Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

b-it Bonn-Aachen International Center for Information Technology

Introduction to Machine Learning

Foundation Course SS24

March 26, 2024

Vicky Prince Victor

What is Machine Learning

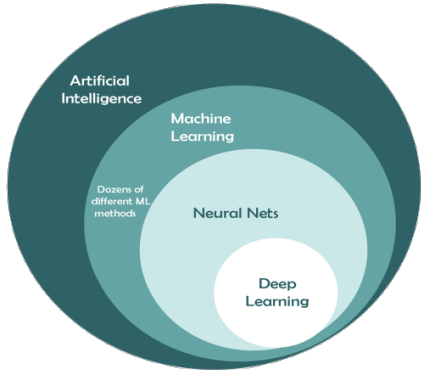

Figure 1: Relationship between Artificial Intelligence, Machine Learning, Neural Network and Deep Learning.

What is Machine Learning

“Learning is any process by which a system improves performance from experience.” - Herbert Simon

Definition by Tom Mitchell (1998): Machine Learning is the study of algorithms that

- improve their performance P
- at some task T
- with experience E.

A well-defined learning task is given by <P, T, E>

What is Machine Learning

Traditional Programming

Machine Learning

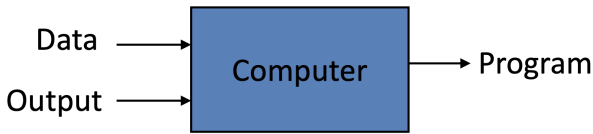

Figure 2: Enter Caption

What is Machine Learning

- Machine Learning is a subset of artificial intelligence (AI) that involves the development of algorithms and statistical models that enable computers to learn patterns from data without explicit programming.
- In essence, it allows machines to improve their performance on a task over time through experience.

Why Machine Learning

- Machine Learning is used to tackle complex problems for which traditional programming approaches may be impractical or inefficient.
- It excels in handling tasks like image and speech recognition, natural language processing, recommendation systems, and more.
- ML enables computers to make predictions, identify patterns, and learn from data, offering valuable insights and automation capabilities.

When to use Machine Learning

- Tasks involve large amounts of data.
- The task is too complex for explicit programming.
- The problem involves patterns or trends that are difficult to define with traditional programming.
- Continuous improvement or adaptation to changing conditions is necessary.

Applications of Machine Learning

- Computer Vision
- Information Retrieval
- Financial Prediction
- Medical Diagnosis
- Autonomous Driving
- etc ...

Basic Example of How Machine Learning Works

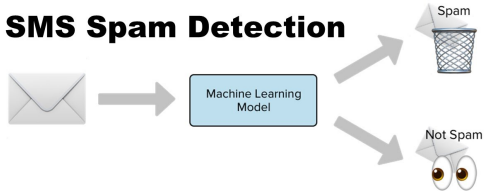

Figure 3: SMS Spam Detection using Machine Learning

Basic Example of How Machine Learning Works

- Consider a spam filter as a simple example.
- In traditional programming, rules might be explicitly defined to identify spam based on specific keywords.
- In contrast, a machine learning approach involves training a model on a dataset of emails, learning the characteristics of spam and non-spam emails.
- The model can then generalize this learning to identify and filter out new, unseen emails based on patterns it has learned.

Autonomous Driving

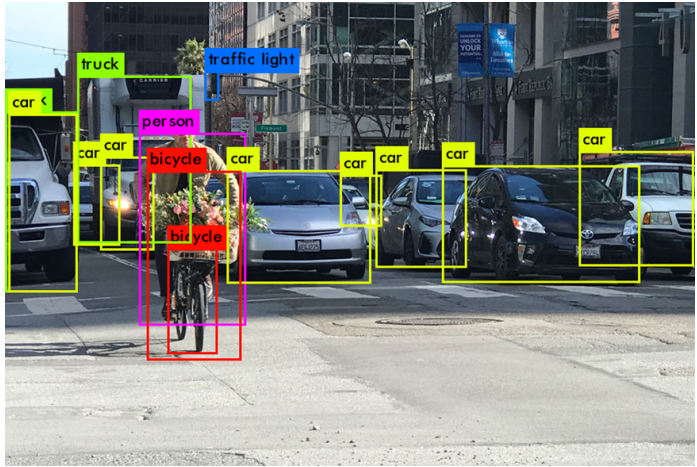

Figure 4: Data Labelling in Autonomous Driving

Steps Involved in a Machine Learning Project

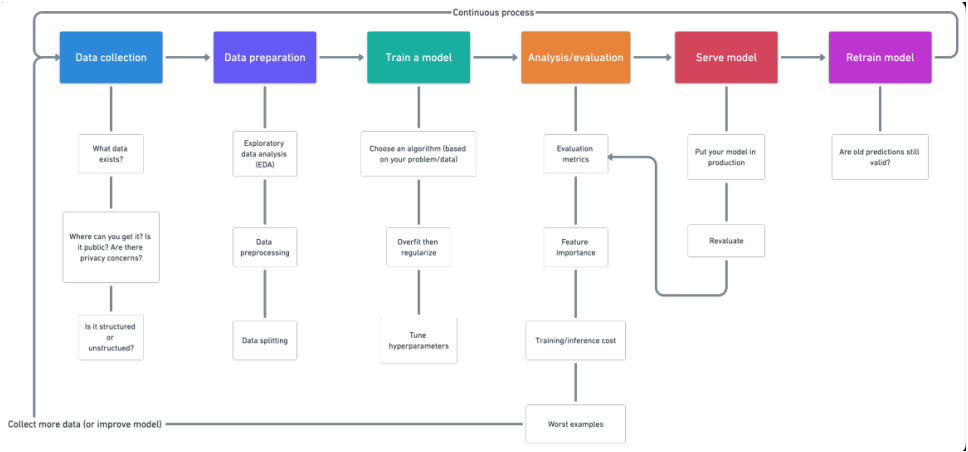

Figure 5: Steps involved in ML Projects [1]

Data Collection

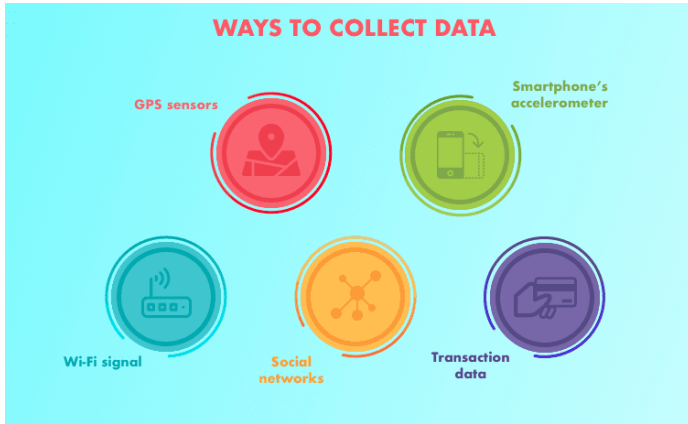

Figure 6: Data Collection [1]

Types of Data

- Numerical Data
- Categorical Data
- Time Series Data
- Text Data
- Image Data
- Audio Data
- Video Data
- Graph and Network Data

Data Preprocessing

- Data Cleaning
  - Handling Missing Values
  - Removing Outliers
- Data Integration - Merging data from multiple sources
- Data Normalization/Standardization
- Data Reduction
- Feature Engineering

Train a Model

- Split the dataset into training and test set.
- Select the algorithm depending on the type of dataset.
- Model learns on the training set.
- Test set is used for evaluating the performance of the model

Overfitting and Underfitting

- **Overfitting:**
  - Happens when a machine learning model learns the training data too well, including its noise and outliers, resulting in poor performance on new, unseen data.
  - High training accuracy but poor performance on new, unseen data.
- **Underfitting:**
  - Happens when a model is too simple to capture the underlying patterns in the training data, leading to poor performance on both the training and testing sets.
  - Low training accuracy and inadequate generalization to new data.

Machine Learning Paradigms

- Supervised (inductive) learning

  Given: training data + desired outputs (labels)

- Unsupervised learning

  Given: training data (without desired outputs)

- Reinforcement learning

  Rewards from sequence of actions

Supervised Learning

- Supervised learning is one of the primary categories of machine learning, where the model is trained on a labeled dataset.
- For each instance in the training dataset, the correct output or target value is known.
- The goal of supervised learning is to learn a mapping from inputs to outputs, enabling predictions or decisions for unseen data based on the learned relationship.

Supervised Learning

- Types of Problems
  1. Classification
  2. Regression

- Real World Applications
  - Image Recognition: Identifying objects, people, or actions in images.
  - Speech Recognition: Translating spoken language into text.
  - Fraud Detection: Identifying fraudulent activities in banking and financial systems.
  - Personalized Recommendations: Recommending products, movies, or music to users based on past preferences.

Classification

- Given (x_1, y_1), (x_2, y_2), ..., (x_n, y_n)
- Learn a function f(x) to predict y given x
- y is categorical

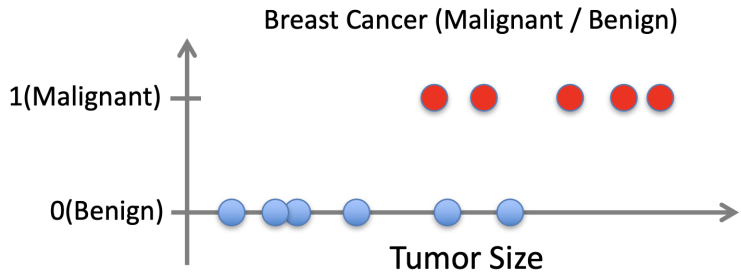

Figure 7: Example of Classification

Regression

- Given (x_1, y_1), (x_2, y_2), ..., (x_n, y_n)
- Learn a function f(x) to predict y given x
- y is real-valued

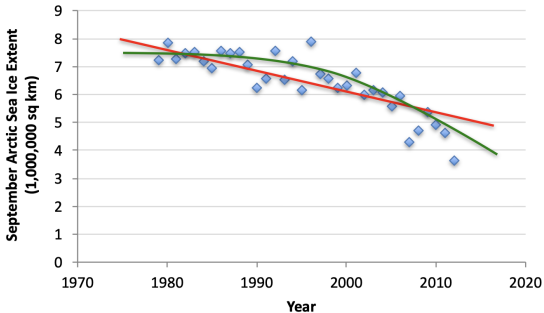

Figure 8: Example of Regression [4]

Unsupervised Learning

- Unsupervised learning is another fundamental category of machine learning, which involves working with data that does not have labeled responses.
- The goal of unsupervised learning is to model the underlying structure or distribution in the data in order to learn more about the data itself.
- Model tries to identify patterns and relationships directly from the input data.

Unsupervised Learning

- Types of Problems
  1. Clustering
  2. Dimensionality Reduction

- Real World Applications
  - Anomaly Detection: Identifying unusual data points in datasets, which is useful in fraud detection, network security, and fault detection.
  - Content Recommendation: Grouping similar content for recommendations in streaming services, e-commerce, and more.
  - Genomic Sequencing: Identifying patterns in DNA sequences without prior knowledge of what to look for.

Unsupervised Learning

- Given x_1, x_2, x_3, ..., x_n (without labels)
- Output hidden structure behind the data

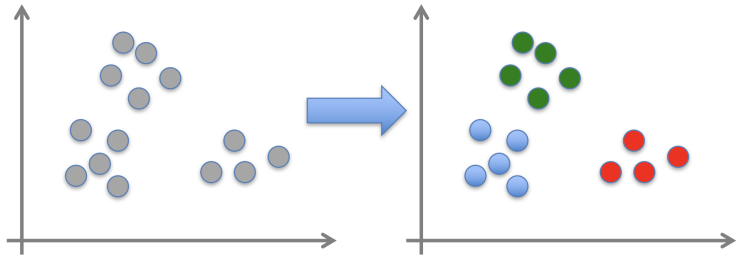

Figure 9: Clustering

Supervised and Unsupervised ML Algorithms

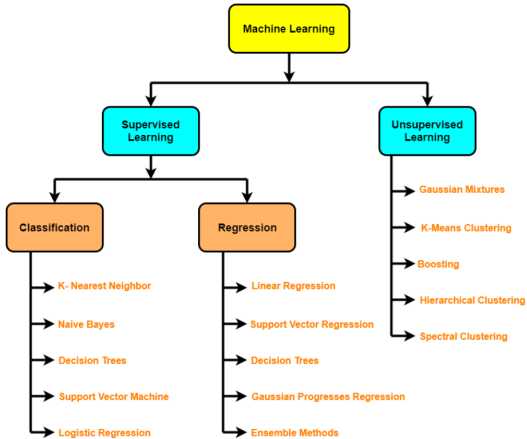

Figure 10: Supervised and Unsupervised ML Algorithms [1]

Reinforcement Learning

- Reinforcement Learning (RL) is a type of machine learning where an agent learns to make decisions by interacting with an environment.
- The agent receives feedback in the form of rewards or punishments, enabling it to learn optimal strategies over time

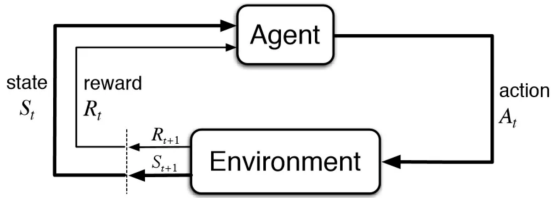

Figure 11: Reinforcement Learning [2]

Reinforcement Learning

- **State:** Represents the current situation or configuration of the environment.
- **Action:** The decision or move made by the agent in response to a given state.
- **Policy:** The strategy or set of rules guiding the agent's decision-making process.
- **Reward:** The immediate feedback the agent receives after taking an action, influencing its learning.
- **Return:** The cumulative reward obtained by the agent over a sequence of actions in an episode or trajectory.

Autonomous Driving

Figure 12: Reinforcement Learning in Autonomous Driving

Autonomous Driving

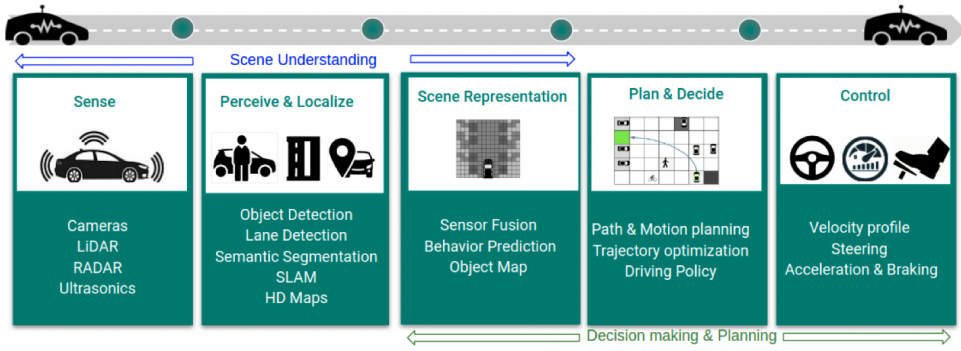

Figure 13: Modern autonomous driving systems pipeline [3]

References

[1] https://www.analyticsvidhya.com/blog/2021/04/steps-to-complete-a-machine-learning-project
[2] https://towardsdatascience.com/reinforcement-learning-101-e24b50e1d292
[3] https://arxiv.org/pdf/2002.00444.pdf
[4] Data from G. Witt. Journal of Statistics Education, Volume 21, Number 1 (2013)