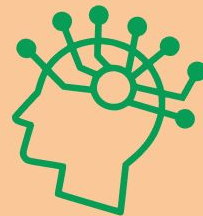




welcome to ML Study Jam

session #1

Google Developer Student Clubs
Simon Fraser University





Timeline

20
June

Session 1

Intro to Machine Learning

Data Processing

27
June

Session 2

Feature Engineering

Intro To Deep Learning

4
July

Session 3

ML Application

Computer Vision

11
July

Session 4

Practice Project





Intro to Machine Learning

ML Algorithms

Data Processing





Intro to Machine Learning

ML Algorithms

Data Processing





What is Machine Learning?



- Definition: Machine Learning is a subset of Artificial Intelligence that enables computers to learn and make predictions or decisions without being explicitly programmed
- ML enables systems to automatically analyze and extract patterns from data

Me: *uses machine learning*

Machine: *learns*

Me:



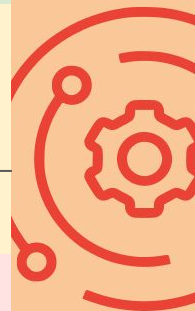


Core Concepts and Terminologies

- Features: Input variables or attributes used to make predictions.
- Labels: The output or target variable to predict.
- Training Data: Labeled data used to train the ML model.
- Testing Data: Unlabeled data used to evaluate the ML model's performance.
- Prediction: Making an output or decision based on input data.



Color (Feature)	Weight (Feature)	Fruit Name (Label)
Red	150	Apple
Yellow	120	Banana
Red	180	Apple
Green	130	Apple
Yellow	110	Banana
Green	150	Apple
Yellow	110	Banana
Yellow	90	Banana
Yellow	100	Banana
Red	190	Apple

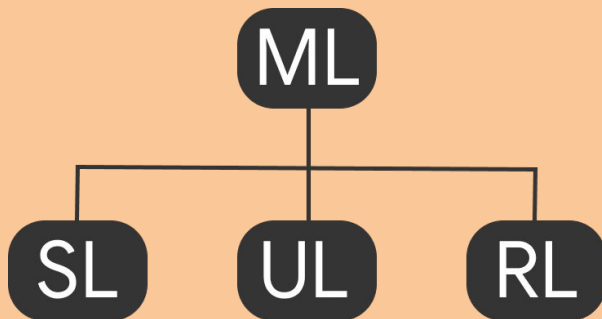




Types of Machine Learning

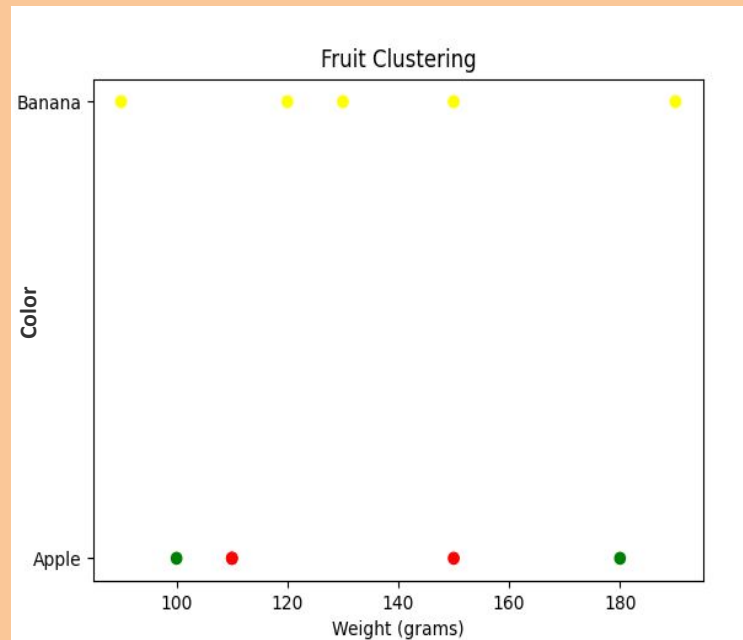
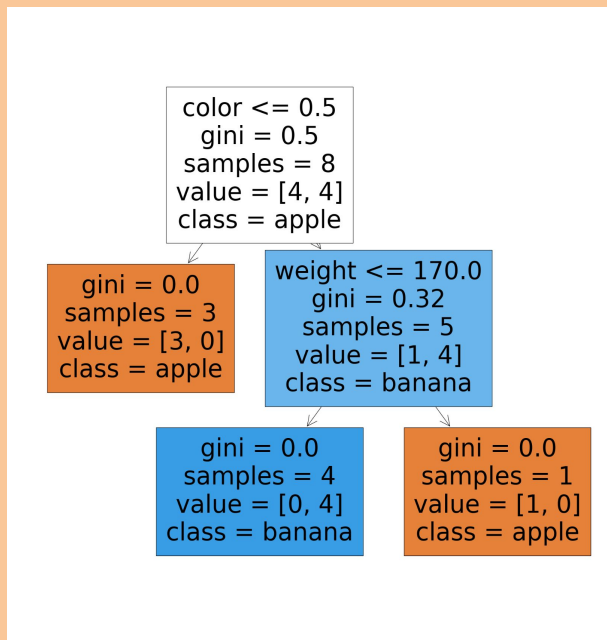


- Supervised Learning: Learning from labeled data with input-output pairs.
- Unsupervised Learning: Learning from unlabeled data to discover patterns or structures.
- Reinforcement Learning: Learning through trial and error interactions with an environment.





Supervised vs Unsupervised Learning





Intro to Machine Learning

ML Algorithms

Data Processing





Machine Learning Algorithms

- Regression: Predicting continuous numeric values.
- Classification: Assigning data to predefined categories or classes.
- Clustering: Grouping similar data points based on their characteristics.

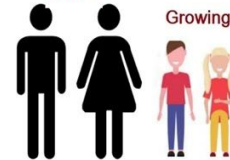
Classification vs. Regression vs. Clustering

170 CMS 168 CMS



Regression
(Shall we predict??)

Grown



Classification
(We shall classify!)

Cluster A



Clustering
(We shall cluster, i.e. group)



Popular Machine Learning Algorithms



- Linear Regression: Predicting a continuous value based on linear relationships.
- Logistic Regression: Classifying data into discrete categories using a logistic function.
- Decision Trees: Creating a tree-like model for classification or regression.
- K-means Clustering: Grouping data points into clusters based on similarity.





Intro to Machine Learning

ML Algorithms

Data Processing





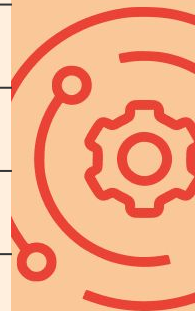
Data Processing

- Manipulation and transformation of raw data to make it suitable for analysis and modeling.
- Typical data Processing tasks:
 - Handling missing values
 - Dealing with outliers
 - Data normalization and scaling
- Crucial for improving the quality and performance of ML models.





Color (Feature)	Weight (Feature)	Fruit Name (Label)
Red	150	Apple
Yellow	120	Banana
Red		Apple
Green		Apple
20/June/2023		Banana
	150	Apple
Yellow	0.11 kg	Banana
1	90	Orange
Yellow	100	Banana
Red	0.42 lb	Apple





Data Processing Workflow

- Data loading and inspection.
- Data cleaning and preprocessing.
- Feature engineering and selection.
- Data transformation and normalization.
- Splitting data into training and testing sets.





Data Processing with Pandas

- Powerful library for data manipulation and analysis in Python.
- It provides data structures and functions for efficient data processing.
- Common data processing tasks, including data cleaning, filtering, and transformation.





Numerical Computing with NumPy



- Fundamental library for numerical computing in Python.
- Features and benefits:
 - N-dimensional array objects (ndarrays) for efficient storage and manipulation of large datasets.
 - Mathematical operations and functions for array-based computations.
 - Linear algebra, Fourier transforms, and random number generation capabilities.
- Provides a solid foundation for numerical operations in Machine Learning.





Exercise Time

<https://colab.research.google.com/drive/1aoZXlOyjb9KB1xa2sLXQIZ98O7yrKUw2?usp=sharing>





Kaggle Resources

- <https://www.kaggle.com/learn/intro-to-machine-learning>
- <https://www.kaggle.com/learn/pandas>





Thank you!

Google Developer Student Clubs
Simon Fraser University

