AMERICAN INTERNATIONAL UNIVERSITY-BANGLADESH

Faculty of Science and Technology
AIUB
DEPARTMENT OF COMPUTER SCIENCE

CSC01893: COMPUTER VISION AND PATTERN RECONGNITION

Lecture: # 2 Week: # 1-2 Semester: Summer 2020-2021

BASICS OF CLASSIFICATION AND CLUSTERING

Instructor: Dr. Hossain Md Shakhawat, Assistant Professor,
Department of Computer Science, Faculty of Science & Technology.
shakhawat@aiub.edu

COURSE OUTLINE

- **Fundamentals of Image Processing**
- **Fundamentals of Computer Vision**
- **Fundamentals of Pattern Recognition**
- **Fundamentals of Machine Learning/Deep Learning**

- **Books & References:**
  - Computer Vision: Algorithms and Applications, 2nd ed by Richard Szeliski
  - *Digital image processing*; (Latest Edition) by Raphael C. Gonzalez.
  - Pattern recognition by Sergios Theodoridis

Supervised Learning

Supervised learning is a machine learning approach that's defined by its use of labeled datasets. These datasets are designed to train or "supervise" algorithms into classifying data or predicting outcomes accurately. Using labeled inputs and outputs, the model can measure its accuracy and learn over time. Two types : classification and regression:

Classification problems use an algorithm to accurately assign test data into specific categories, such as separating apples from oranges. Linear classifiers, support vector machines, decision trees and random forest are all common examples.

Regression is another type of supervised learning method that uses an algorithm to understand the relationship between dependent and independent variables. Regression models are helpful for predicting numerical values based on different data points, such as sales revenue projections for a given business. Some examples are linear regression, logistic regression and polynomial regression.

Unsupervised Learning

Unsupervised learning uses machine learning algorithms to analyze and cluster unlabeled data sets. These algorithms discover hidden patterns in data without the need for human intervention (hence, they are "unsupervised"). Unsupervised learning models are used for three main tasks: clustering, association and dimensionality reduction:

Clustering is a data mining technique for grouping unlabeled data based on their similarities or differences. Example, K-means clustering

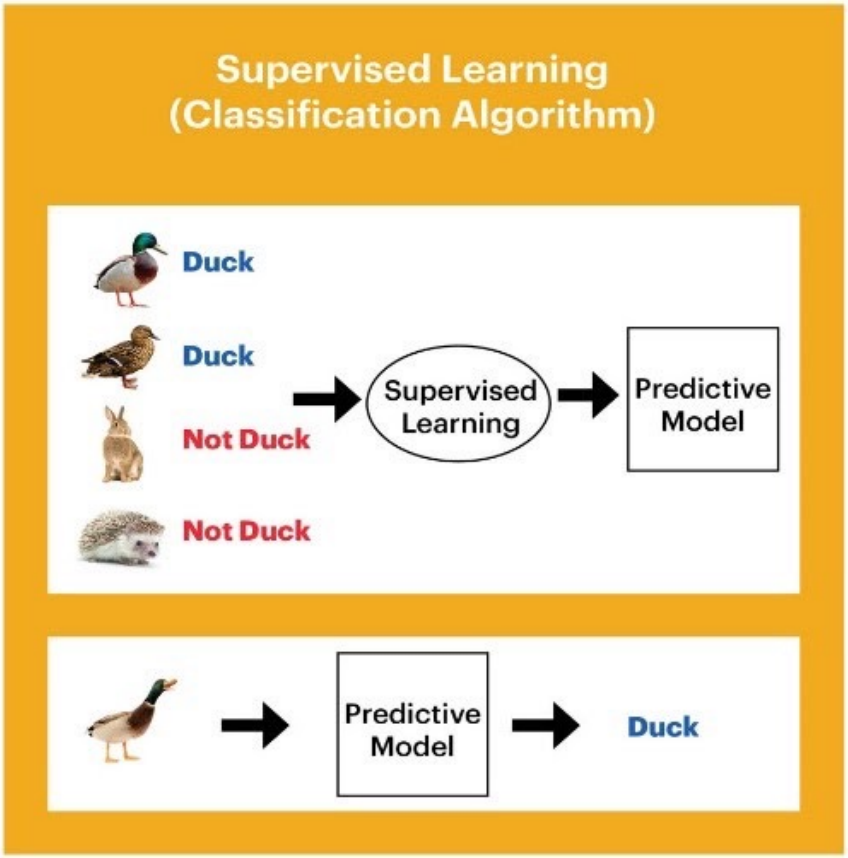
Association is another type of unsupervised learning method that uses different rules to find relationships between variables in a given dataset. These methods are frequently used for market basket analysis and recommendation engines, along the lines of "Customers Who Bought This Item Also Bought" recommendations.

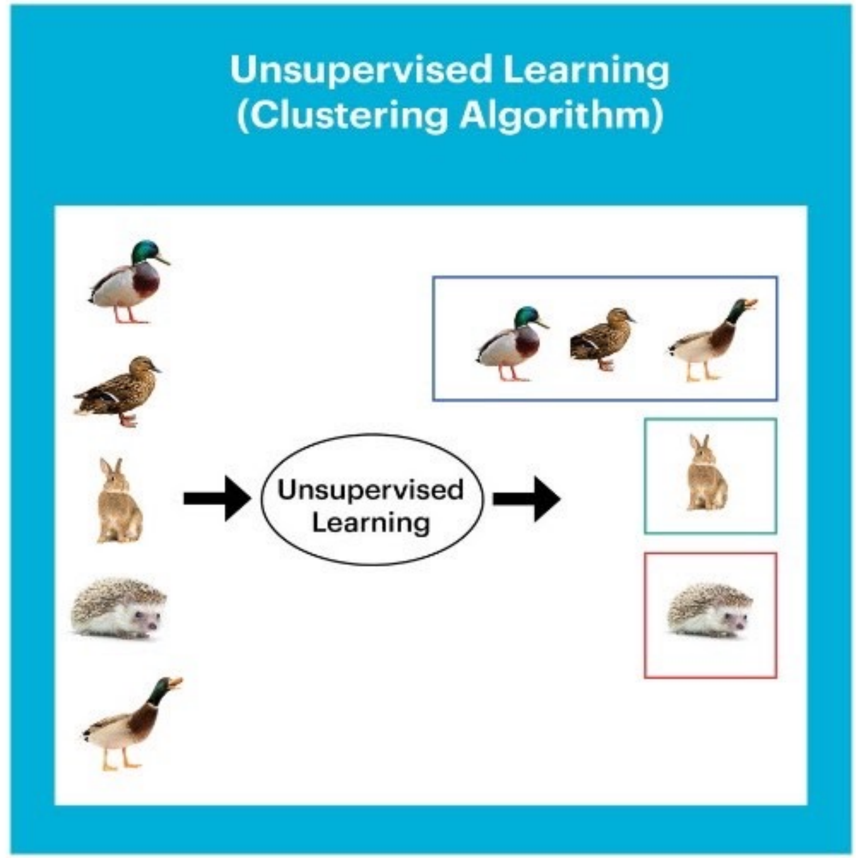
Dimensionality reduction is a learning technique used when the number of features (or dimensions) in a given dataset is too high. It reduces the number of data inputs to a manageable size while also preserving the data integrity. Often, this technique is used in the preprocessing stage, such as when autoencoders remove noise from visual data to improve picture quality.

SUPERVISED VS UNSUPERVISED METHODS

	Unsupervised	Supervised
Continuous	Clustering & Dimensionality Reduction SVD PCA K-Means	Regression Linear Polynomial Decision Trees Random Forests Nerual Networks
Categorical	Association Analysis Apriori FP-Growth Hidden Markov Model	Classification KNN Trees Logistic Regression Naive-Bayes SVN Nerual Networks

Supervised vs Unsupervised Learning





CLASSIFIERS/ CLASSIFICATION

Parametric vs non-parametric classifiers

Linear classifier vs non-linear classifier

PARAMETRIC VS NON-PARAMETRIC CLASSIFIERS

A machine learning algorithm can be classified as either parametric or non-parametric.

A **parametric** algorithm has a **fixed number of parameters**. A parametric algorithm is **computationally faster but** makes **stronger assumptions** about the data; the algorithm may work well if the assumptions turn out to be correct, but it may perform badly if the assumptions are wrong. A common example of a parametric algorithm is **linear regression**.

In contrast, a **non-parametric** algorithm uses a **flexible number of parameters**, and the number of parameters often **grows as it learns from more data**. A non-parametric algorithm is **computationally slower but** makes **fewer assumptions** about the data. A common example of a non-parametric algorithm is **K-nearest neighbor**.

To summarize, the **trade-offs** between parametric and non-parametric algorithms are in **computational cost** and **accuracy**.

LINEAR VS NON-LINEAR CLASSIFIERS

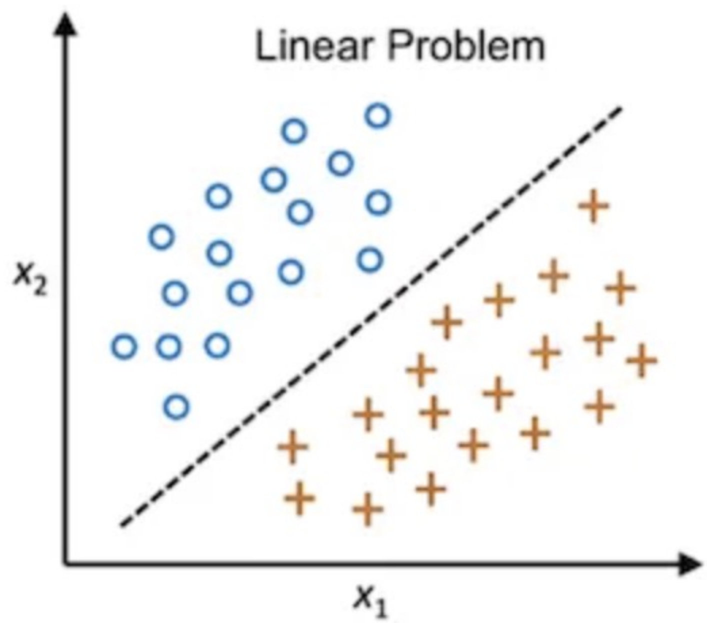
- Linear Classification refers to categorizing a set of data points to a discrete class based on a linear combination of its explanatory variables. On the other hand, Non-Linear Classification refers to separating those instances that are not linearly separable.

- Some of the classifiers that use linear functions to separate classes are *Linear Discriminant Classifier, Naive Bayes, Logistic Regression, Perceptron, SVM (linear kernel)*.

- Non-Linear Classification refers to categorizing those instances that are not linearly separable.
- Some of the classifiers that use non-linear functions to separate classes are Quadratic Discriminant Classifier, Multi-Layer Perceptron (MLP), Decision Trees, Random Forest, and K-Nearest Neighbours (KNN).

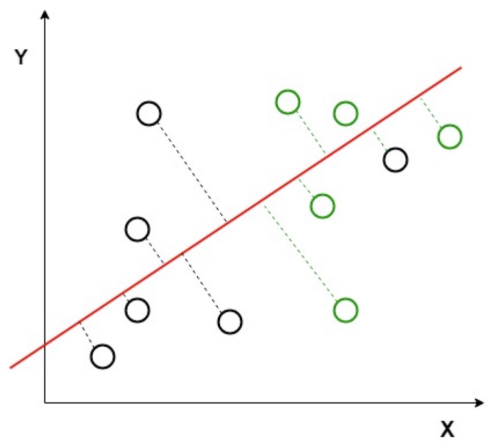
LINEAR PROBLEM

- In the figure, we have two classes, namely 'O' and '+.' To differentiate between the two classes, an arbitrary line is drawn, ensuring that both the classes are on distinct sides.
- Since we can tell one class apart from the other, these classes are called ‘linearly-separable.’
- However, an infinite number of lines can be drawn to distinguish the two classes.
- The exact location of this plane/hyperplane depends on the type of the linear classifier.



LINEAR CLASSIFIERS

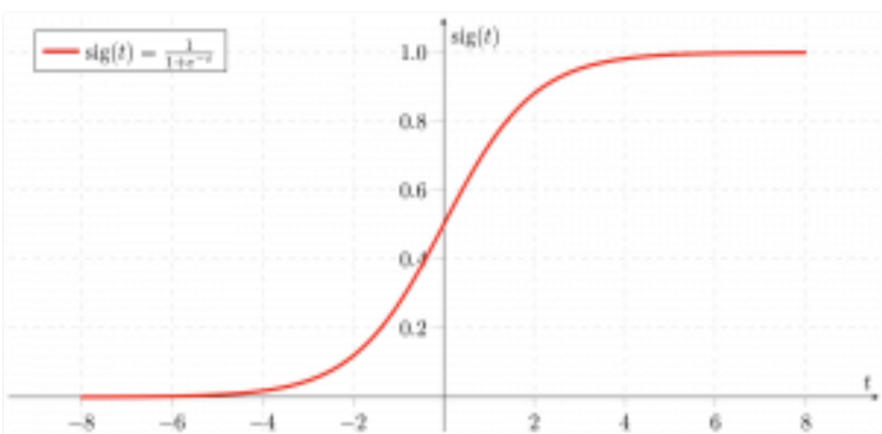
Linear Discriminant Classifier



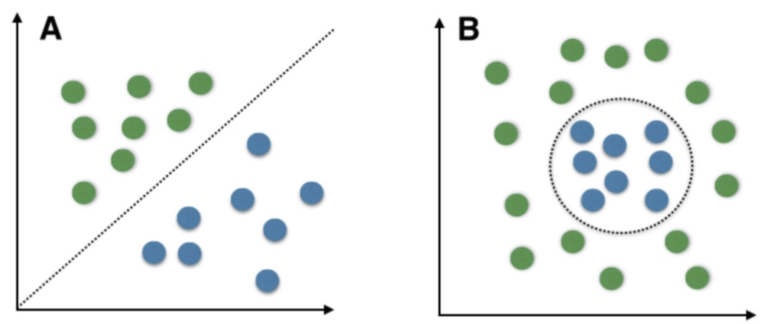
Naive Bayes

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

Logistic Regression



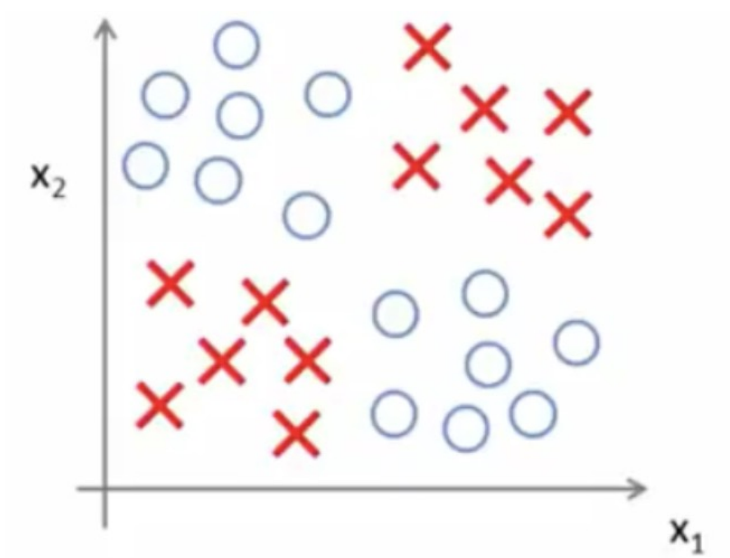
Support vector machine



A: Linearly Separable Data B: Non-Linearly Separable Data

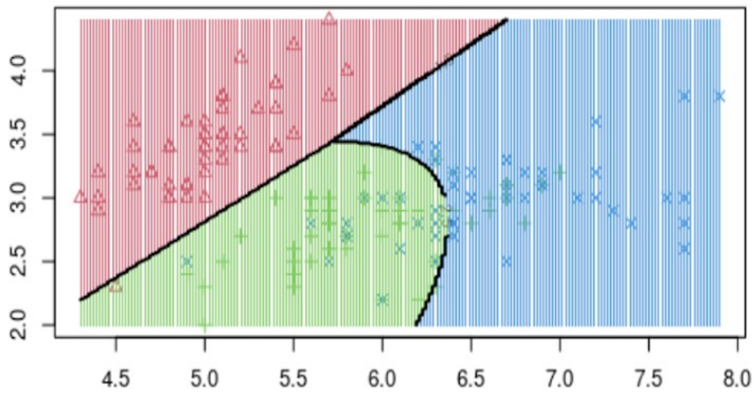
Non-linear problem

- In this figure, to differentiate between the two classes, it is impossible to draw an arbitrary straight line to ensure that both the classes are on distinct sides.
- We notice that even if we draw a straight line, there would be points of the first-class present between the data points of the second class.
- In such cases, piece-wise linear or non-linear classification boundaries are required to distinguish the two classes

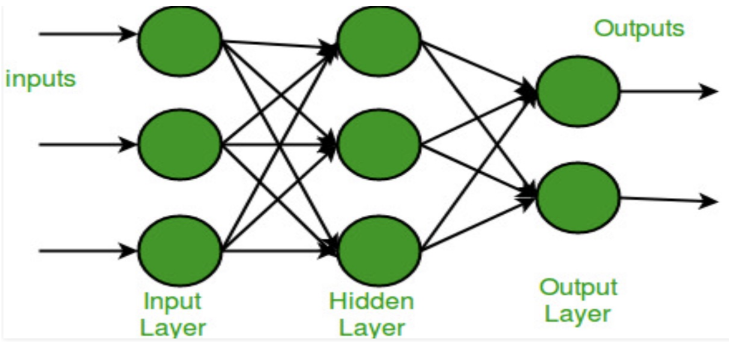


Non-linear Classifiers

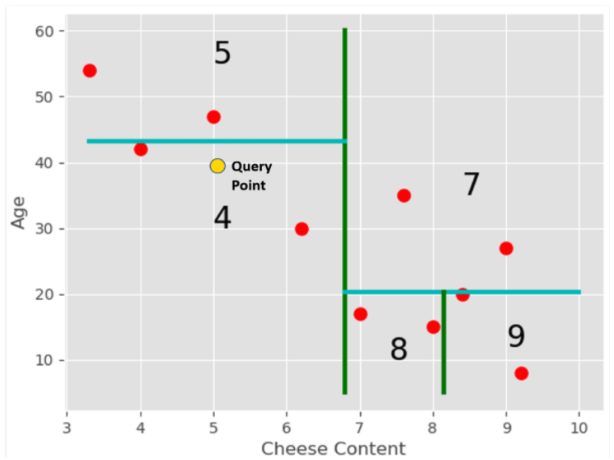
Quadratic Discriminant Classifier



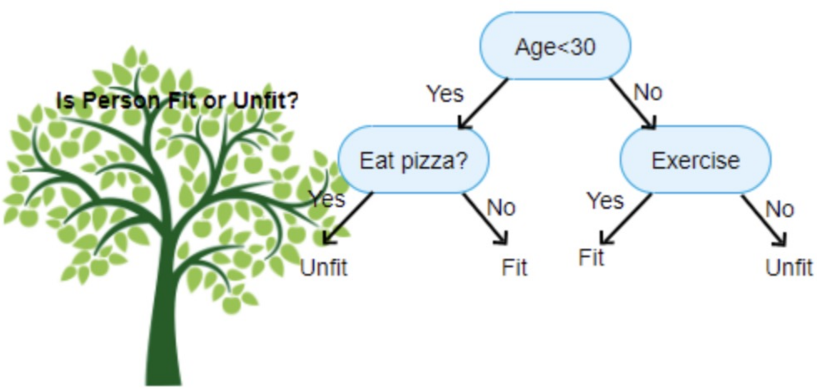
Multi-Layer Perceptron (MLP)



K-Nearest Neighbours



Decision Tree



KEY TAKEAWAYS

1.What is the difference between Linear Classification and Non-Linear Classification?

The main difference is that in the case of Linear Classification, data is classified using a hyperplane. In contrast, kernels are used to organize data in the Non-Linear Classification case.

2.Name a few linear classifiers.

Some of the popular linear classifiers are:
i) Naive Bayes
ii) Logistic Regression
iii) Support Vector Machine (linear kernel)

3.What are the most popular non-linear classifiers?

Some of the popular non-linear classifiers are:
i) Multi-Layer Perceptron (MLP)
ii) Decision Tree
iii) Random Forests
iv) K-Nearest Neighbors

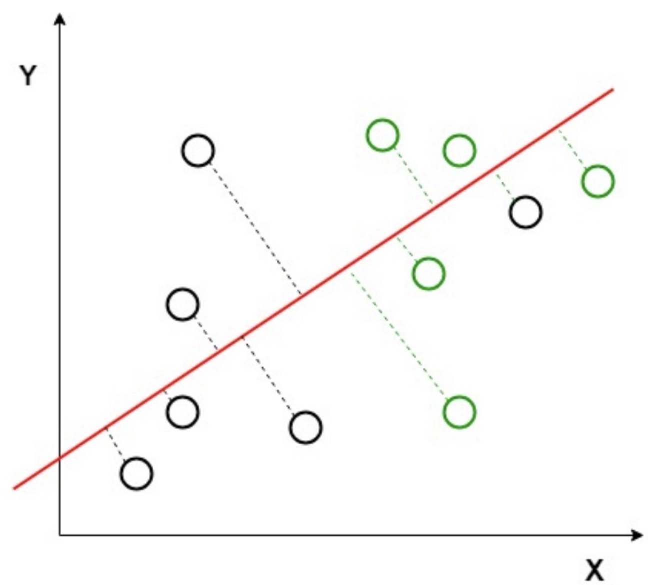
LDA

- It is a dimensionality reduction technique in the domain of Supervised Machine Learning.
- It is crucial in modeling differences between two groups, i.e., classes. It helps project features in a high dimensions space in a lower-dimensional space.

- Technique - Linear Discriminant Analysis (LDA) is used, which reduced the 2D graph into a 1D graph by creating a new axis. This helps to maximize the distance between the two classes for differentiation.

- In the graph, we notice that a new axis is created, which maximizes the distance between the mean of the two classes. As a result, variation within each class is also minimized.

- However, the problem with LDA is that it would fail in case the means of both the classes are the same. This would mean that we would not be able to generate a new axis for differentiating the two.



NAÏVE BAYES

It is based on the Bayes Theorem and lies in the domain of Supervised Machine Learning.
Every feature is considered equal and independent of the others during Classification. Naive Bayes indicates the likelihood of occurrence of an event. It is also known as conditional probability.

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

A: event 1

B: event 2

P(A|B): Probability of A being true given B is true - posterior probability

P(B|A): Probability of B being true given A is true - the likelihood

P(A): Probability of A being true - prior

P(B): Probability of B being true - marginalization

However, in the case of the Naive Bayes classifier, we are concerned only with the maximum posterior probability, so we ignore the denominator, i.e., the marginal likelihood. Argmax does not depend on the normalization term.