CSE 471: Machine Learning

Categorization of Machine Learning

Tajkia Nuri Ananna

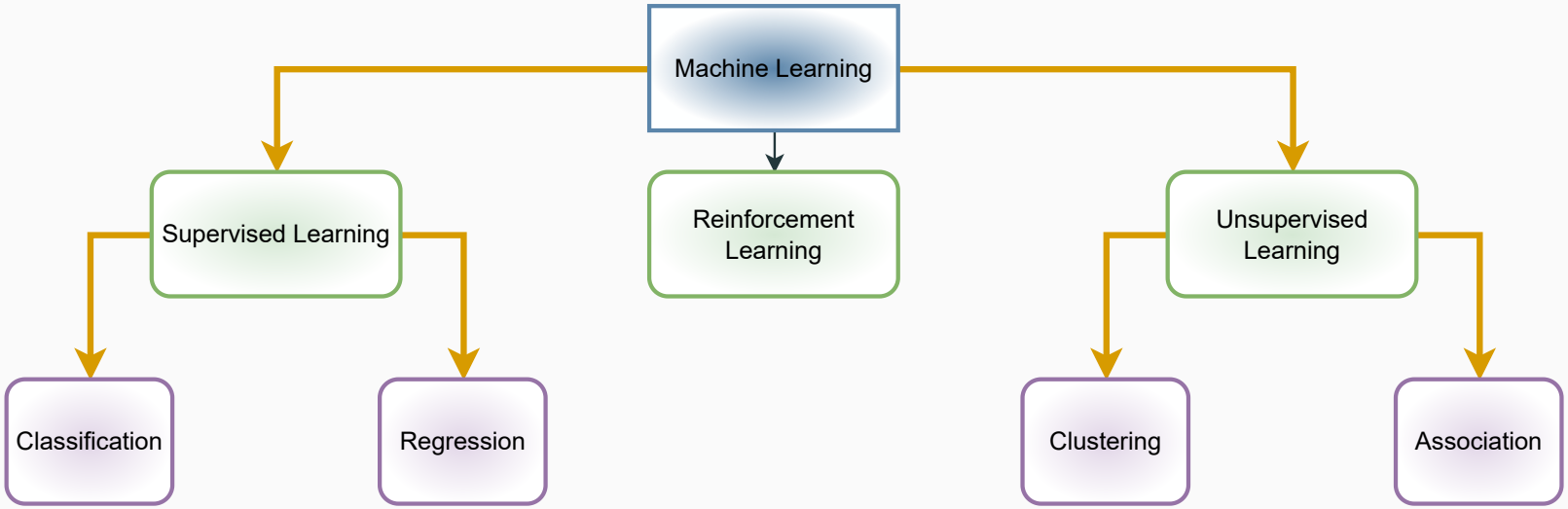
January 30, 2023

Lecturer, Metropolitan University

Table of contents

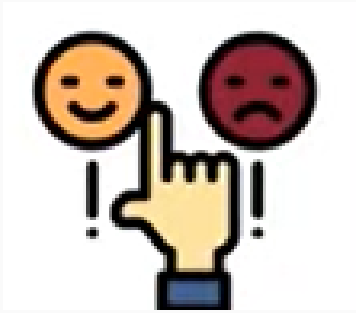
Categorization of Machine Learning

Categorization of Machine Learning i



Supervised Machine Learning

Supervised Machine Learning



Classification

Used When output is categorical like yes or no

Classification

- When the output variable is categorical (Example: Yes/No, True/Flase, Red/Blue)
- Used to predict/Classify the discrete values such as Male or Female, True or False, Spam or Not Spam, etc
- **Examples:** Email spam or not spam, Tumor cancerous or benign

Supervised Machine Learning

Regression

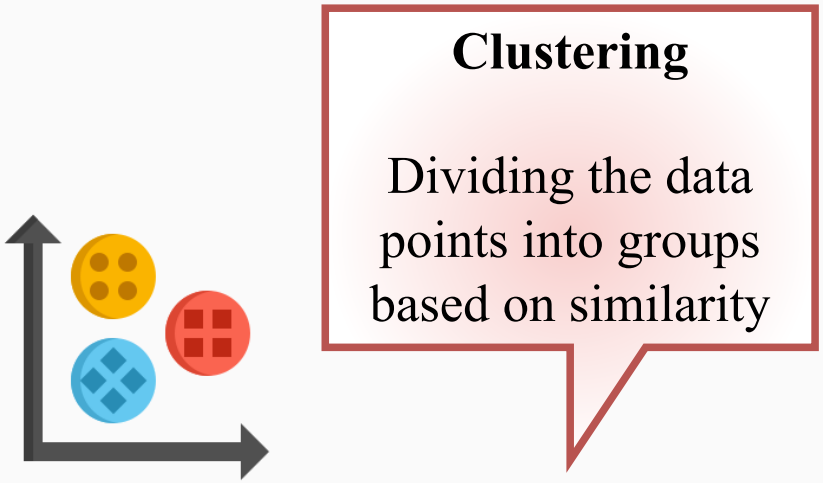
Used when a numerical value needs to be predicted

Regression

- When the output variable is numerical
- Used to predict the continuous values such as price, salary, age, etc
- **Examples:** Predicting house or property price, Predicting stock market price, Prediction of road accidents due to rash driving

Unsupervised Machine Learning

Unsupervised Machine Learning



Clustering

- Clustering is about grouping data points according to their similarities
- Examples: Groups news into coohesive groups,

Unsupervised Machine Learning

Regression

Used when a numerical value needs to be predicted

Association

- It is about discovering some interesting relationships between variables in large databases.
- Market basket analysis
- **Example:** People that buy a new house also tend to buy new furniture. It discovers the probability of the co-occurrence of items in a collection.

Supervised Classification Algorithms

Supervised Classification Algorithms

There are two types of classifications.

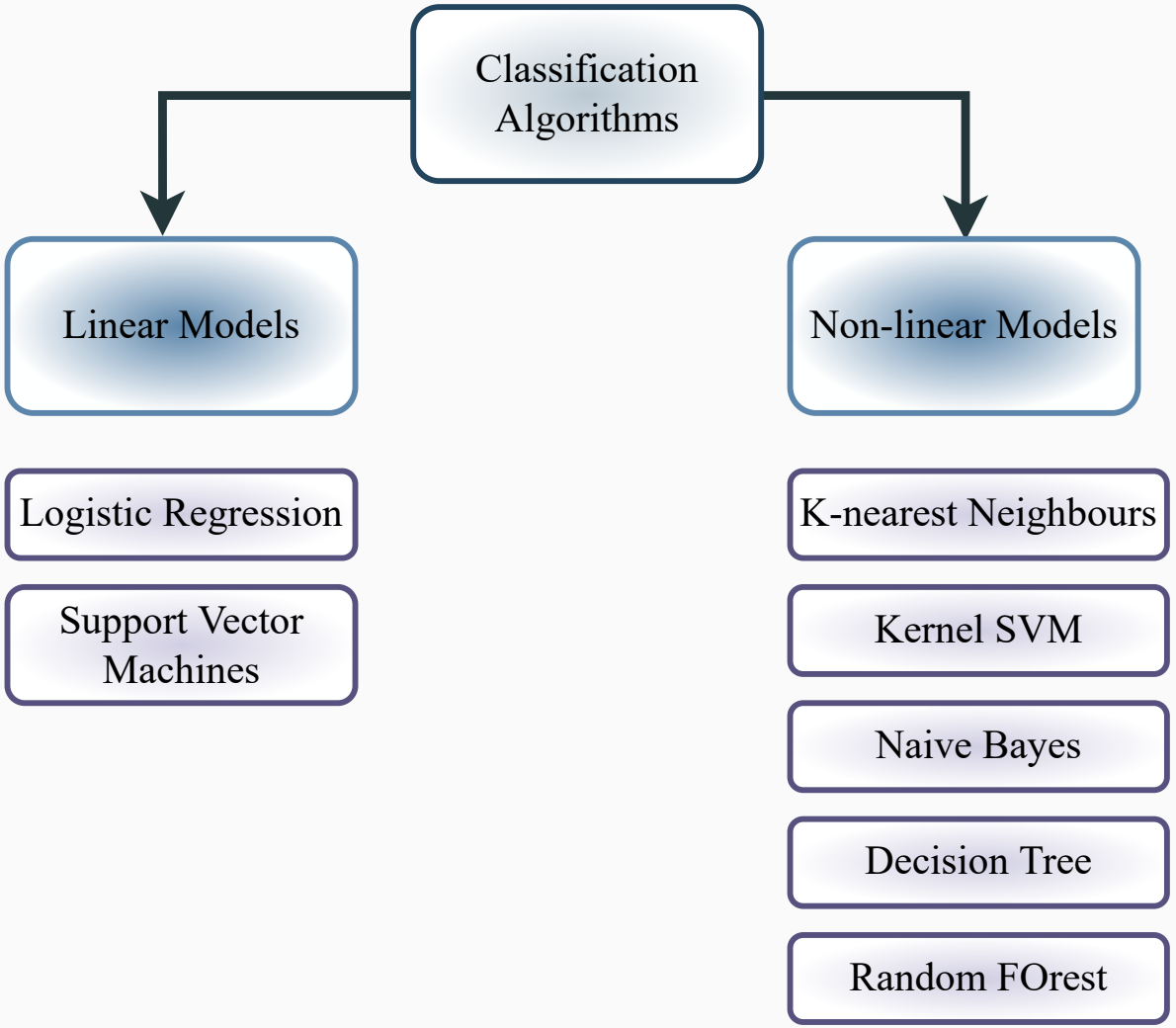
Binary Classification

- Classification problem has only two possible outcomes
- Examples: YES or NO, MALE or FEMALE, SPAM or NOT SPAM, CAT or DOG, etc.

Multiclass Classification

- Classification problem has more than two outcomes
- Examples: Classifications of types of crops, Classification of types of music.

Types of Classification Algorithms



Supervised Regression Algorithms

Supervised Regression Algorithms

Terminologies related to Regression Analysis:

Dependent Variable

The main factor in Regression analysis which we want to predict or understand is called the dependent variable. It is also called target variable.

Independent Variable

The factors which affect the dependent variables or which are used to predict the values of the dependent variables are called independent variable, also called as a predictor.

Outliers

Outlier is an observation which contains either very low value or very high value in comparison to other observed values. An outlier may hamper the result, so it should be avoided.

Supervised Regression Algorithms

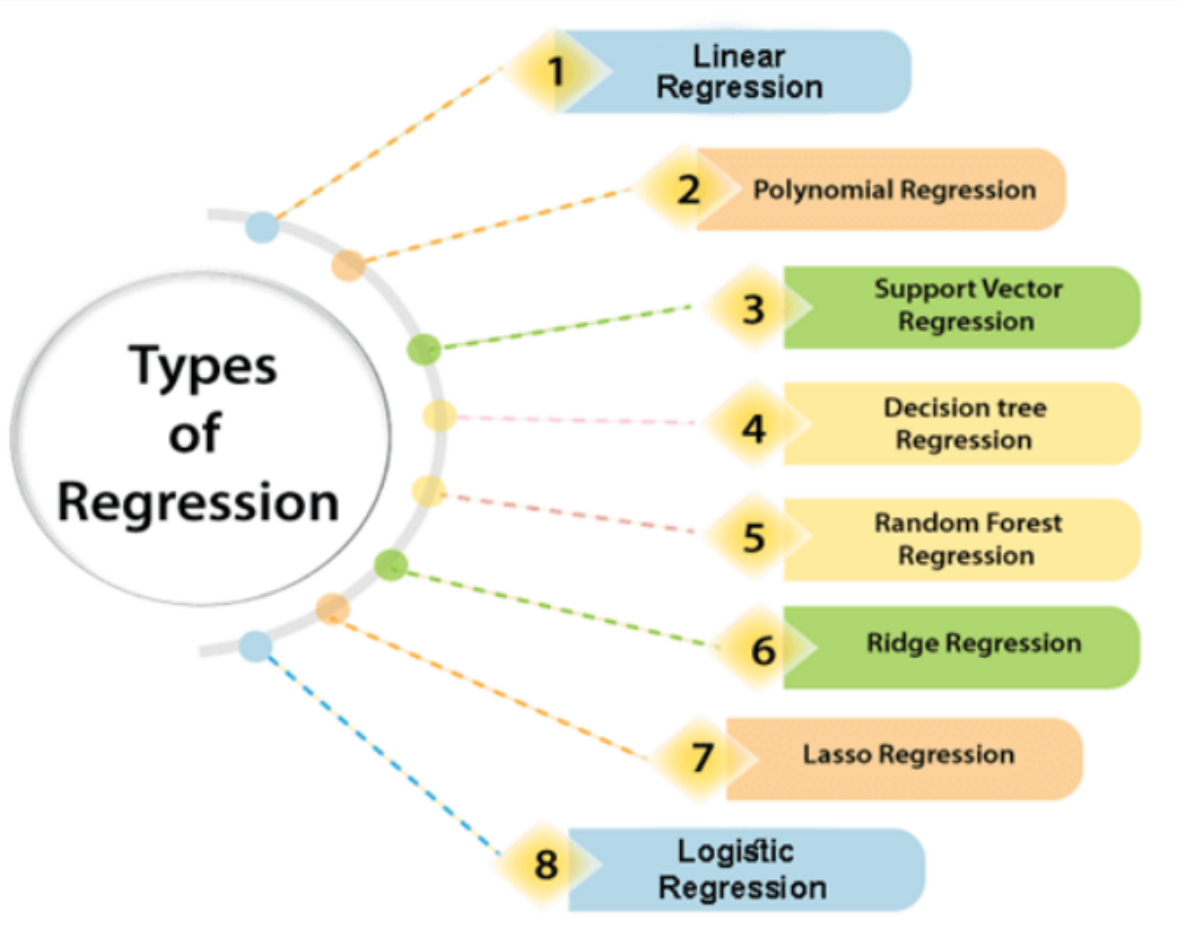
Multicollinearity

- If the independent variables are highly correlated with each other than other variables, then such condition is called Multicollinearity.
- Highly Correlated: change in one variable would cause change to another
- It should not be present in the dataset, because it creates problem while ranking the most affecting variable.

Underfitting and Overfitting

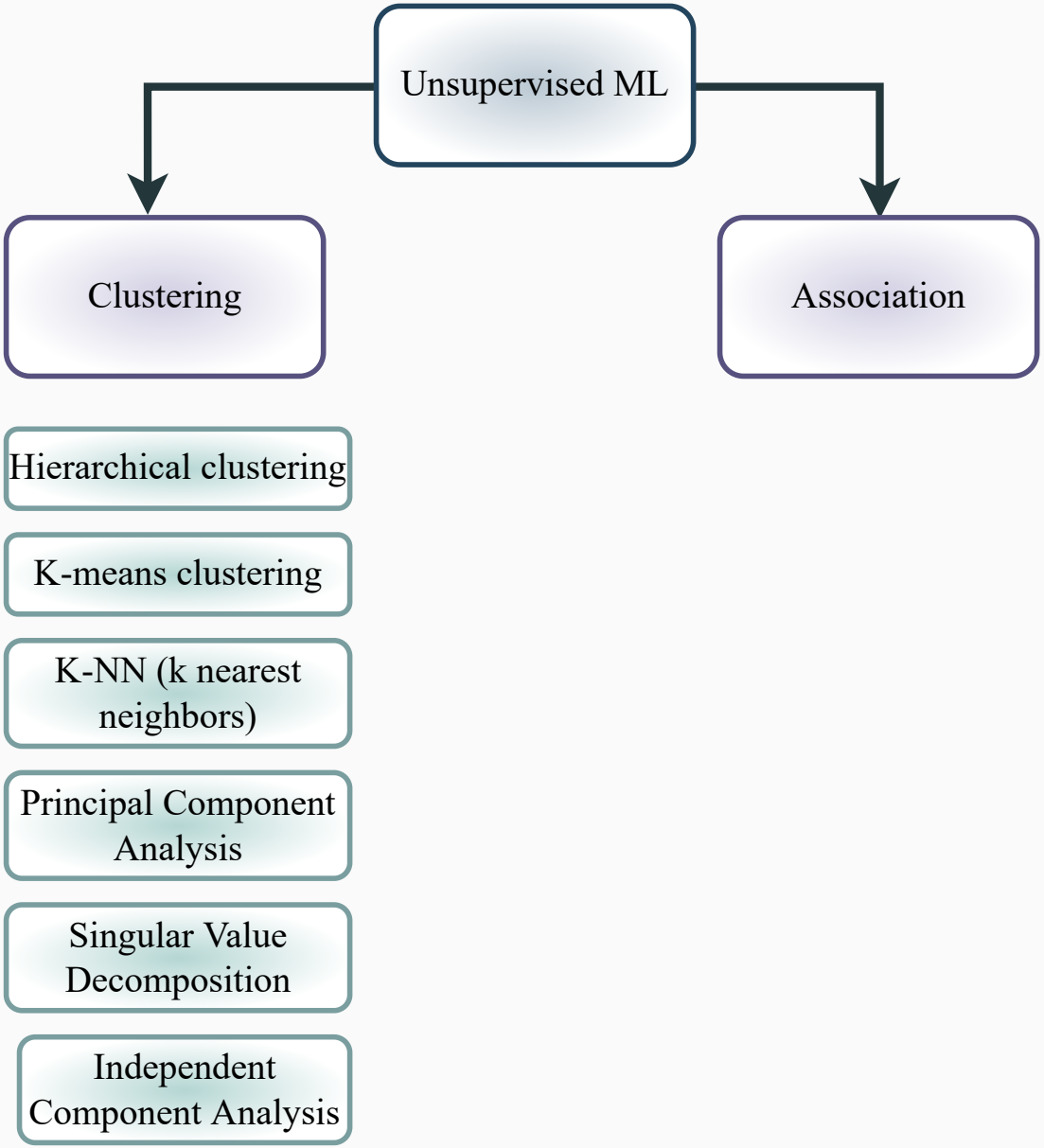
- **Overfitting:** The algorithm works well with the training dataset but not well with test dataset
- **Underfitting:** The algorithm does not perform well even with training dataset.

Types of Regression Algorithms



Unsupervised Machine Learning

Unsupervised Machine Learning




Thank
YOU!