Diagnostic Methods

Non Model Diagnostic Methods

K Nearest Neighbors (KNN)

Outline

- Overview
- K Nearest Neighbors(KNN)
  - Definition
  - KNN Algorithm
  - Advantages/disadvantages
  - KNN in fault diagnosis
    - Medical applications
    - Industrial applications

Overview

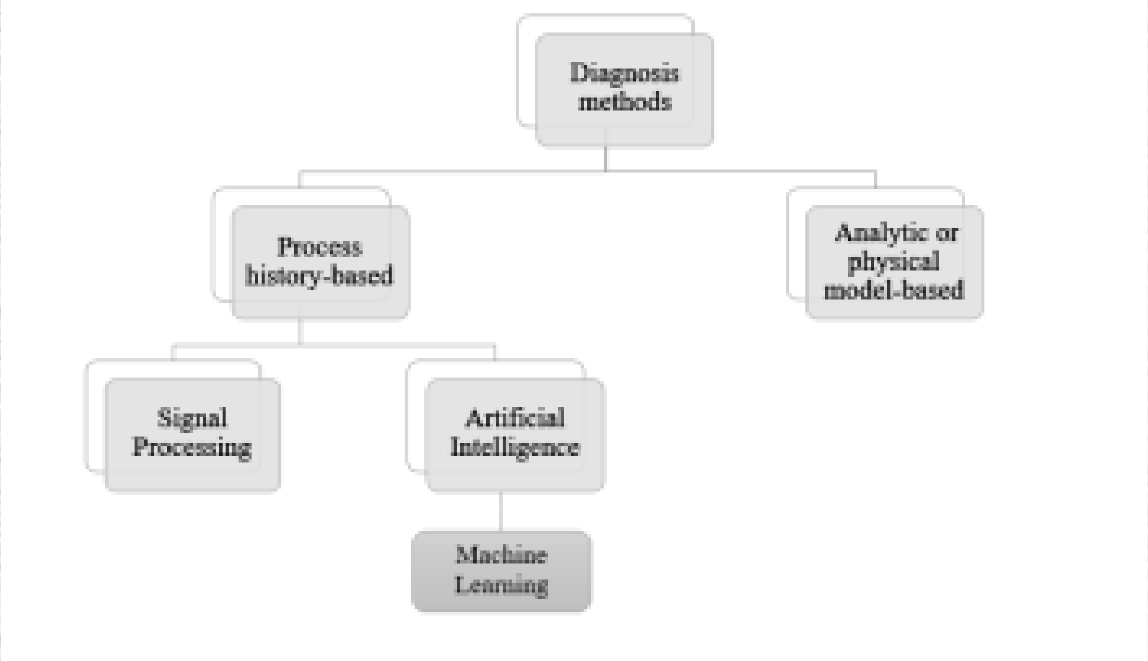
- In the literature, there are two main approaches for fault diagnosis:
  – Approaches that use an analytic or physical model of the system. The use of the techniques of diagnosis with models seems difficult and expensive, by offering less satisfactory performances. Besides, several industrial applications exist where a model is difficult even impossible to obtain due to increased complexity or several reconfigurations involved in the production process.

Overview

- Approaches that rely only on system observations. The artificial intelligence offers tools totally decoupled by the structure of the system, not requiring the preliminary modeling of the latter and allowing a real-time follow-up of its evolution. Besides, on-line reasoning makes that the approach of Artificial intelligence is stronger in changes in operating modes and have several reconfigurations.

Overview

- The different methods in fault diagnosis:



Overview

- Nearest neighbor algorithms are among the "simplest" supervised machine learning algorithms and have been well studied in the field of pattern recognition.
- KNN is also called a lazy learning algorithm since the processing of the training examples is postponed until making predictions.
  - meaning that it does not build a model using the training set until a query of the data set is performed.

Definitions

- An instance is assigned to the most common class among the instances similar to it.
  - How to measure similarity between instances?
  - How to choose the most common class?
- Tell me who your friends are and I'll tell you who you are!

Definitions

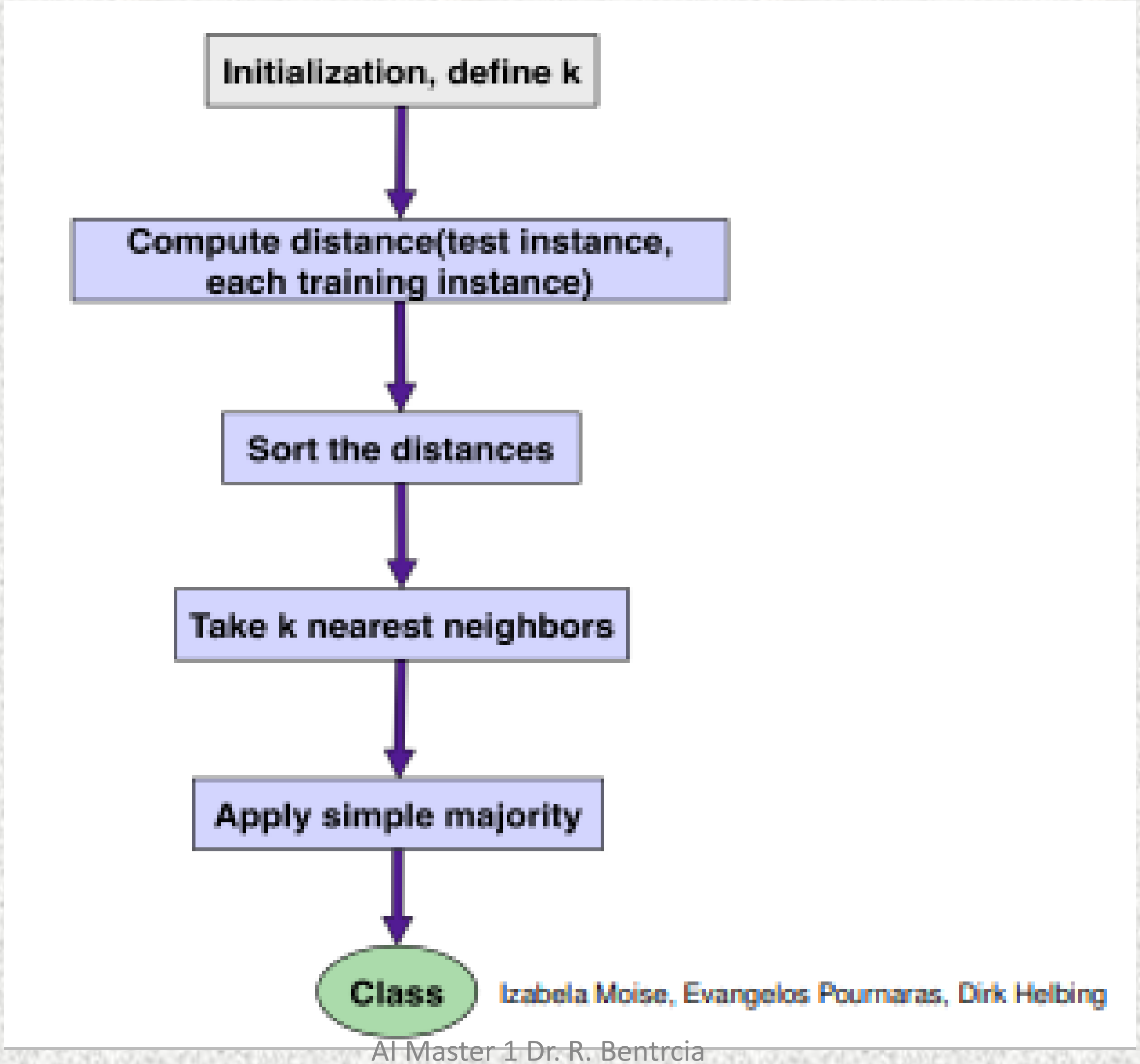
- A k-nearest-neighbor algorithm, often abbreviated k-nn, is an approach to data classification that estimates how likely a data point is to be a member of one group or the other depending on what group the data points nearest to it are in.

KNN Algorithm

- Steps to calculate KNN algorithm:

1. Determine parameter K = number of nearest neighbors
2. Calculate the distance between the query-instance and all the training samples
3. Sort the distance and determine nearest neighbors based on the K-th minimum distance
4. Gather the category of the nearest neighbors
5. Use simple majority of the category of nearest neighbors as the prediction value of the query instance

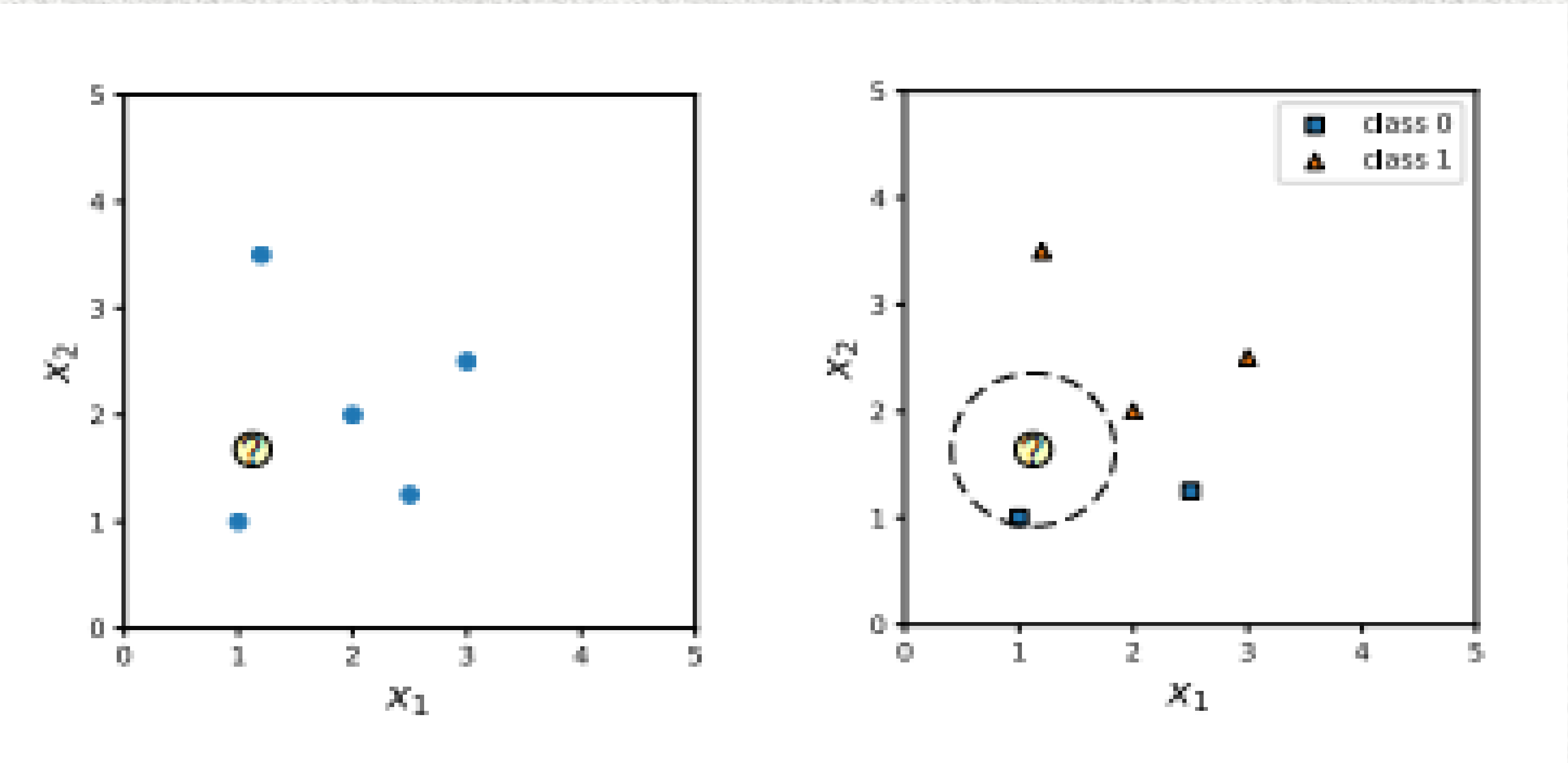
KNN Algorithm



KNN Algorithm

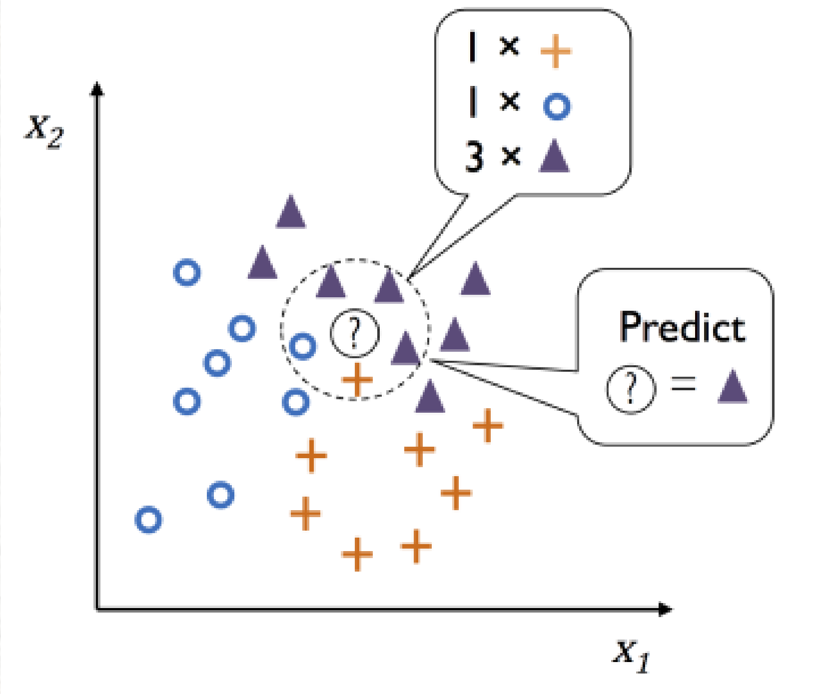
- Illustration of the nearest neighbor classification algorithm in two dimensions (features x1 and x2).
- On the left, the training examples are shown as blue dots, and a query point that we want to classify is shown as a question mark.
- On the right, the class labels are, and the dashed line indicates the nearest neighbor of the query point, assuming a Euclidean distance metric.
- The predicted class label is the class label of the closest data point in the training set (here: class 0).

KNN Algorithm



KNN Algorithm

- Illustration of KNN for a 3-class problem with k=5:



KNN Algorithm

- Distance Metrics:

1. Euclidean Distance

$$D(X,Y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

2. Manhattan Distance

$$D = \sum_{i=1}^{n}|x_i - y_i|$$

3. Minkowski Distance

$$D = \left(\sum_{i=1}^{n}|x_i - y_i|^p\right)^{1/p}$$

KNN Algorithm

- Choosing k: Classification is sensitive to the correct selection of k,
  - if k is too small ⇒ overfitting → algorithm performs too good on the training set, compared to its true performance on unseen test data
  - small k? → less stable, influenced by noise
  - larger k? → less precise, higher bias
- The optimal K value usually found is the **square root of N,** where N is the total number of samples.

Advantages/ Disadvantages

- Advantage:
  - Simple to implement and use.
  - Robust to noisy data by averaging k-nearest. neighbors.
  - KNN classification is based solely on local information.
- Disadvantage:
  - Curse of dimensionality: distance can be dominated by irrelevant attributes.
  - O(n) for each instance to be classified.
  - Large memory requirements.

KNN in fault diagnosis

- Medical applications:
  - The application of K-Nearest Neighbor (KNN) algorithm on an expert system for diagnose of lung disease.
  - KNN is used to predict heart disease.
  - Prediction of COVID-19 Possibilities using KNN Classification Algorithm.

KNN in fault diagnosis

- Industrial applications:
  - KNN is utilized for detection, classification, and location of short circuit faults.
  - KNN for fault diagnosis on a main engine journal-bearing