Regression & Classification

Presented By: Azeem Waqar
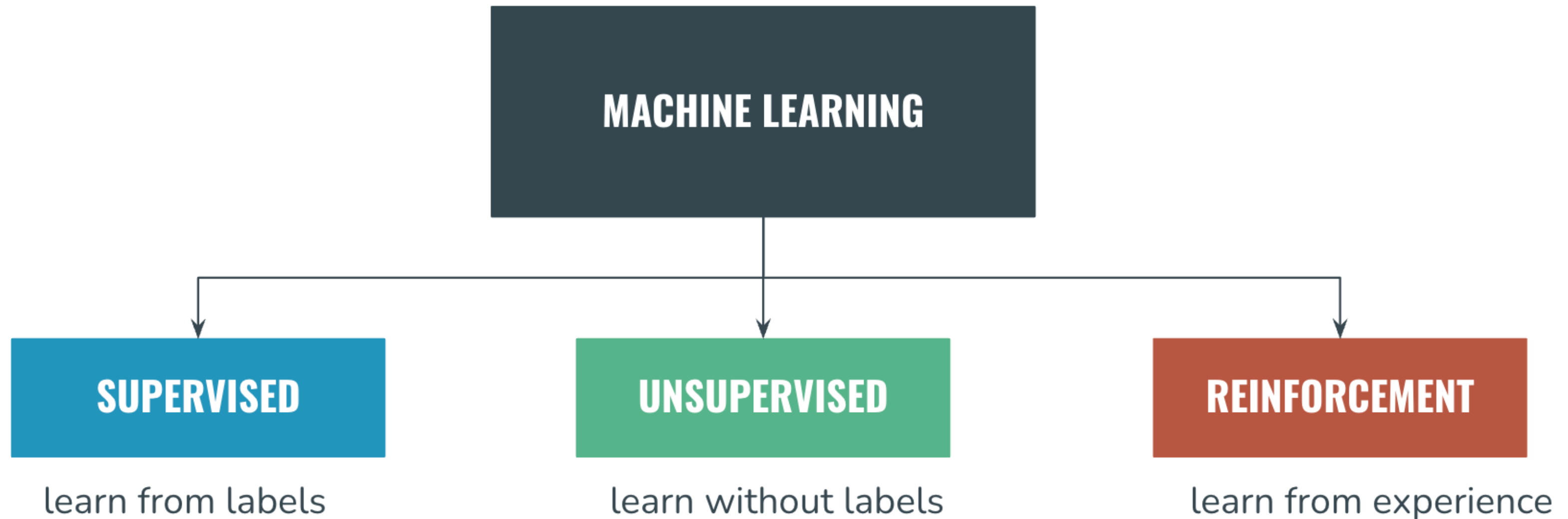
Agenda

- Data
- Value Types
- Regression
- Classification
- Categorical Encoding
- Train Test Split
- Accuracy Metrices

Types of Machine Learning

Let's talk about Data first!

Data....

- A combination of rows and columns.
- Every row represents an object.
- Every column represents a feature of an object.

Name	Age (years)	Weight (kg)	Height (cm)	BMI
Hamza	24	77	166	27.9
Asfan	19	68	170	21.7
Laiba	22	54	154	22.8

Types of attributes/columns

- Independent Features
- Dependent Feature

Find the Dependent Feature.

$$BMI = \frac{\text{weight (kg)}}{\text{height (m}^2)}$$

Out of 3 attributes (weight, height, BMI), which one is Dependent ????

Value Types

- Continuous
  - uncountable
  - between a range of 2 number
- Discrete
  - countable

Exercise

Price (Rs)	Category	Sold
280	Food item	0
320	Clothing	1
210	Phone Acessory	1

Regression
&
Classification

Regression

Regression is a type of supervised machine learning algorithm used for predicting continuous numeric values.

Examples of Regression

- House Price Prediction
- Stock Price Prediction
- Air Quality Index
- Customer Life Time Prediction

Regression Algorithms

- Linear Regression
- Polynomial Regression
- Decision Tree Regressor
- Random Forest Regressor

and many more...

Classification

Classification is a type of supervised machine learning algorithm used for predicting categorical or discrete class labels.

Examples of Classification

- Spam Email Detection
- Cancer Disease Prediction
- Credit Card Fraud Detection

Classification Algorithms

- K Nearest Neighbors
- Decision Tree
- Random Forest
- SVM

Encoding

Why encode?

- Computers do not understand strings/words
- Categorical values must be encoded to numerical values

Price (Rs)	Category	Sold
280	Food item	0
320	Clothing	1
210	Phone Acessory	1

Which column is categorical??

Price (Rs)	Category	Sold
280	Food item	0
320	Clothing	1
210	Phone Acessory	1

I am... Trust me

Category
Food item
Clothing
Phone Acessory

Category
Food item
Clothing
Phone Acessory

→ 0

Category
Food item
Clothing
Phone Acessory

Food item → 0

Clothing → 1

Category
Food item
Clothing
Phone Acessory

→ 0

→ 1

→ 2

Category
0
1
2

Encoding Techniques

- One hot encoding
- Label encoding
- Ordinal encoding

and many more...

Train Test Split

What is it?

- Train-test split is a technique used in machine learning to evaluate the performance of a model on unseen data.
- The dataset is divided in 2 parts, one is used for training other for training.

Name	Age (years)	Weight (kg)	Height (cm)	BMI
Hamza	24	77	166	27.9
Asfan	19	68	170	21.7
Haseeb	21	55	153	23.1
Hadi	22	67	168	24.2
Laiba	22	54	154	22.8

Name	Age (years)	Weight (kg)	Height (cm)	BMI
Hamza	24	77	166	27.9
Asfan	19	68	170	21.7
Haseeb	21	55	153	23.1

train data

test data

| Hadi | 22 | 67 | 168 | 24.2 |
| Laiba | 22 | 54 | 154 | 22.8 |

Name	Age (years)	Weight (kg)	Height (cm)	BMI
Hamza	24	77	166	27.9
Asfan	19	68	170	21.7
Haseeb	21	55	153	23.1
Hadi	22	67	168	24.2
Laiba	22	54	154	22.8

Evaluation Metrices

What are they?

Evaluation metrics, also known as performance metrics or scoring metrics, are measures used to assess the performance or effectiveness of a machine learning model or algorithm.

For regression

- mean squared error
- root mean squared error
- mean absolute squared error
- r squared error

For classification

- accuracy
- precision
- recall
- F1- Score

Python Libraries

- **Pandas**
  - Data manipulation and analysis library
  - Data cleaning & Data exploration

- **Scikit Learn**
  - Train test split
  - Machine learning algorithms
  - Accuracy metrices

let's summarize

Questions?

Thank you
for coming!

Special Thanks to Devhack!!