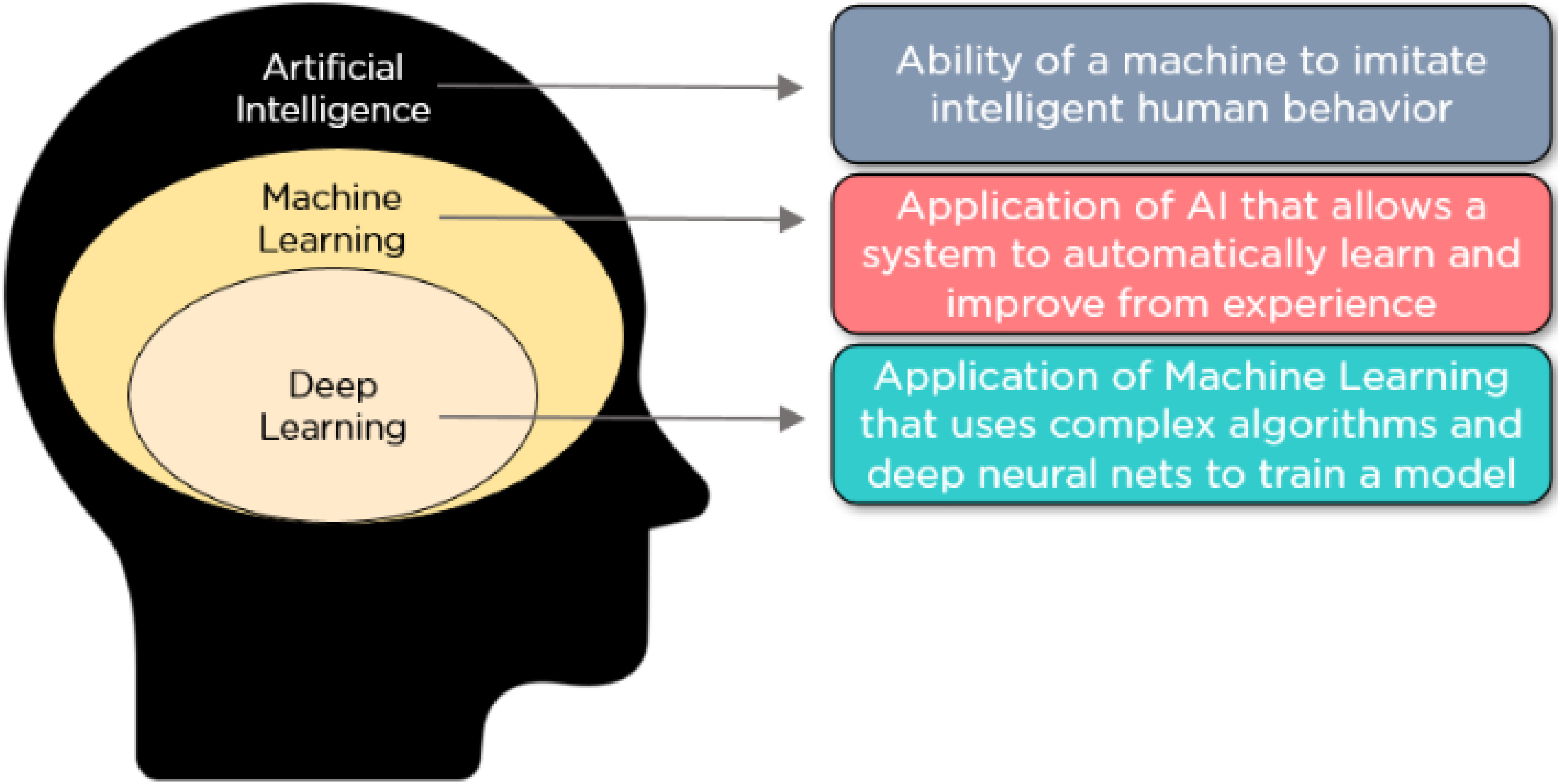
Data Science: An interdisciplinary field that uses statistical methods, algorithms, and technology to extract insights and knowledge from structured and unstructured data.

Artificial Intelligence (AI): The simulation of human intelligence in machines, enabling them to perform tasks that typically require human cognition, such as learning, reasoning, and problem-solving.

Machine Learning (ML): A subset of AI focused on developing algorithms that allow computers to learn from and make decisions based on data without being explicitly programmed.

Deep Learning (DL): A subset of machine learning that uses **neural networks with many layers** (deep neural networks) to model and solve complex patterns and relationships in data, often achieving state-of-the-art results in tasks like image and speech recognition.

DS vs AI vs ML vs DL



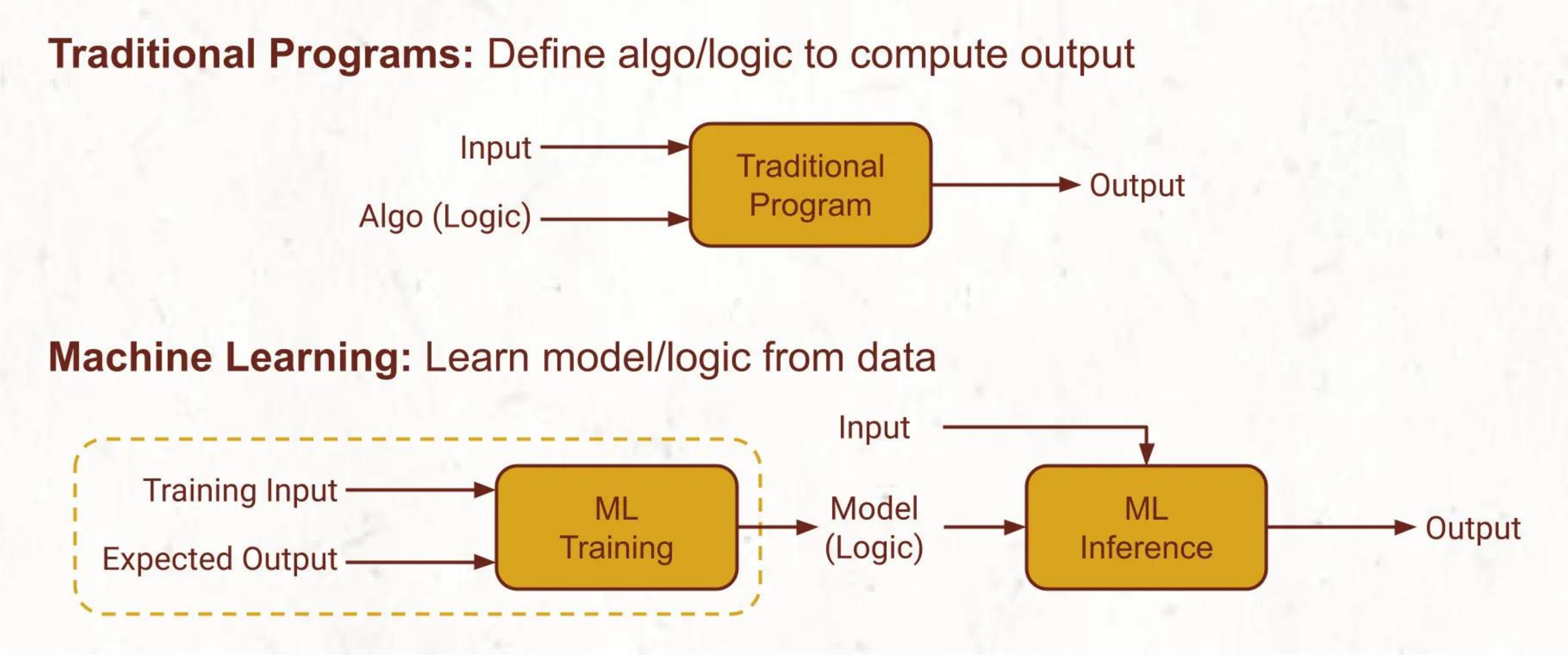
Ability of a machine to imitate intelligent human behavior

Application of AI that allows a system to automatically learn and improve from experience

Application of Machine Learning that uses complex algorithms and deep neural nets to train a model

Source: https://medium.com/@ramaleelamadishetti/regularization-in-deep-learning-6610207409e2

MACHINE LEARNING



Traditional Programs: Define algo/logic to compute output

Input

Algo (Logic)

Traditional Program

Output

Machine Learning: Learn model/logic from data

Training Input

Expected Output

ML Training

Model (Logic)

Input

ML Inference

Output

Source: https://ml4devs.substack.com/p/011-machine-learning-vs-traditional-software-development

MACHINE LEARNING

- “A computer program is said to learn if its performance at a task **T**, as measured by a performance **P**, improves with experience, ***E***.”

  -- Tom Mitchell (1997)

- Three major components:

  1. ***Task, T***
  2. ***Performance metric, P***
  3. ***Experience, E***

MACHINE LEARNING (EXAMPLES)

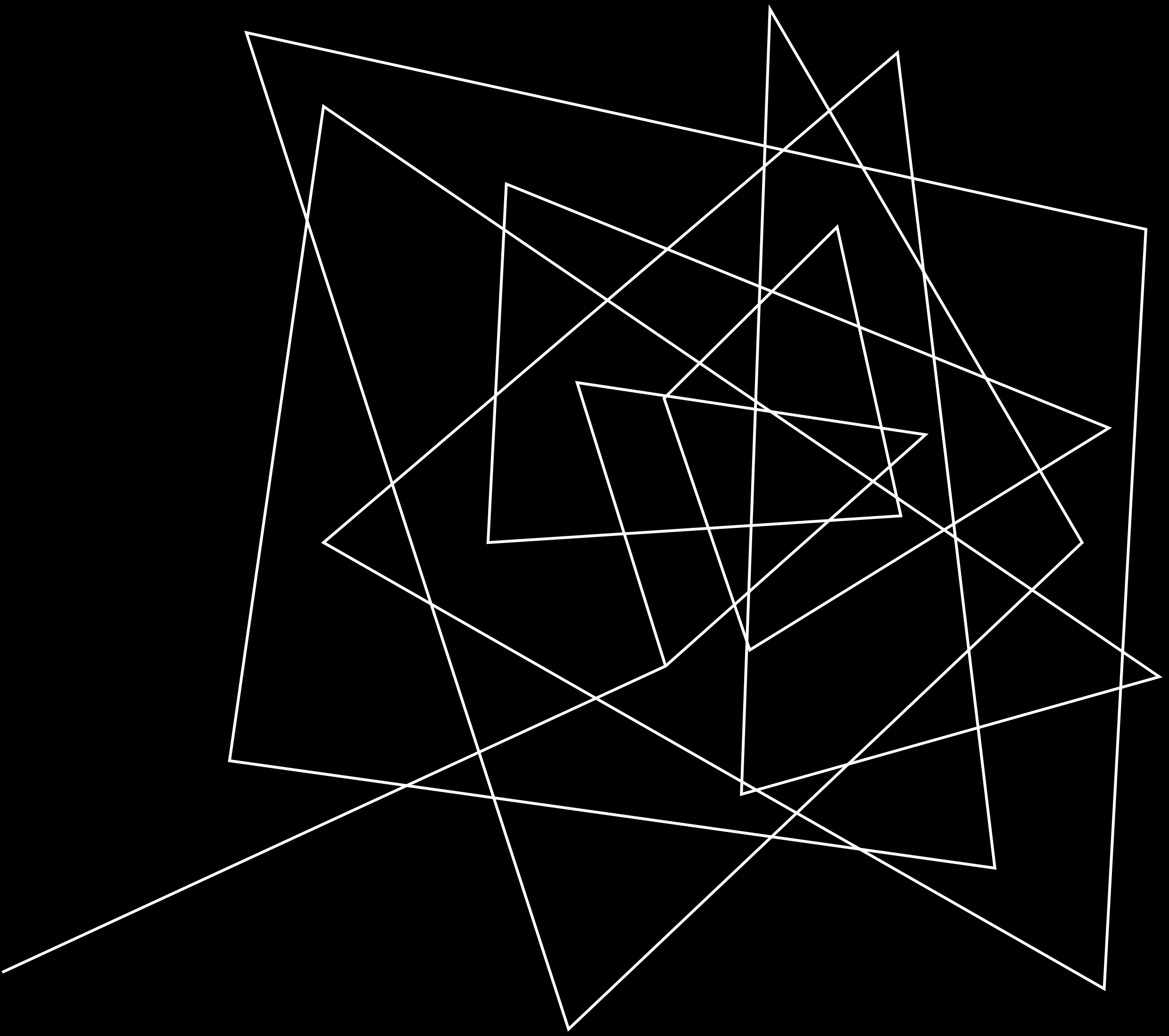
Example 1: Learning to approve loans

1. **Task, T:** Decide whether to extend someone a loan
2. **Performance metric, P:** Number of people who default on their loan
3. **Experience, E:** Interviews with loan officers.

Example 2: Predicting Stock Prices

1. **Task, T:** Predict the price of a stock market share on a date in the future
2. **Performance metric, P:** Difference between predicted price & actual price
3. **Experience, E:** Historical data on stock prices

*Typically, experience **E** refers to training data, and performance metric **P** depends on the quality of the output generated using that data.*

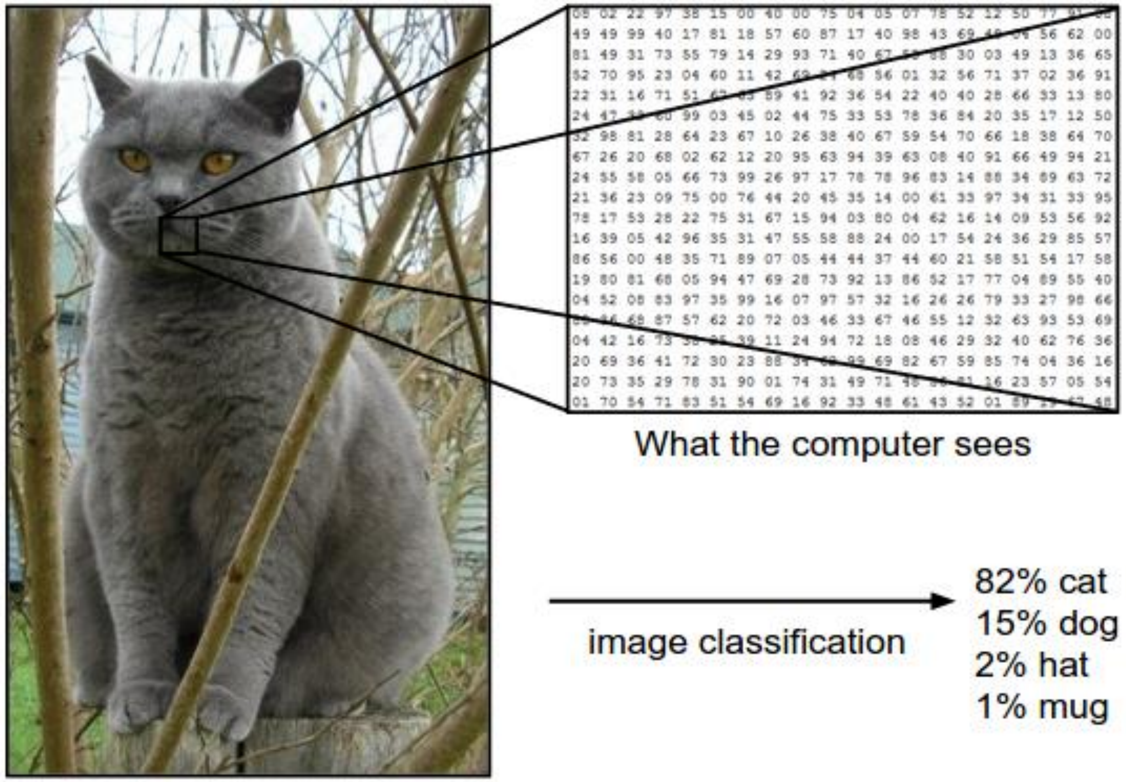


WHY
MACHINE LEARNING?

WHY ML?

- **Handling Complex Problems:** Traditional programming requires explicitly defining rules for every possible scenario. ML excels in situations where it's impractical or impossible to define all the rules, such as image recognition, natural language processing, or predicting trends.
- **Pattern Recognition:** ML is particularly strong in recognizing patterns in data that are not easily observable through traditional programming.
- **Adaptability:** ML models can adapt to new, unseen data by adjusting their parameters based on learned patterns, making them more flexible and robust in dynamic environments compared to hard-coded logic.
- **Automation:** ML can automate decision-making processes that are too complex for manual intervention, such as predicting maintenance needs in industrial settings or optimizing supply chain logistics.

WHY ML (IN COMPUTER VISION)?



08	02	22	97	38	15	00	40	00	75	04	05	07	78	52	12	50	77	91	08
49	49	99	40	17	81	18	57	60	87	17	40	98	43	69	48	04	56	62	00
81	49	31	73	55	79	14	29	93	71	40	67	53	88	30	03	49	13	36	65
52	70	95	23	04	60	11	42	69	24	68	56	01	32	56	71	37	02	36	91
22	31	16	71	51	67	63	89	41	92	36	54	22	40	40	28	66	33	13	80
24	47	32	60	99	03	45	02	44	75	33	53	78	36	84	20	35	17	12	50
32	98	81	28	64	23	67	10	26	38	40	67	59	54	70	66	18	38	64	70
67	26	20	68	02	62	12	20	95	63	94	39	63	08	40	91	66	49	94	21
24	55	58	05	66	73	99	26	97	17	78	78	96	83	14	88	34	89	63	72
21	36	23	09	75	00	76	44	20	45	35	14	00	61	33	97	34	31	33	95
78	17	53	28	22	75	31	67	15	94	03	80	04	62	16	14	09	53	56	92
16	39	05	42	96	35	31	47	55	58	88	24	00	17	54	24	36	29	85	57
86	56	00	48	35	71	89	07	05	44	44	37	44	60	21	58	51	54	17	58
19	80	81	68	05	94	47	69	28	73	92	13	86	52	17	77	04	89	55	40
04	52	08	83	97	35	99	16	07	97	57	32	16	26	26	79	33	27	98	66
88	36	68	87	57	62	20	72	03	46	33	67	46	55	12	32	63	93	53	69
04	42	16	73	38	25	39	11	24	94	72	18	08	46	29	32	40	62	76	36
20	69	36	41	72	30	23	88	34	62	99	69	82	67	59	85	74	04	36	16
20	73	35	29	78	31	90	01	74	31	49	71	48	86	81	16	23	57	05	54
01	70	54	71	83	51	54	69	16	92	33	48	61	43	52	01	89	19	67	48

What the computer sees

image classification

82% cat
15% dog
2% hat
1% mug

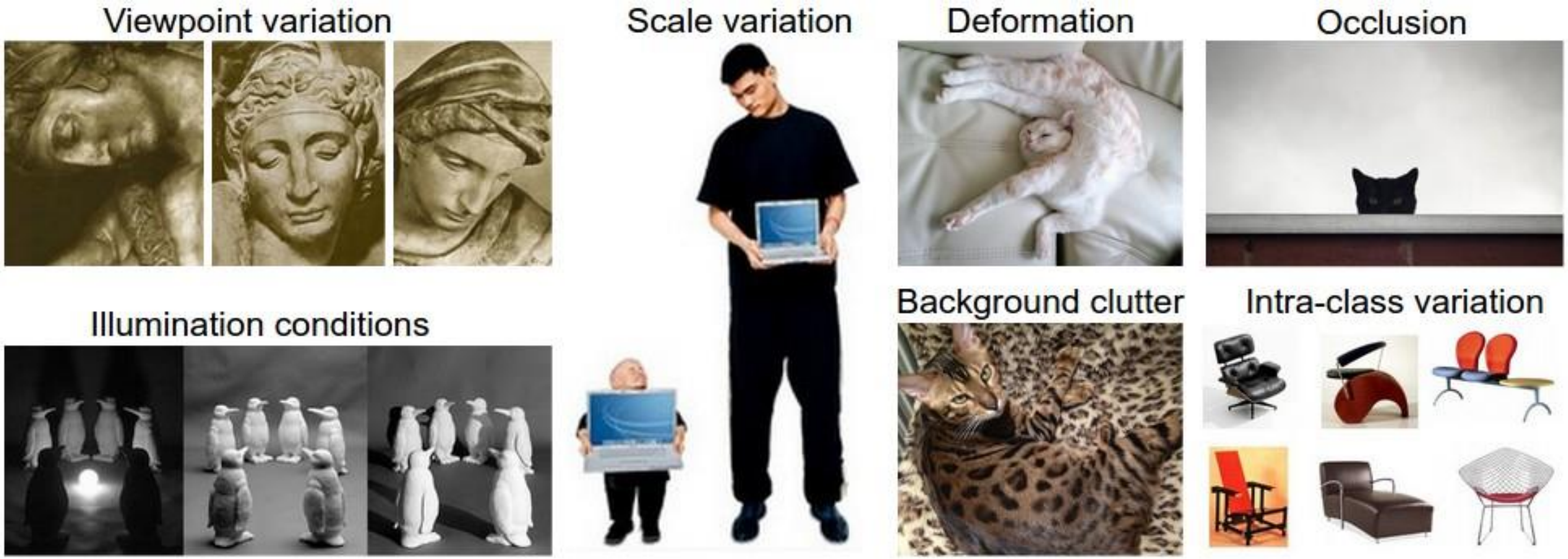
Source: Stanford cs231n Course Notes (https://cs231n.github.io/classification/)

- Very difficult to find patterns manually from pixels.

- Distinguishing between different classes can be challenging too solely from pixel values.

- Pixel values can vary wildly within the same class due to certain variations (color of the cat for example).

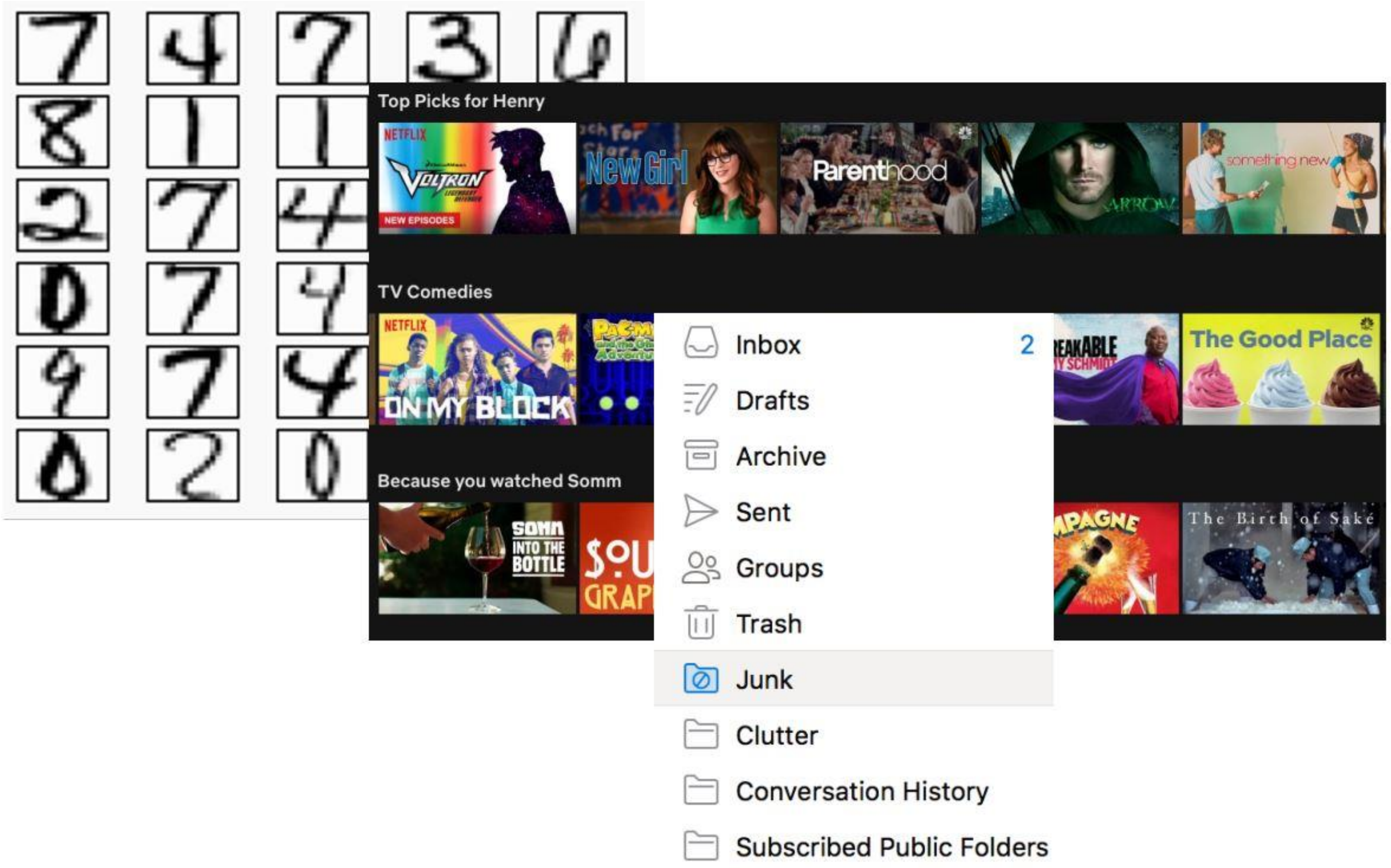
WHY ML (IN COMPUTER VISION)?



Source: Stanford cs231n Course Notes (https://cs231n.github.io/classification/)

Almost Impossible to code for all possible variations!

ML IN PRACTICE

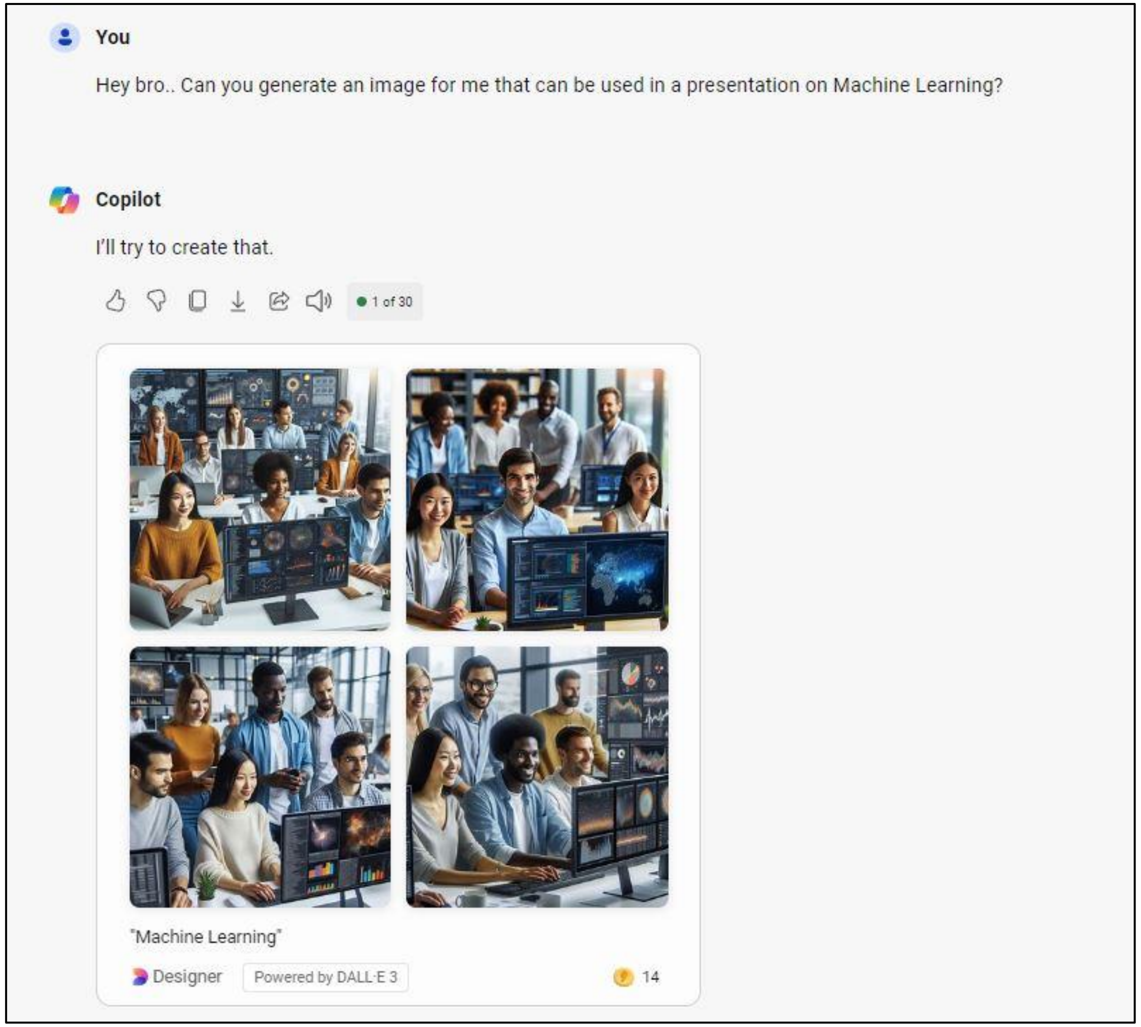


Source: CMU's Introduction to Machine Learning (10-601) Lectures

ML IN PRACTICE

Source: CMU's Introduction to Machine Learning (10-601) Lectures

ML IN PRACTICE (NOWADAYS)

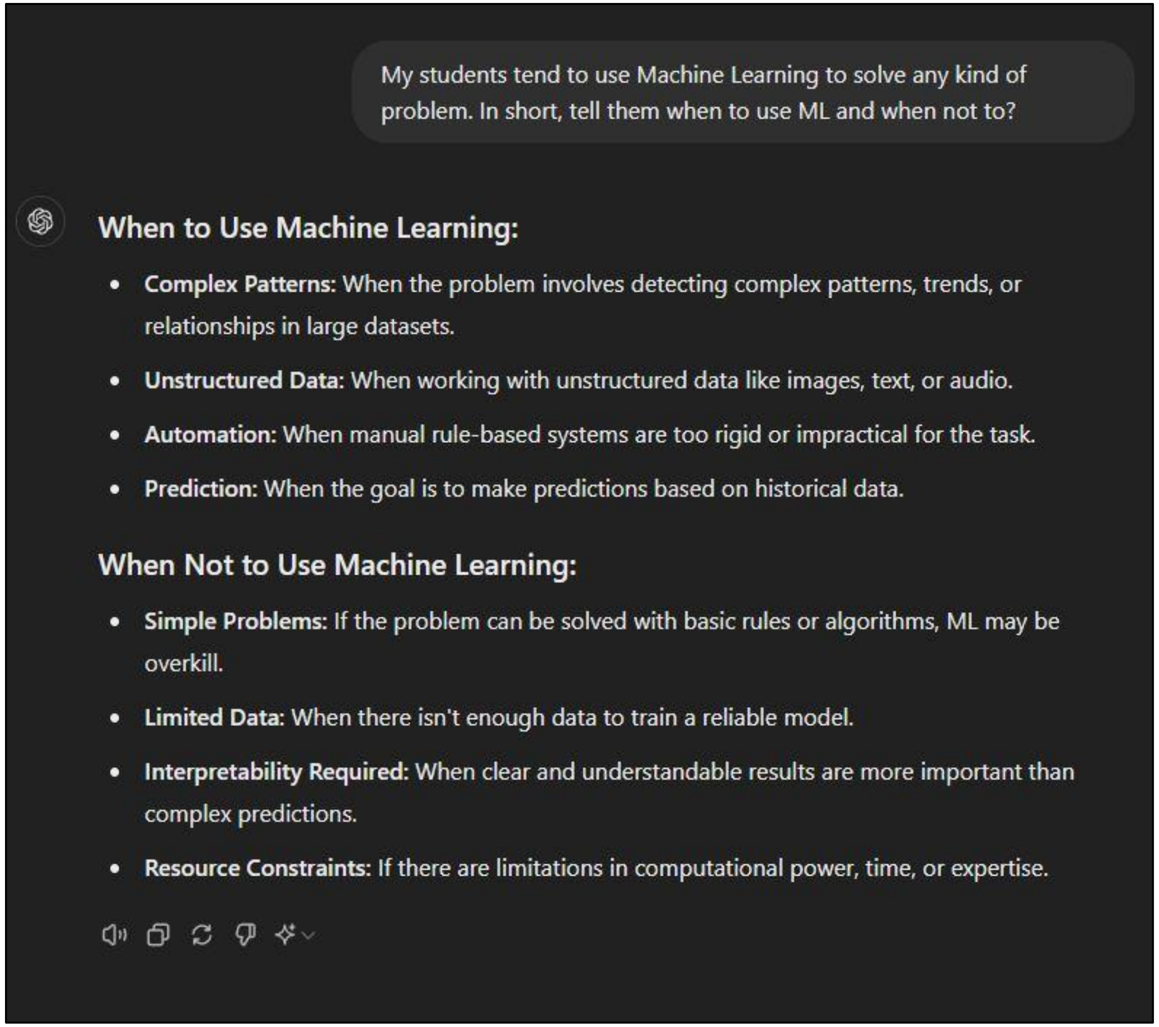


You

Hey bro.. Can you generate an image for me that can be used in a presentation on Machine Learning?

Copilot

I'll try to create that.



My students tend to use Machine Learning to solve any kind of problem. In short, tell them when to use ML and when not to?

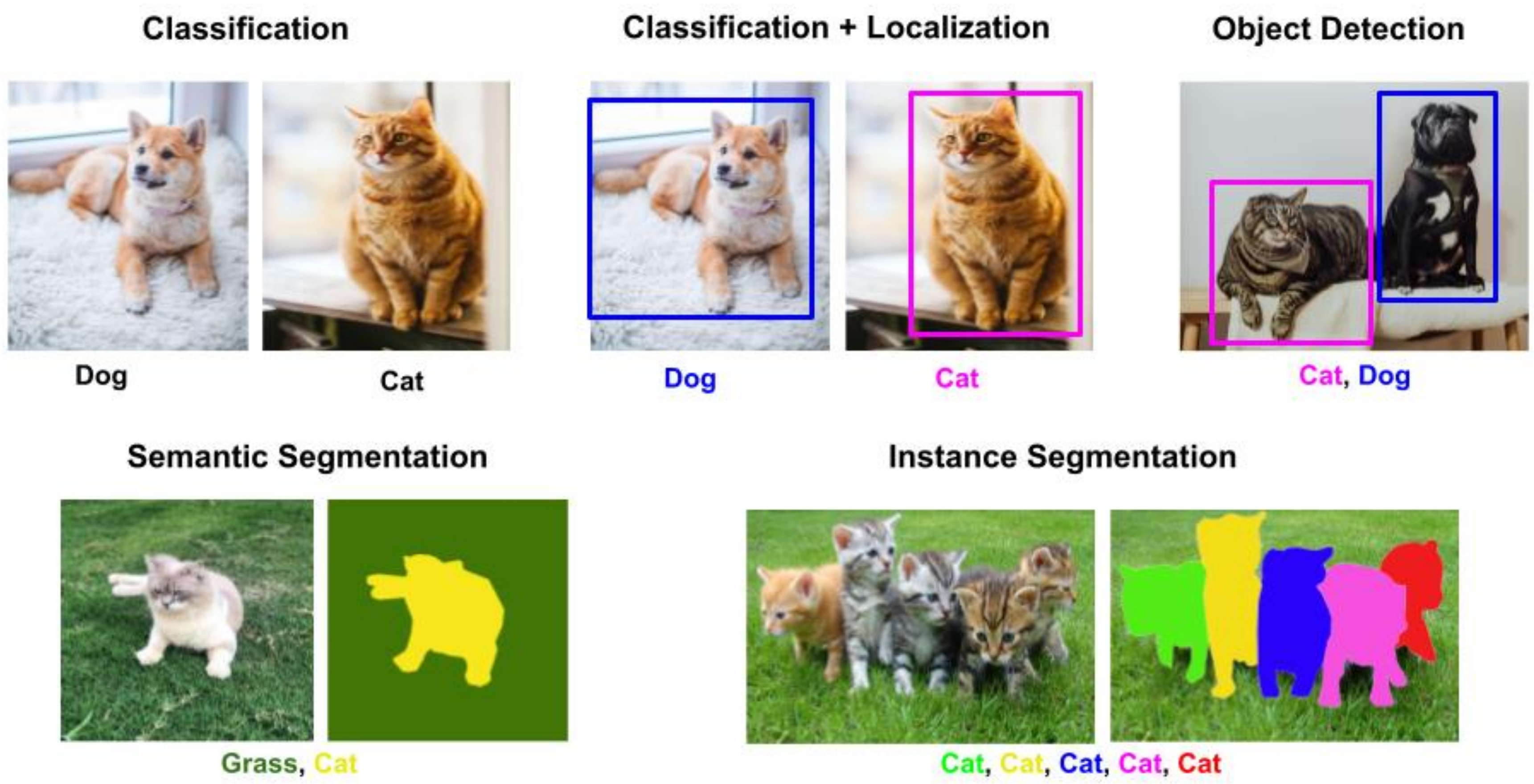
When to Use Machine Learning:

- **Complex Patterns:** When the problem involves detecting complex patterns, trends, or relationships in large datasets.
- **Unstructured Data:** When working with unstructured data like images, text, or audio.
- **Automation:** When manual rule-based systems are too rigid or impractical for the task.
- **Prediction:** When the goal is to make predictions based on historical data.

When Not to Use Machine Learning:

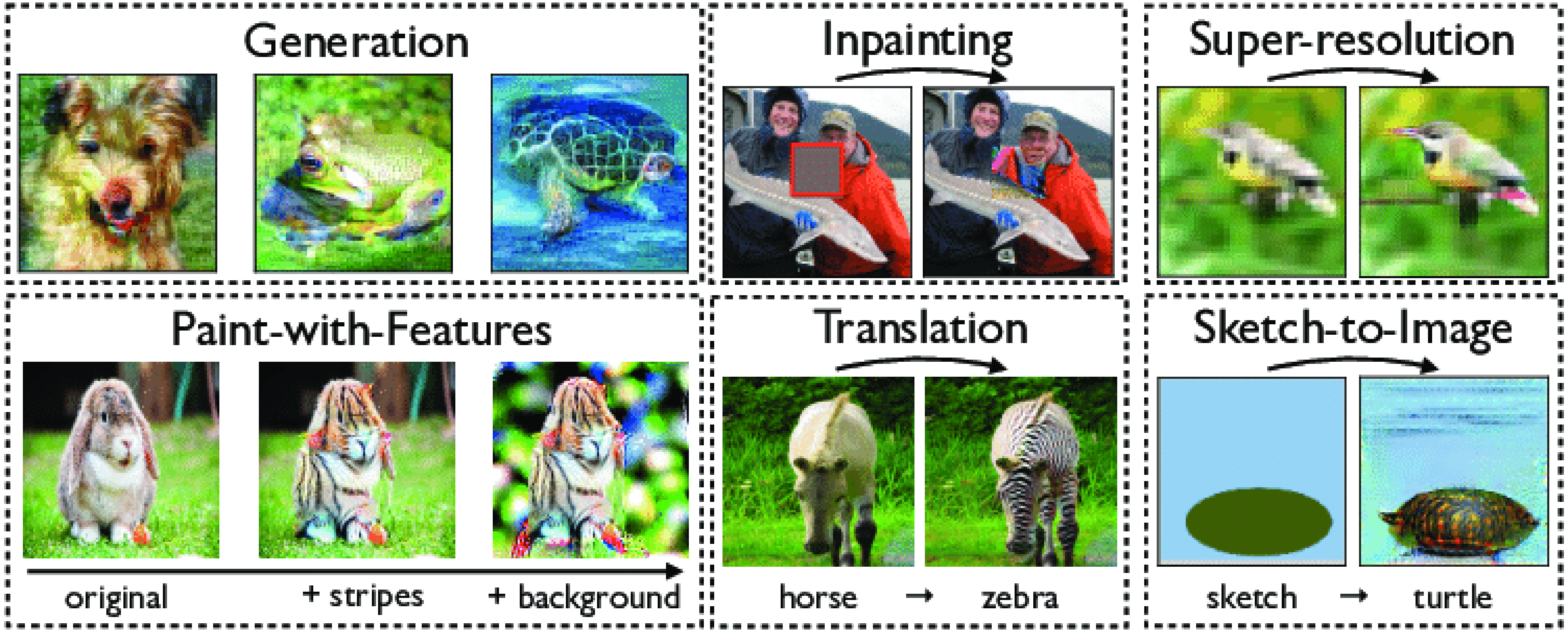
- **Simple Problems:** If the problem can be solved with basic rules or algorithms, ML may be overkill.
- **Limited Data:** When there isn't enough data to train a reliable model.
- **Interpretability Required:** When clear and understandable results are more important than complex predictions.
- **Resource Constraints:** If there are limitations in computational power, time, or expertise.

ML IN COMPUTER VISION

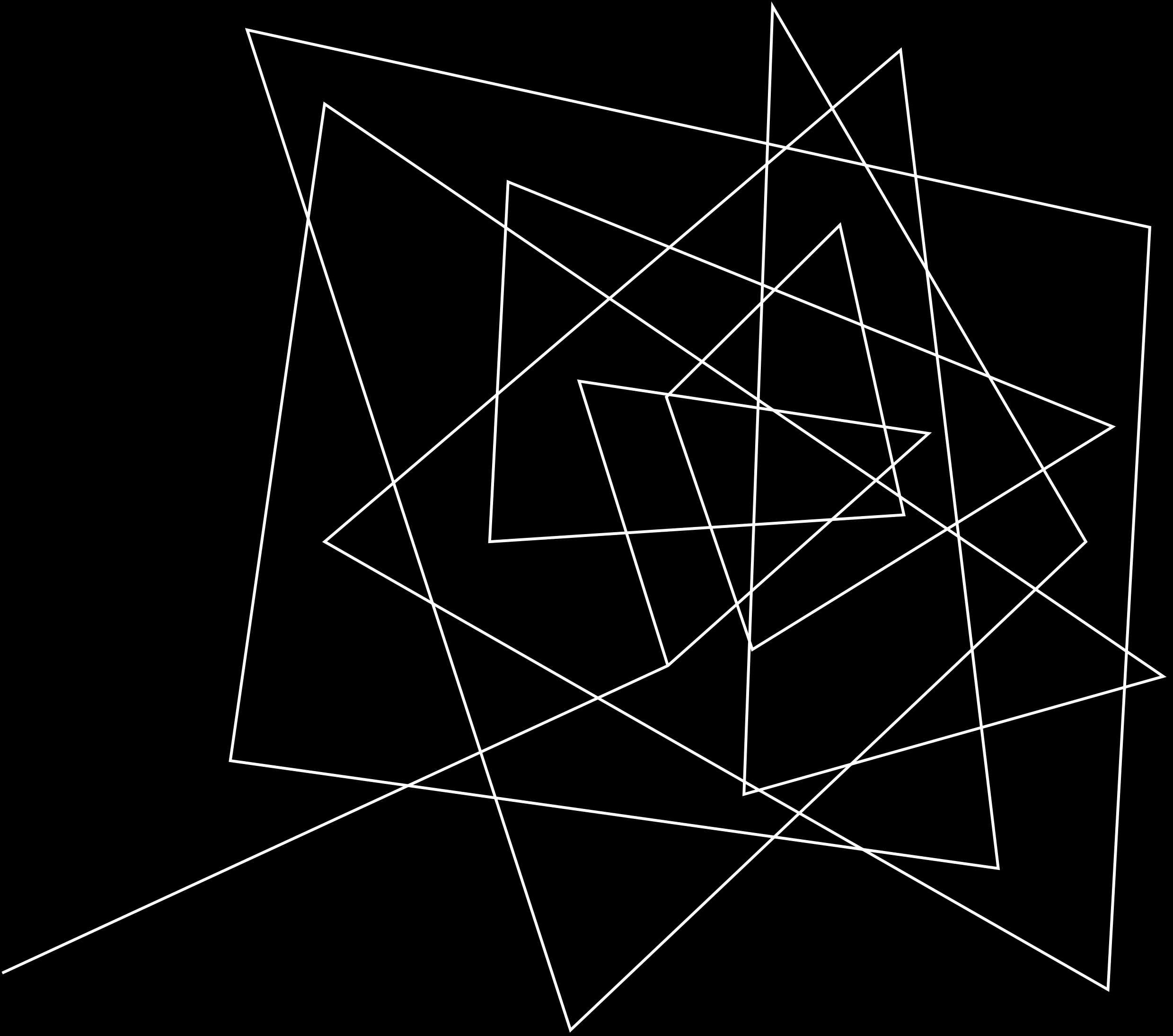


Source: https://www.tagxdata.com/

ML IN COMPUTER VISION



Source: S. Santurkar, A. Ilyas, D. Tsipras, L. Engstrom, B. Tran, and A. Madry, “Image synthesis with a single (robust) classifier,” in Advances in Neural Information Processing Systems 33 (NeurIPS), Dec. 2019, pp. 1262–1273



ML NOTATIONS & TERMINOLOGIES

ML NOTATIONS & TERMINOLOGIES

- **Data:** The foundation of machine learning. Algorithms learn from large datasets to identify patterns and make decisions.
- **Training:** The **recursive process** of feeding data to a machine learning model so it can learn from it. The model adjusts its internal parameters to minimize errors in predictions. An individual instance from the training data is a *training example/sample*. The whole dataset used for training is the *training set*.
- **Test Set:** A separate subset of the dataset that is used to **evaluate** the performance of a trained machine-learning model. It contains data that the model has never seen before, allowing for an unbiased assessment of how well the model generalizes to new, unseen data.
- **Model/Hypothesis:** A mathematical representation of a real-world process that a machine learning algorithm learns during training. The model is then used to make predictions or decisions based on new data.

ML NOTATIONS & TERMINOLOGIES

- **Features/Input:** Individual measurable properties or characteristics of the data used by the model to make predictions.
- **Labels/Output:** The output or **target variable** that the model aims to predict. In supervised learning, the model learns by comparing its predictions to these labels. In this course, we will denote original labels by $\boldsymbol{t}$ and predicted labels by **y.**
- **Parameters/Weights:** Parameters (or weights) are the **internal variables of a machine learning model** that are learned from training data. They determine how the input data is transformed into an output and are adjusted during the training process to minimize the error between the predicted and actual outputs.
- **Learning Algorithm:** The method or process used to adjust the model's parameters based on the training data. It dictates how the model learns from data by minimizing the error through techniques like gradient descent or optimization methods, ultimately improving the model's performance on unseen data.
- **Overfitting**: A situation where a model performs well on the training data but poorly on new, unseen data because it has learned noise or irrelevant details in the training set.

ML NOTATIONS & TERMINOLOGIES

Data (Training Set + Test Set)

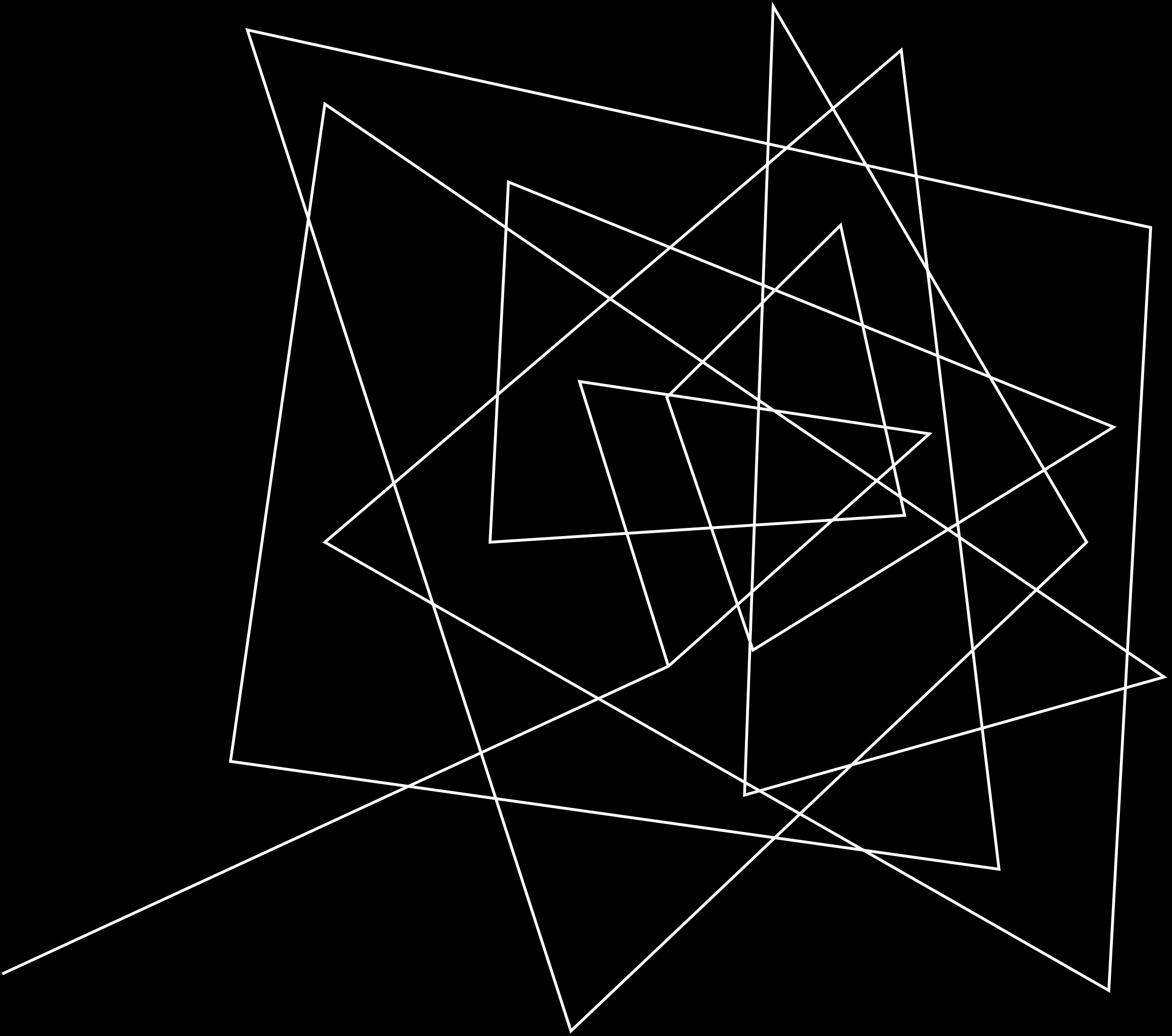
i^{th} Training Example/Sample $\mathbf{x}^{(i)}$, $t^{(i)}$ [here, i = 2]

Living Area (feet^2)	#Bedrooms	Price (X1000$)
2104	3	**400**
1600	3	**330**
2400	3	**369**
1416	2	**232**
3000	4	**516**
.	.	.
.	.	.

Input/Features, x

Target/Labels, t

Task: Predicting the price of the flat from the given Living area and # of Bedrooms.



TYPES OF ML

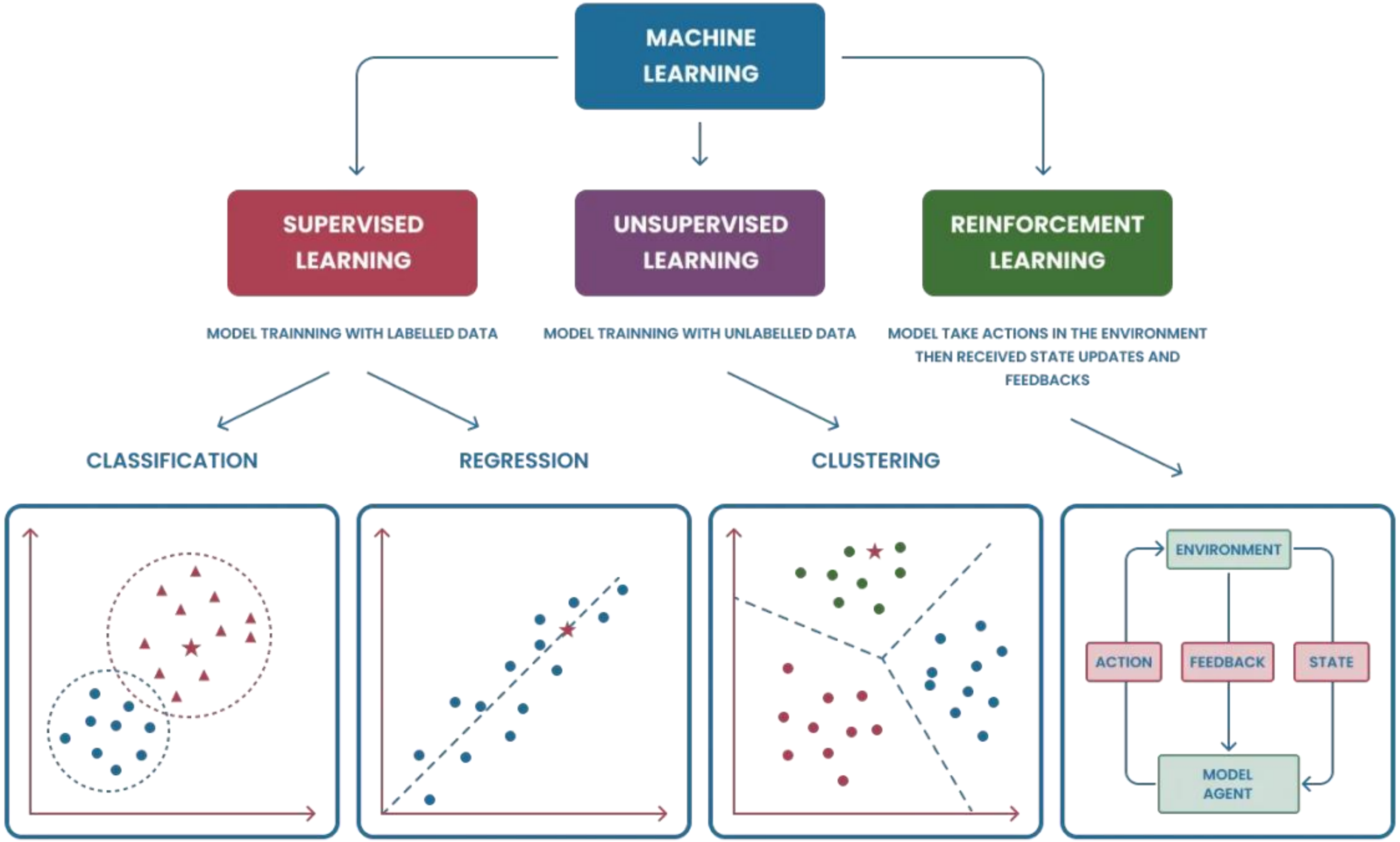
TYPES OF ML

There are **three major types** of Machine Learning:

1. **Supervised Learning**
2. **Unsupervised Learning**
3. **Reinforcement Learning**

Other types, such as **Semi-Supervised Learning**, **Self-Supervised Learning**, and **Transfer Learning**, are considered variations or hybrid approaches, but the three mentioned above are the core categories.

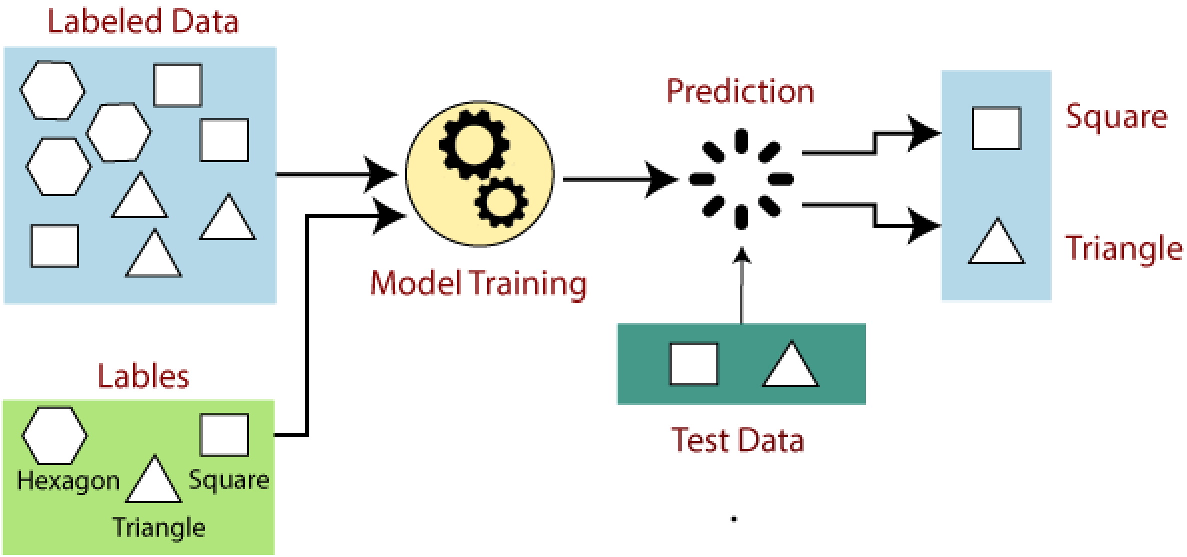
TYPES OF ML



Source: https://www.linkedin.com/pulse/types-machine-learning-models-algorithms-dr-rabi-prasad-padhy-w7nqc

SUPERVISED LEARNING

In **Supervised Learning**, the model is trained on a labeled dataset, where the input data is paired with the correct output. The goal is for the model to learn a mapping from inputs to outputs, enabling it to make predictions on new, unseen data. Examples include **classification** (e.g., spam detection) and **regression** (e.g., predicting house prices).

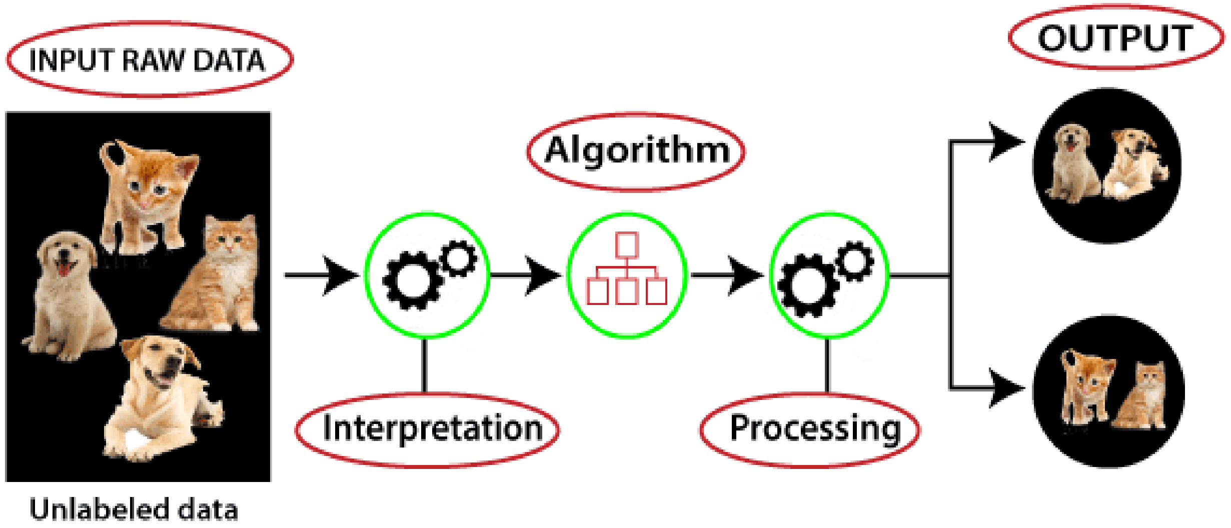


Source: https://www.javatpoint.com/supervised-machine-learning

Supervised Learning Process

UNSUPERVISED LEARNING

Unsupervised Learning involves training a model on **data without explicit labels**. The model tries to find patterns or structures in the data on its own. Common tasks include clustering (e.g., customer segmentation) and dimensionality reduction (e.g., principal component analysis).

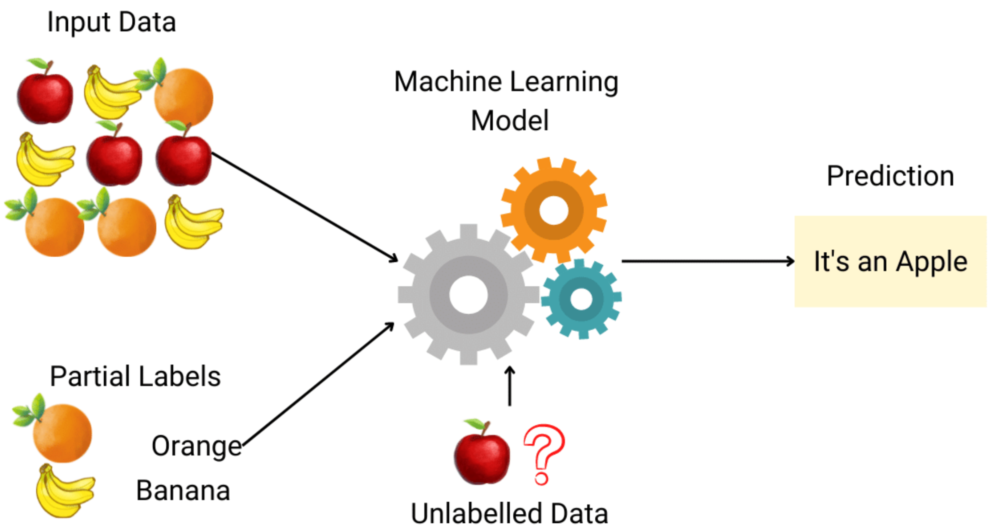


Source: https://www.javatpoint.com/unsupervised-machine-learning

Unsupervised Learning Process

SEMI-SUPERVISED LEARNING

Semi-Supervised learning is a hybrid of supervised and unsupervised learning, where the model is trained on a small amount of labeled data and a larger amount of unlabeled data. It's useful when labeling data is expensive or time-consuming, but large amounts of unlabeled data are available.

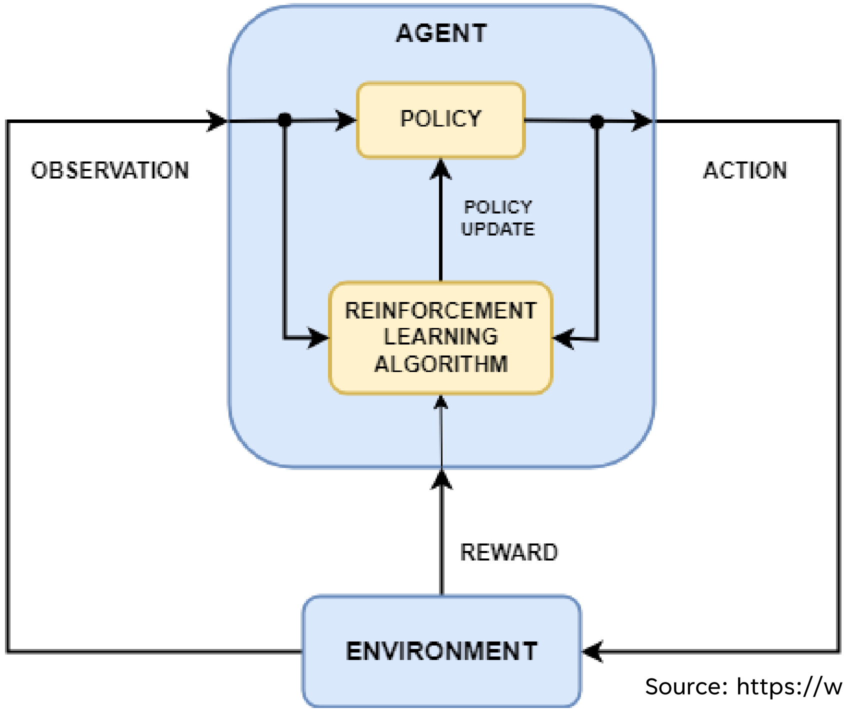


Source: https://www.enjoyalgorithms.com/blogs/supervised-unsupervised-and-semisupervised-learning

Semi-Supervised Learning Process

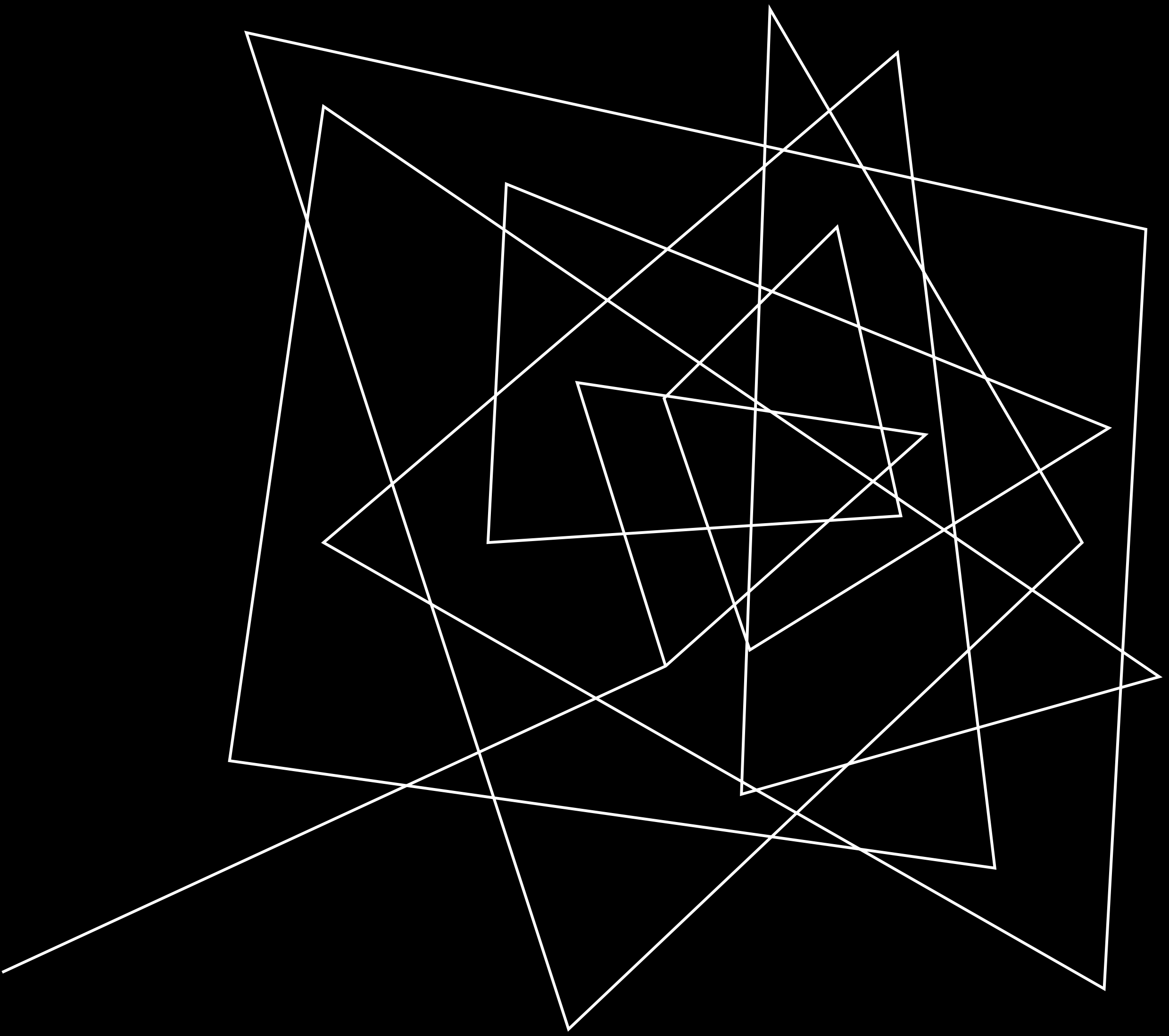
REINFORCEMENT LEARNING

In **Reinforcement Learning**, an agent learns to make decisions by interacting with an environment and receiving feedback in the form of rewards or penalties. The goal is to learn a strategy that maximizes cumulative rewards over time. It’s commonly used in robotics, gaming, and autonomous systems.



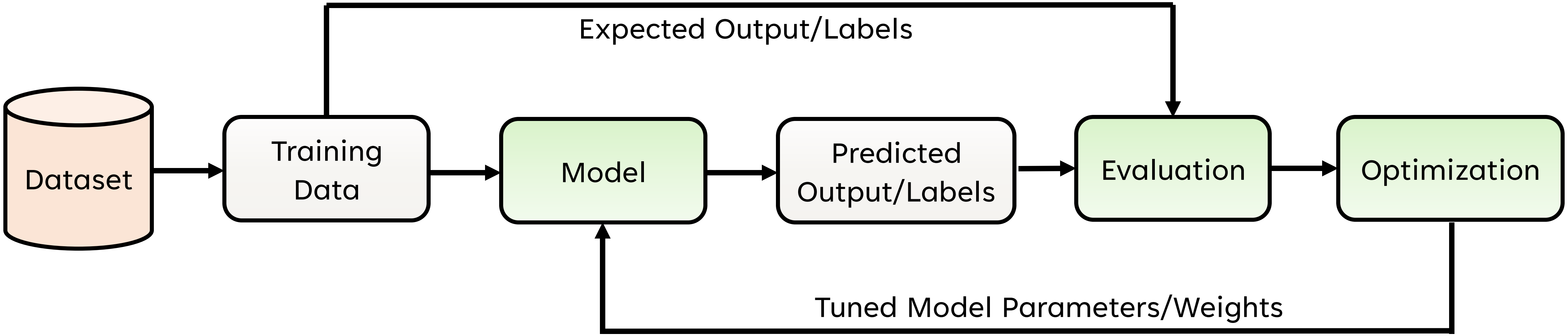
Source: https://www.mathworks.com/help/reinforcement-learning/

Reinforcement Learning Framework



OVERVIEW OF A GENERIC ML ALGORITHM

OVERVIEW OF A GENERIC ML ALGORITHM



The Diagram above represents **ML Workflow** in general. An **ML Algorithm** is made up of three elements related to three stages in this workflow. They are:

- **Model**: The mathematical representation of the problem (Structure of the algorithm/model).
- **Evaluation:** Loss Function (the model parameters are tuned to minimize this value).
- **Optimization:** Learning Algorithm (dictates the model tuning process).

OVERVIEW OF A GENERIC ML ALGORITHM

The three important elements of a Machine Learning algorithm are:

Model (Hypothesis): The mathematical representation of the problem, which is trained on data to make predictions or decisions. This includes the algorithm's structure, such as linear regression, decision trees, or neural networks.

Learning / Optimization Algorithm: The method used to adjust the parameters (weights) of the model based on the training data. Common learning algorithms include gradient descent, backpropagation, and various optimization techniques.

Cost Function (Objective Function): A function that measures the difference between the model's predictions and the actual outcomes. The learning algorithm aims to minimize this loss to improve the model's accuracy. At the individual sample level, it is known as **Loss Function**. Examples include mean squared error (MSE) for regression tasks and cross-entropy loss for classification tasks.

OVERVIEW OF A GENERIC ML ALGORITHM (MODEL)

Data (Training Set + Test Set)

Living Area (feet²)	#Bedrooms	Price (X1000$)
2104	3	**400**
1600	3	**330**
2400	3	**369**
1416	2	**232**
3000	4	**516**
.	.	.
.	.	.

$\mathbf{x}^{(2)}, t^{(2)}$

Input/Features, x **Target/Labels, *t***

$\boldsymbol{\theta}$ denotes the **Parameters(Weights)** of the Model

Here,

- Training set = $\{(\mathbf{x}^{(i)}, t^{(i)})$, for i = 1, 2, . . ,n},
  where, n = number of samples in the training set.
- $\mathbf{x}^{(i)} = \begin{bmatrix} x_1^{(i)} \\ x_2^{(i)} \end{bmatrix}$ As, #Features = 2 here.
- So, $\mathbf{x}^{(2)} = \begin{bmatrix} \mathbf{1600}, \\ \mathbf{3} \end{bmatrix}$ & $t^{(2)} = \mathbf{330}$ in this case.
- Let's Assume, $\mathbf{y = h_\Theta(x)}$

➢ Here, $\underbrace{\mathbf{h_\Theta : x \longrightarrow y} \text{ (mapping y from x)}}_{\text{Hypothesis / Model}}$

➢ Example: Assume, $\mathbf{y = 0.1x_1 + 50x_2 + 12}$

Here, $\mathbf{h_\Theta(x)} = \sum_{i=0}^{n} \Theta_i x_i$

Where, $\Theta_1 = 0.1$, $\Theta_2 = 50$, $\Theta_0 = 12$

OVERVIEW OF A GENERIC ML ALGORITHM (COST FUNCTION)

Data (Training Set + Test Set)

Living Area (feet²)	#Bedrooms	Price (X1000$)
2104	3	**400**
1600	3	**330**
2400	3	**369**
1416	2	**232**
3000	4	**516**
.	.	.
.	.	.

$\mathbf{x}^{(2)}, t^{(2)}$

Input/Features, x **Target/Labels, t**

We Assumed, $\mathbf{y = 0.1x_1 + 50x_2 + 12}$

Here, $\mathbf{h_\Theta(x)} = \sum_{i=0}^{n} \Theta_i x_i$

Where, $\Theta_1 = 0.1$, $\Theta_2 = 50$, $\Theta_0 = 12$

- In Vector form, $\mathbf{h_\Theta = \Theta^T x}$ or, $\mathbf{x^T\Theta}$
- How good is this model/hypothesis? We need an *Evaluation Metric*. This *evaluation metric* is the **Objective Function** or **Cost Function, $J(\theta)$**.
- The predicted label(price) should be as close as possible to the actual label. So, $\mathbf{y = h_\Theta(x)}$ can be compared to t.
- The *objective* of the ML algorithm will be to minimize the difference (loss) between $\mathbf{h_\Theta(x)}$ & t. Hence, **Minimizing the value of $J(\theta)$.**

Example: Let, $J(\theta) = \frac{1}{n}\sum_{i=1}^{n} J_i(\theta)$ Where, $J_i(\theta) = [t^{(i)} - h_\theta(\mathrm{x}^{(i)})]^2$. This is known as MSE (Mean Squared Error).

OVERVIEW OF A GENERIC ML ALGORITHM (OPTIMIZING ALGORITHM)

Data (Training Set + Test Set)

Living Area (feet^2)	#Bedrooms	Price (X1000$)
2104	3	**400**
1600	3	**330**
2400	3	**369**
1416	2	**232**
3000	4	**516**
.	.	.
.	.	.

$\mathbf{x}^{(2)}, t^{(2)}$

Input/Features, x **Target/Labels, t**

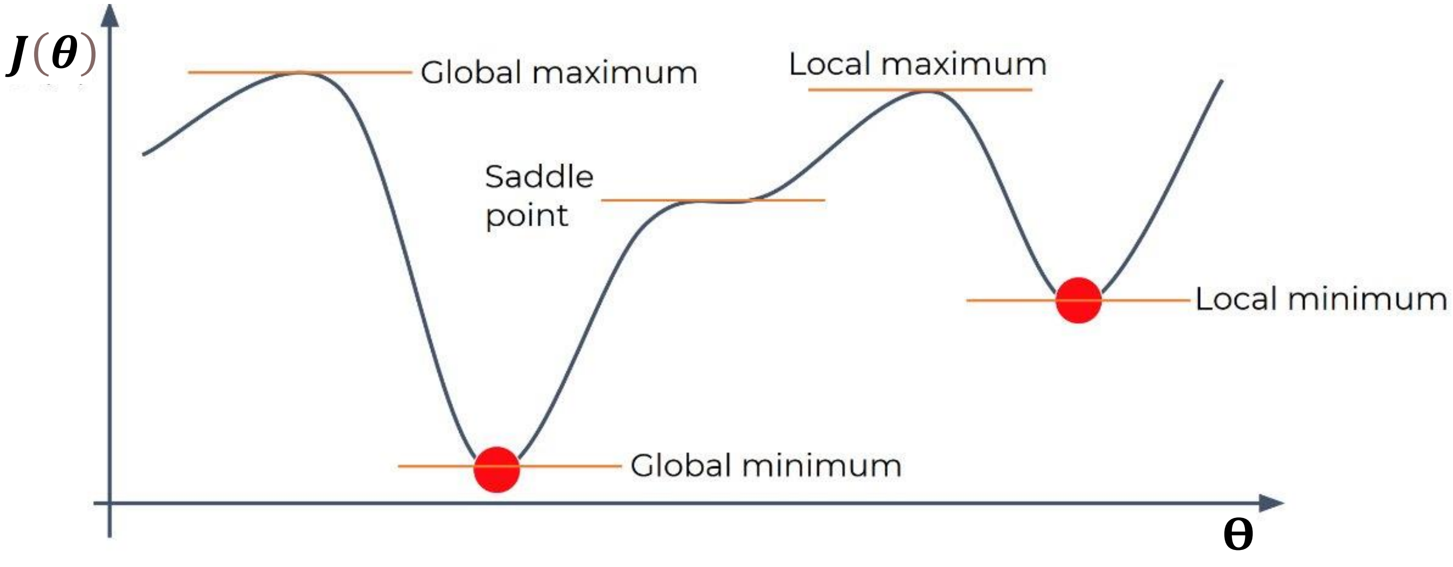
- The *objective* of the ML algorithm will be to minimize the difference (loss) between $\mathbf{h_\theta(x)}$ & $\boldsymbol{t}$. Hence, **Minimizing the value of $J(\theta)$.**

- Let, after *evaluation*, some value of $J(\theta)$ was observed. Now, we need some method to keep reducing this value of the Cost function. That's where the **Learning/Optimizing Algorithm** comes in.

- The **Optimizing Algorithm** typically tunes the parameters(weights). So, the purpose of this algorithm is to change $\boldsymbol{\theta = [\theta_0, \theta_1, \ldots, \theta_n]}$ in such a way so that $J(\theta)$ is reduced.

- The basic idea is to calculate the *gradient (or slope)* of the loss function with respect to each parameter (weight), and then **update the parameters in the direction that reduces the loss.**

OVERVIEW OF A GENERIC ML ALGORITHM (OPTIMIZING ALGORITHM)



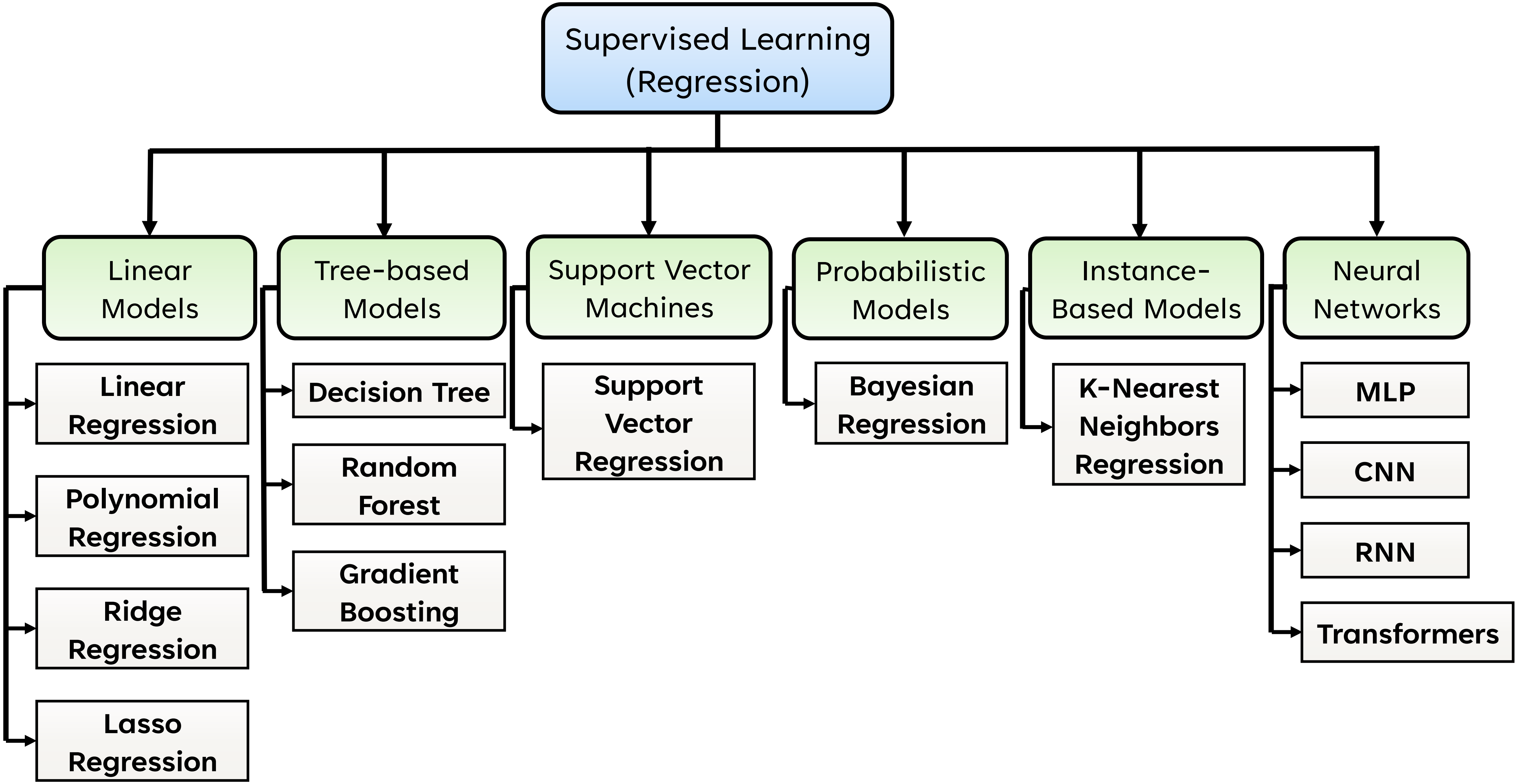
- In the graph above, the x-axis represents the combination(simplified) of the parameters **θ = [θ_0, θ_1, . . . , θ_n]** and the y-axis represents the value of the cost function, $J(\theta)$.

- Basically, the *Optimization Algorithm* aims to find the global minima or at least get as close as possible to it, so that the cost is minimal. **To find the combination of the parameters that results in that minimal value, calculating the gradients(derivatives) can be crucial, as they contain information on which direction should the parameters be changed to reach the minima.**

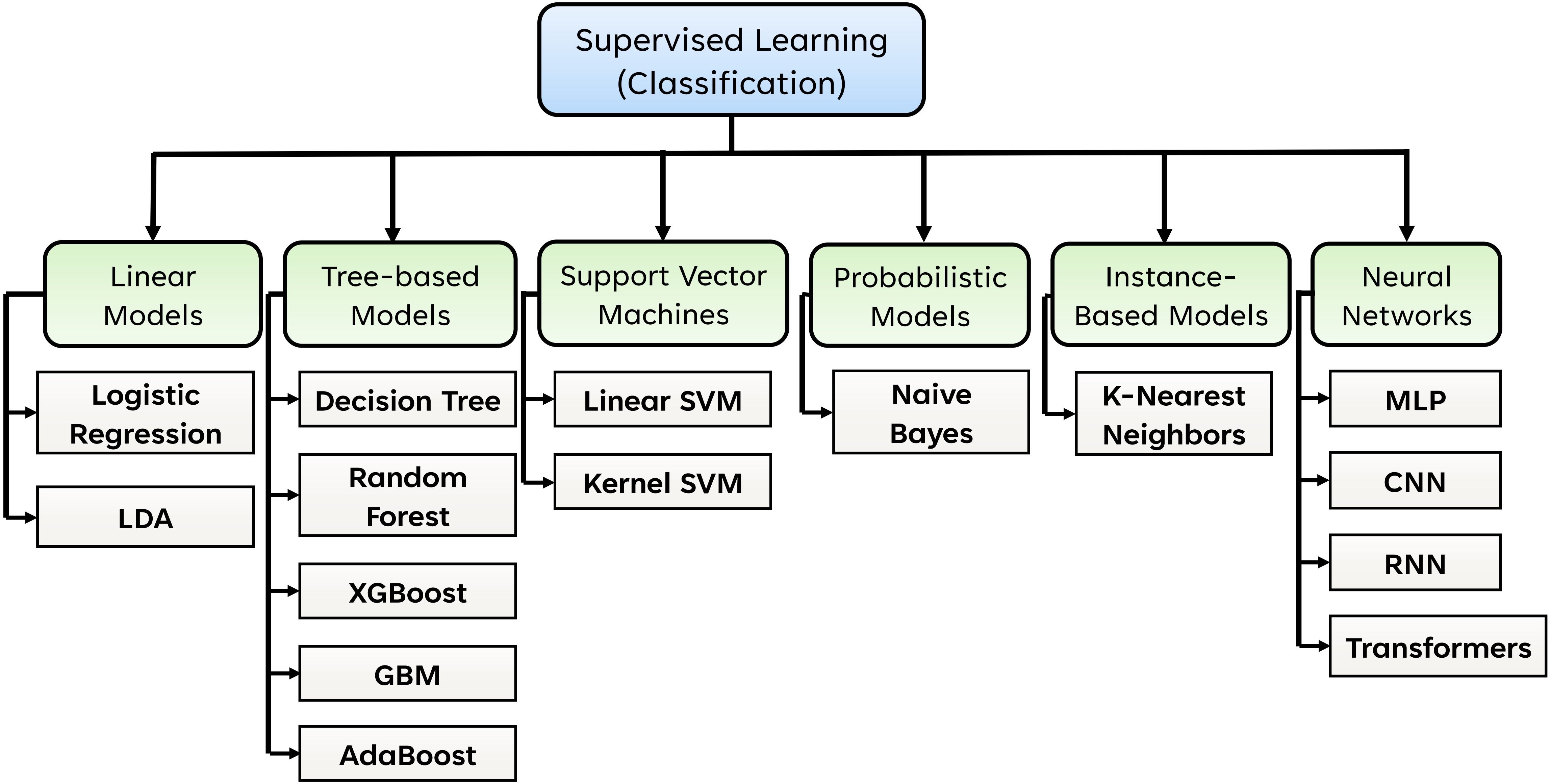
- Many Optimization Algorithms utilize the derivatives due to this fact, in order to find the minima.

POPULAR
ML ALGORITHMS

SUPERVISED LEARNING (REGRESSION)



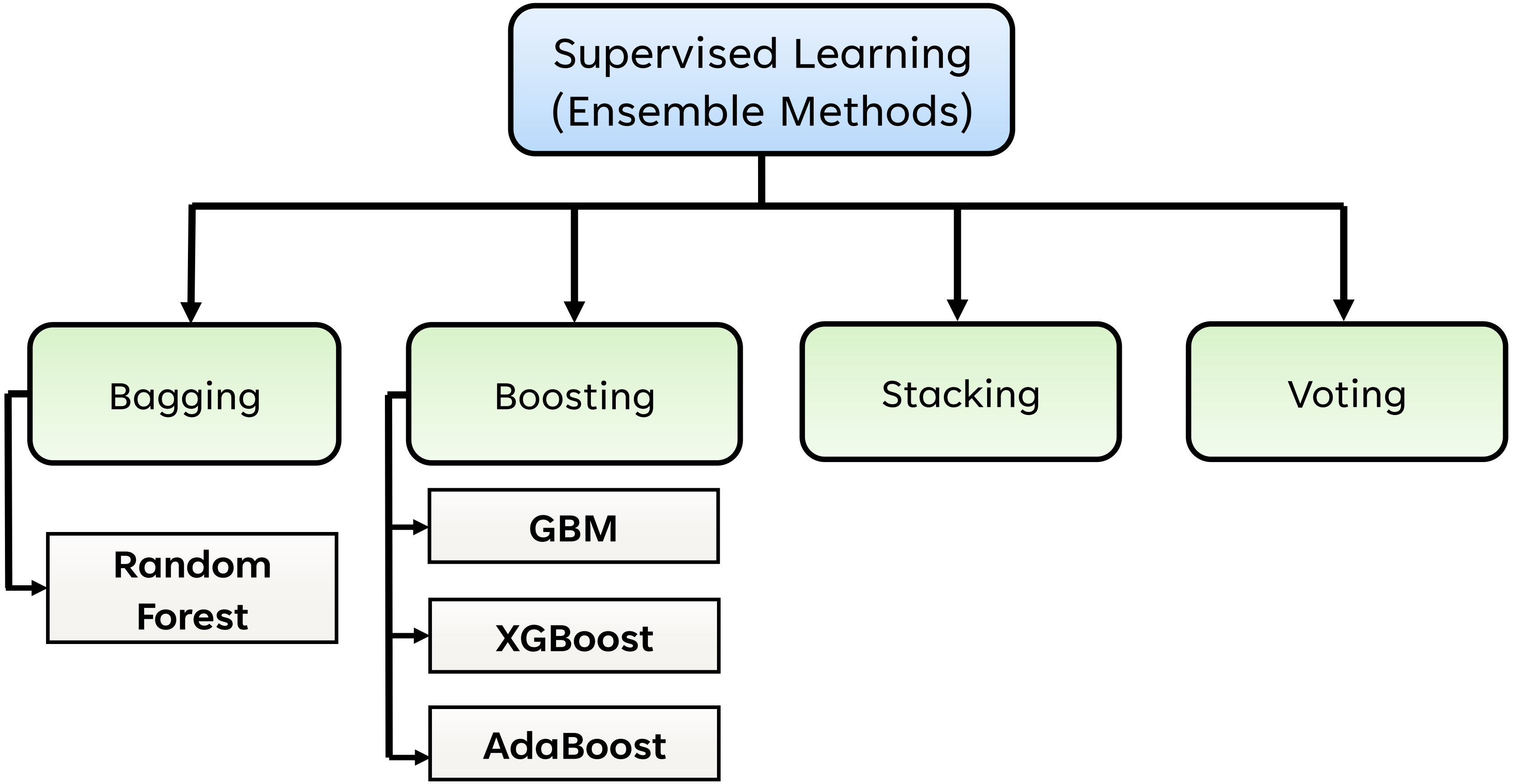
SUPERVISED LEARNING (CLASSIFICATION)



Supervised Learning (Classification)

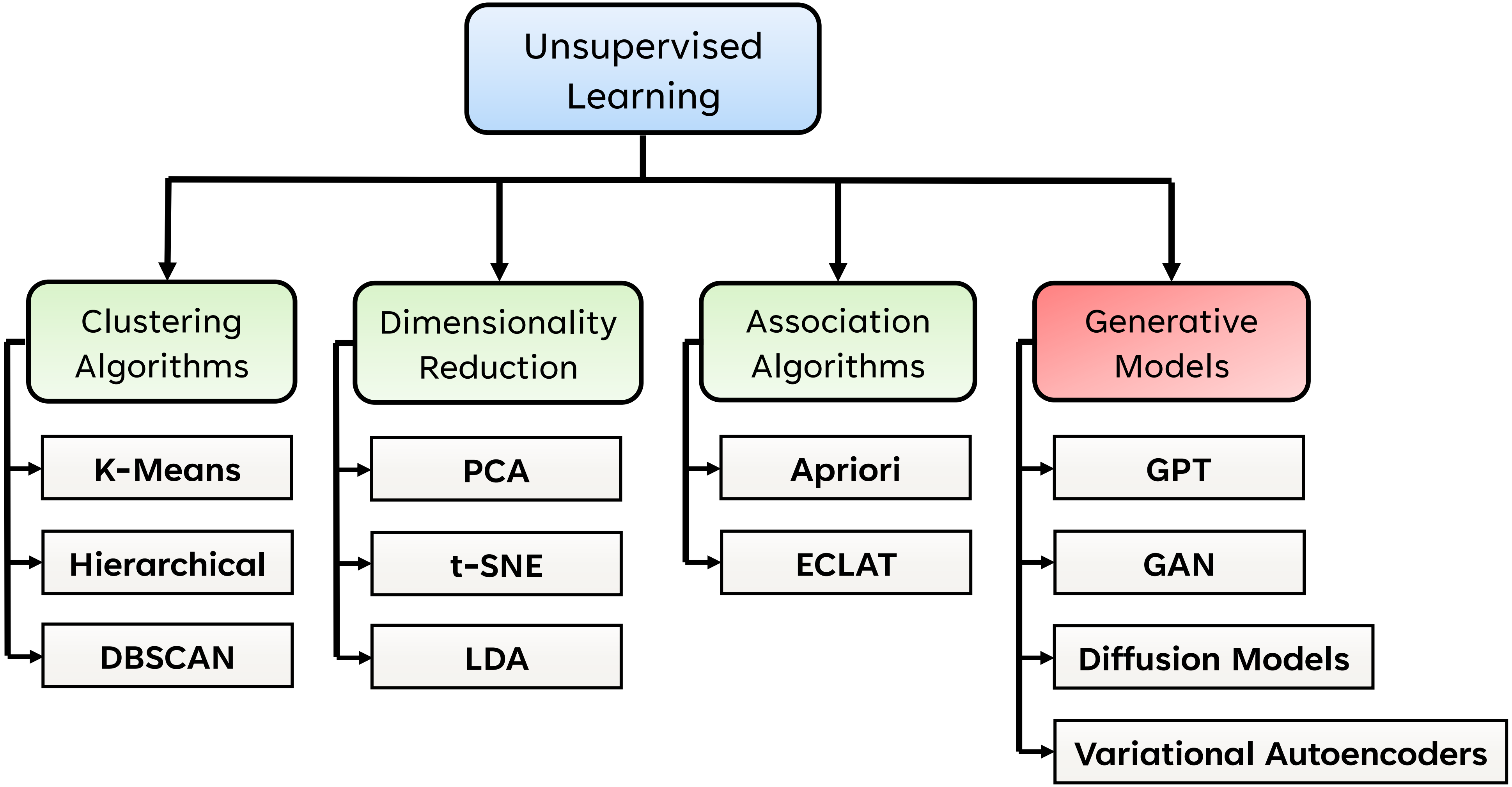
- Linear Models
  - Logistic Regression
  - LDA
- Tree-based Models
  - Decision Tree
  - Random Forest
  - XGBoost
  - GBM
  - AdaBoost
- Support Vector Machines
  - Linear SVM
  - Kernel SVM
- Probabilistic Models
  - Naive Bayes
- Instance-Based Models
  - K-Nearest Neighbors
- Neural Networks
  - MLP
  - CNN
  - RNN
  - Transformers

SUPERVISED LEARNING (ENSEMBLE METHODS)



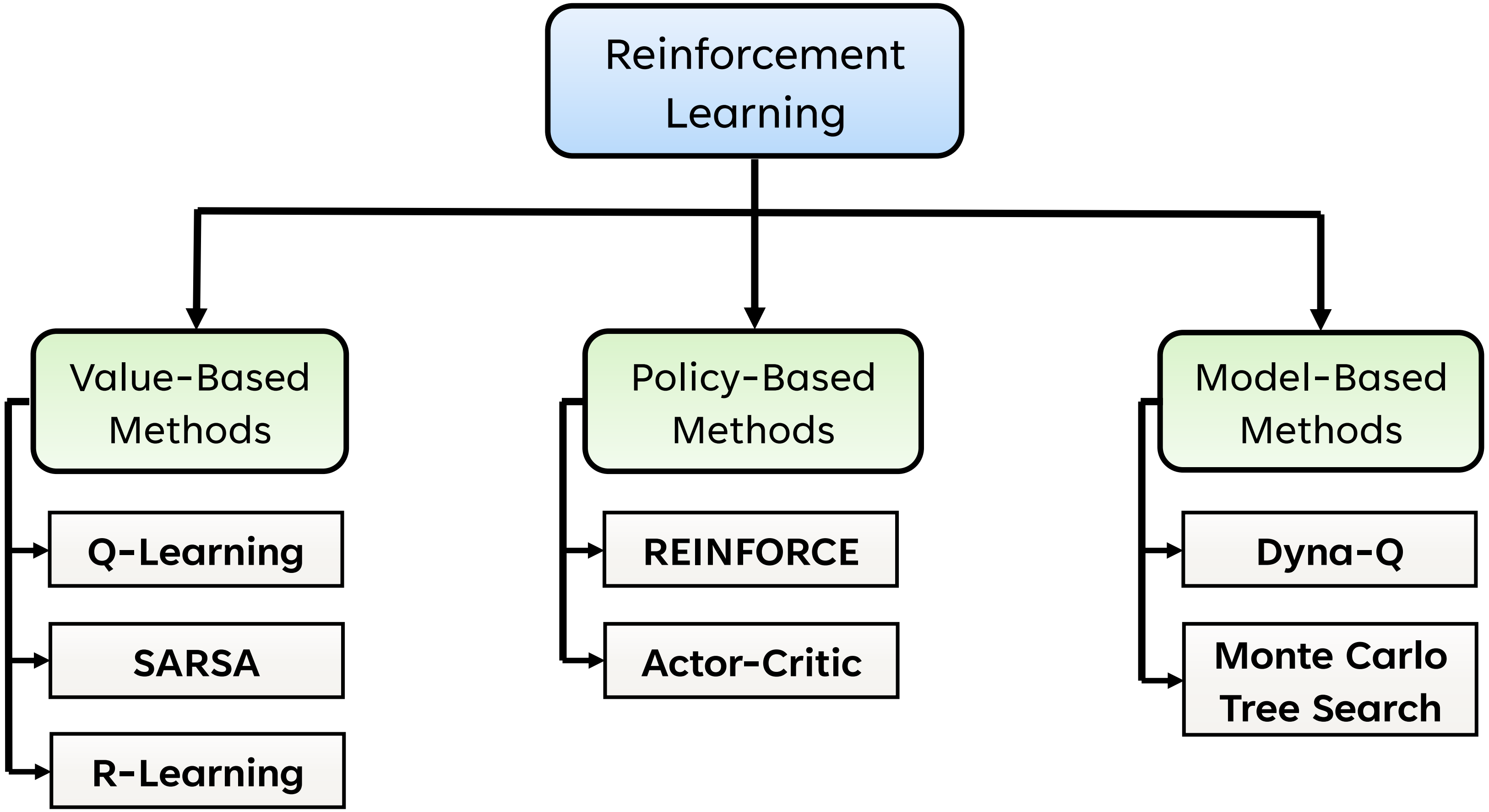
Ensemble: Ensemble methods are techniques in machine learning that combine multiple models (often referred to as "weak learners") to create a stronger, more robust model. The idea is that by aggregating the predictions of several models, the ensemble can achieve better performance.

UNSUPERVISED LEARNING

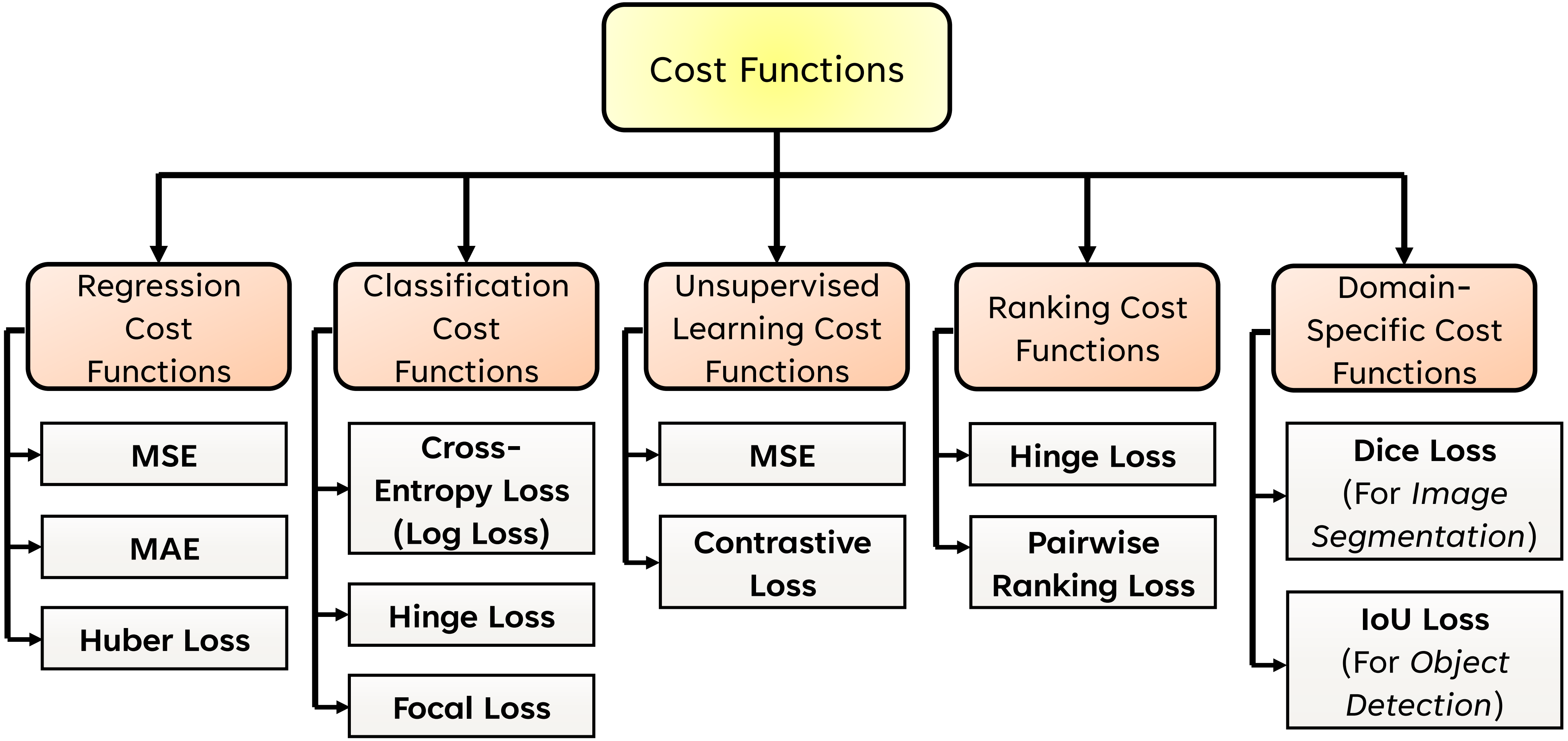


Generative Models are broadly categorized under *Unsupervised Learning*; however, many generative models incorporate *Supervised* elements within their training processes.

REINFORCEMENT LEARNING



COST FUNCTIONS


Cost Functions

Regression Cost Functions
- MSE
- MAE
- Huber Loss

Classification Cost Functions
- Cross-Entropy Loss (Log Loss)
- Hinge Loss
- Focal Loss

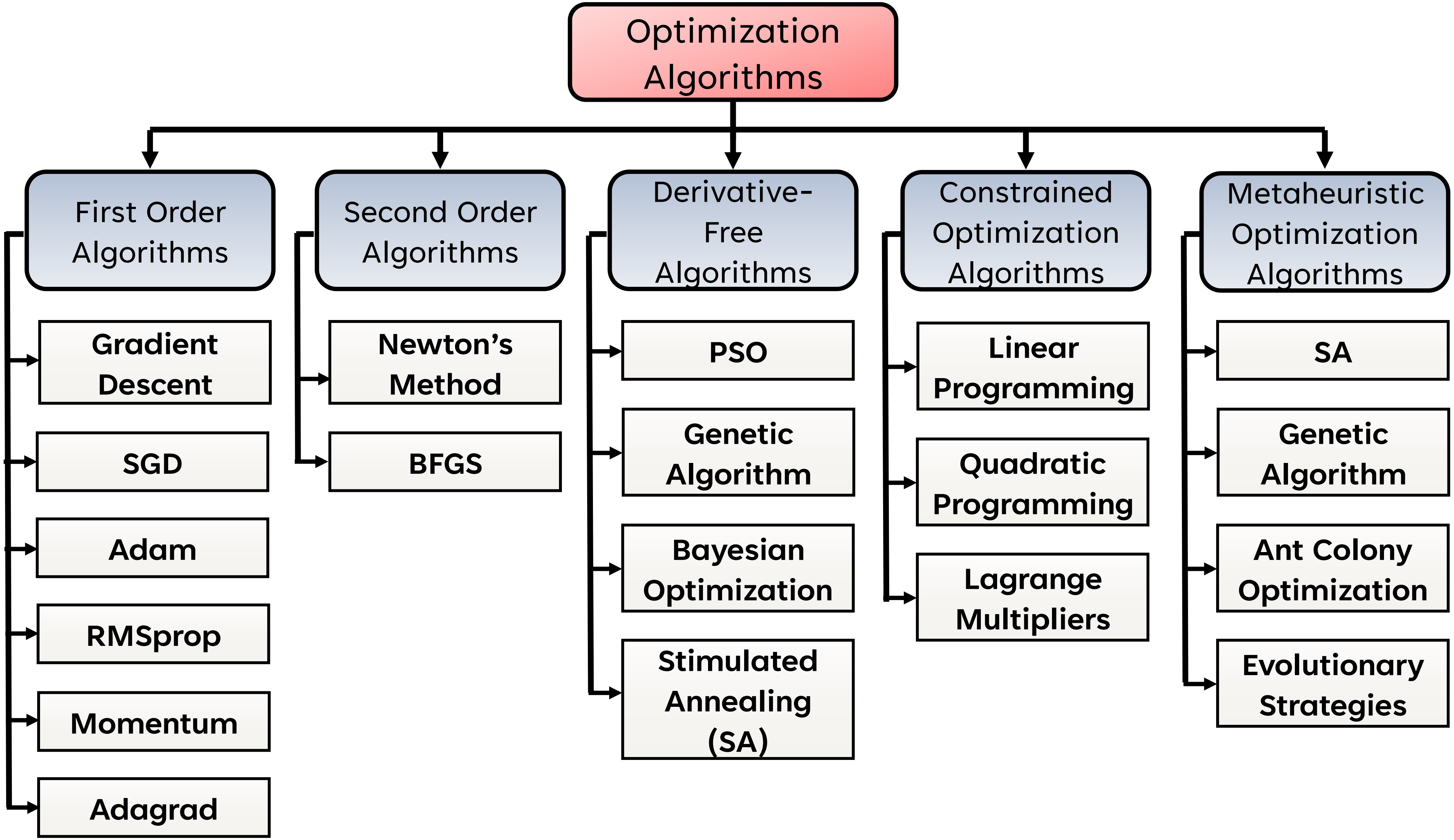
Unsupervised Learning Cost Functions
- MSE
- Contrastive Loss

Ranking Cost Functions
- Hinge Loss
- Pairwise Ranking Loss

Domain-Specific Cost Functions
- Dice Loss (For *Image Segmentation*)
- IoU Loss (For *Object Detection*)

Regularization Terms: Not standalone cost functions but they’re often added to various cost functions to address overfitting. Example: L1 Regularization (Lasso), L2 Regularization (Ridge).

OPTIMIZATION ALGORITHMS



CHALLENGES WITH MACHINE LEARNING

CHALLENGES WITH ML

- **Data Quality and Quantity**: High-quality and large quantities of data are essential for training effective ML models. Collecting, cleaning, and processing this data can be time-consuming and expensive. Also, If the data is biased or unrepresentative, the model may make incorrect predictions or fail to generalize to new data.
- **Model Interpretability:** Many ML models, particularly deep learning models, operate as *"black boxes"*, making it difficult to understand how they make decisions, and which features are significant. This lack of interpretability can be problematic in areas like healthcare or finance, where understanding the reasoning behind decisions is crucial.
- **Computational Resources:** ML models, especially deep learning models, often require significant computational power and memory, which can be costly and time-consuming.
- **Generalization:** Often ML Models may perform well on training data but fail to generalize to unseen data, leading to poor performance in real-world applications. Ensuring that the model performs well across different datasets and environments can be challenging.

CHALLENGES WITH ML

- **Security and Privacy:** ML systems can be vulnerable to adversarial attacks, where small, carefully crafted changes to input data can cause the model to make incorrect predictions. Additionally, using sensitive data in ML models raises privacy concerns.
- **Regulation and Compliance:** In industries like finance, healthcare, and law, ML models must comply with regulatory requirements, which can be difficult given the challenges with interpretability and bias.
- **Ethical Concerns and Bias**: ML models can unintentionally perpetuate or amplify societal biases present in the data. Addressing these biases and ensuring fairness in ML applications is a major challenge.
- **Keeping Up with Rapid Advances:** The field of ML is rapidly evolving, with new algorithms, techniques, and tools being developed constantly. Staying up-to-date with these advances and integrating them into existing workflows can be challenging.

WHEN TO USE MACHINE LEARNING

IS MACHINE LEARNING THE ANSWER TO EVERYTHING?

- *Machine Learning* has become a buzzword in recent years, with its potential to revolutionize industries and solve complex problems. However, this hype has led to a trend where people are eager to apply ML to every problem, often without considering whether it's the most appropriate solution.
- In many cases, simpler and more interpretable methods, such as rule-based systems or statistical analysis, might be more suitable. These traditional methods can be **faster to implement, easier to understand,** and **more cost-effective**, especially when dealing with well-defined problems that don't require the complexity of ML.
- Applying ML without a clear understanding of the problem domain can result in solutions that are not only inefficient but also potentially harmful if the model makes incorrect predictions.
- We should carefully evaluate whether ML is the right approach for a given problem, considering the **availability of data, the complexity of the problem,** and the **potential benefits compared to simpler methods.**

WHEN & WHEN NOT TO USE MACHINE LEARNING

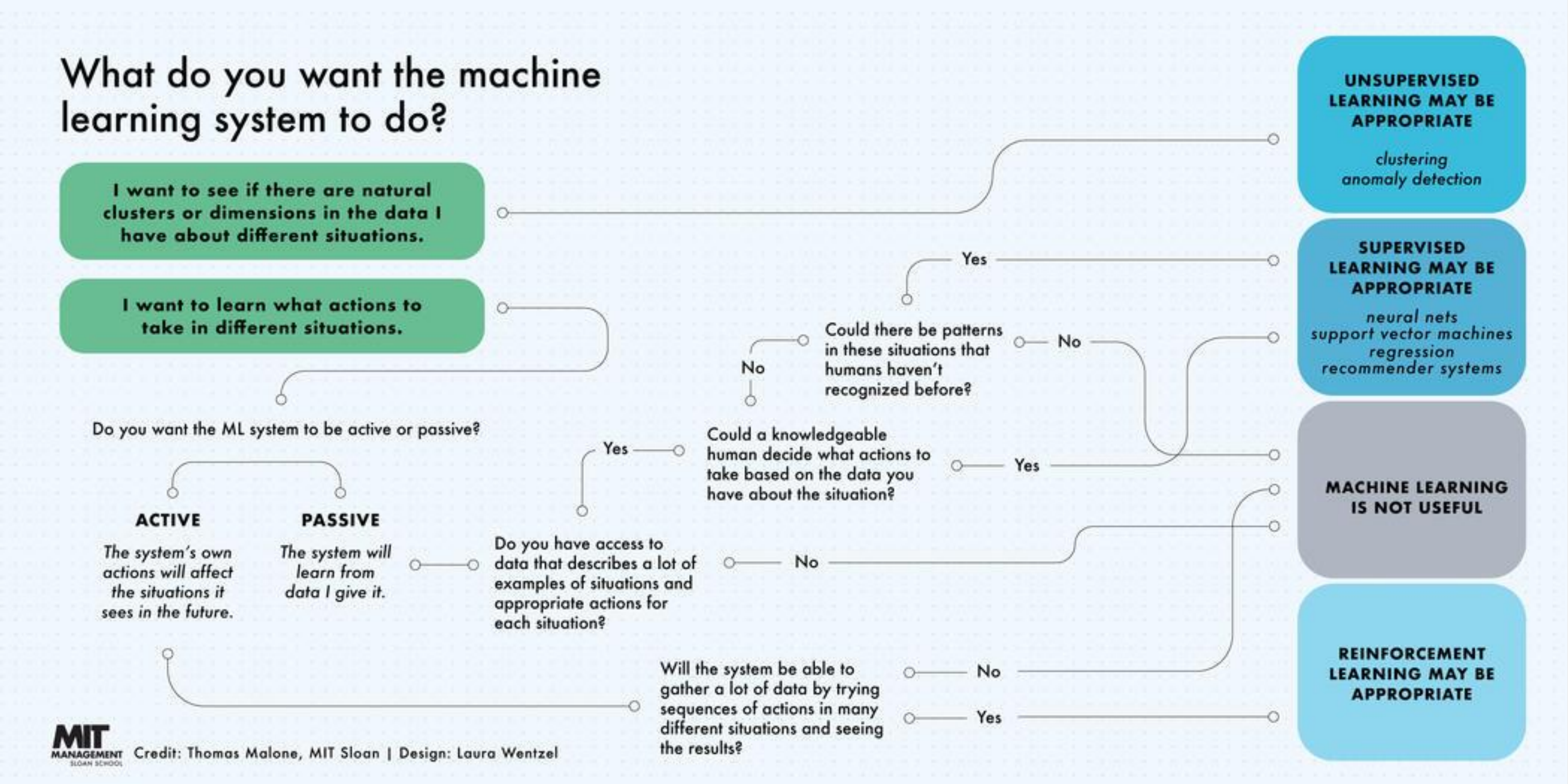
When to use ML

- When it is difficult to define all the rules explicitly.
- When the problem involves Unstructured data like images, text, etc.
- When the problem involves detecting complex patterns, trends, or relationships in large datasets.
- When automation is required & the system needs to adapt over time as more data becomes available or as conditions change.

When not to use ML

- When the problem is straightforward, can be solved using rule-based logic.
- When computational resources are limited or the cost of implementing ML is too high compared to the benefits.
- When the available data is inadequate to train a reliable ML model.
- When clear and interpretable results are more important than accurate predictions. This applies particularly in case of Deep Learning.

WHEN TO USE MACHINE LEARNING



What do you want the machine learning system to do?

I want to see if there are natural clusters or dimensions in the data I have about different situations.

I want to learn what actions to take in different situations.

Do you want the ML system to be active or passive?

ACTIVE

The system's own actions will affect the situations it sees in the future.

PASSIVE

The system will learn from data I give it.

Do you have access to data that describes a lot of examples of situations and appropriate actions for each situation?

Yes

No

Could a knowledgeable human decide what actions to take based on the data you have about the situation?

Yes

No

Could there be patterns in these situations that humans haven't recognized before?

Yes

No

Will the system be able to gather a lot of data by trying sequences of actions in many different situations and seeing the results?

No

Yes

UNSUPERVISED LEARNING MAY BE APPROPRIATE

clustering
anomaly detection

SUPERVISED LEARNING MAY BE APPROPRIATE

neural nets
support vector machines
regression
recommender systems

MACHINE LEARNING IS NOT USEFUL

REINFORCEMENT LEARNING MAY BE APPROPRIATE

MIT MANAGEMENT SLOAN SCHOOL Credit: Thomas Malone, MIT Sloan | Design: Laura Wentzel

MACHINE LEARNING ETHICS

WHY ETHICAL CONCERNS REVOLVING ML?

Machine learning is quickly becoming a critical tool for both businesses and governments. That’s why it is now more important that the technology is used ethically and responsibly. While it has the potential to solve some of society’s most pressing issues, it also raises new **ethical issues** that must be addressed.

Ethical concerns regarding Machine Learning arise from several critical issues related to how these technologies are developed, deployed, and used. **Bias, Transparency,** and **Privacy** are critical ethical concerns in Machine Learning, each impacting how fair, accountable, and respectful of individual rights these systems are.

ML ETHICS: BIAS

ML models can inherit and amplify **biases present in the training data**. It is a no brainer that Machine learning algorithms can only be as good as the data on which they are trained. If the data is skewed in any way, the resulting algorithm/model will be skewed as well. Biases in machine learning models can have far-reaching consequences, from reinforcing harmful stereotypes to escalating existing social inequalities.

Example: A good example is when a facial recognition algorithm only sees data about white men, it might not be able to recognize people of other races or genders correctly. This is owing to the machine learning model being trained on data that only represents the white race.

Biased machine learning is prevalent across several sectors. It has been discovered that most algorithms used to figure out how likely it is that a patient will get a long-term illness are biased against people of color. This has gone as far as situations in the criminal justice system where people of color are hurt more than white people by the predictive algorithms. There are also evidences that algorithmic hiring processes are unfair to women and people with disabilities.

ML ETHICS:
TRANSPARENCY & ACCOUNTABILITY

Transparency & **Accountability** are crucial elements of ethical machine learning. For individuals to really understand how decisions are being made, it is crucial that the inner workings of machine learning algorithms are transparent. In simple terms, in order to hold people accountable for any unfavorable results, there must also be accountability for the results of these algorithms.

Example: A good example is some facial recognition algorithms, having a reputation for being very hard to understand, which makes it hard for users to understand how the algorithm came to its conclusion. Additionally, some predictive policing algorithms have come under fire for being incredibly opaque, making it difficult for the general public to comprehend how police decide who to investigate or arrest.

These issues are going to be very vital in the future especially in sectors like medicine and law enforcement. If the decision-making process is not transparent, there will always be a lack of trust and reliability.

ML ETHICS: PRIVACY

ML systems often rely on large datasets that include personal and sensitive information. The collection, storage, and analysis of this data can lead to **privacy** violations, especially if data is used without proper consent or if it is exposed through security breaches.

Example: There are many examples of machine learning models that threaten privacy. Predictive analytics in social media platforms might use personal data to target users with specific content, raising concerns about how much of an individual's online behavior is being tracked and analyzed without their explicit consent. Some online advertising platforms use machine learning algorithms to track users across the web and collect information about their browsing history, search queries, and other online activities. This can lead to ads that are very relevant to the user, but at the expense of their privacy.

Another example would be the controversy surrounding large corporations collecting and selling private data of their users. It has been a major talking point regarding AI ethics recently.

MACHINE LEARNING: ADDRESSING ETHICAL ISSUES

There are some core principles regarding Machine Learning Ethics that are aimed at preventing potential harm and ensuring that the data created by these systems are used responsibly. Some of those principles are addressed in this section:

- **Fairness:** Ensuring that machine learning algorithms do not discriminate against individuals or groups based on characteristics such as race, gender, or age.
- **Transparency:** Providing clear and understandable explanations of how algorithms make decisions to foster accountability and trust.
- **Privacy:** Safeguarding individuals' personal information through data protection and ensuring it is not misused or exploited.
- **Accountability:** Holding developers and users of machine learning systems responsible for their actions and the negative outcomes sometimes generated by these systems.

CONCLUSION

It is obvious that machine learning has the power to significantly improve a wide range of facets of our lives, including healthcare, transportation, and education. But it’s clear that this technology comes with a huge ethical burden. As machine learning is used more and more in business, it’s important that companies put ethical concerns first.

In conclusion, machine learning offers both potential and challenges. It is our collective responsibility to guarantee that it is created and applied in a way that is morally acceptable and responsible. We can make sure this technology works for everyone by putting an emphasis on ethical issues in our work and promoting ethical ways to use machine learning.

REFERENCES

1. Stanford cs231n Course Notes (https://cs231n.github.io/)
2. CMU's Introduction to Machine Learning (10-601) Lectures (https://www.cs.cmu.edu/%7Etom/10701_sp11/lectures.shtml)
3. MIT OpenCourseWare: 6.867 Machine Learning (https://people.csail.mit.edu/dsontag/courses/ml16/)
4. University of Toronto - CSC411/2515: Machine Learning and Data Mining (https://www.cs.toronto.edu/~rgrosse/courses/csc311_f20/)
5. Applied ML course at Cornell and Cornell Tech (https://github.com/kuleshov/cornell-cs5785-2020-applied-ml/)
6. https://medium.com/@Samietex/the-ethics-of-machine-learning-what-you-need-to-know-8f827568c554
7. https://www.vationventures.com/research-article/machine-learning-ethics-understanding-bias-and-fairness

THANK YOU