Nitte Meenakshi Institute of Technology,

Department of Computer Science and Engineering

18CSE751 Introduction to Machine

Learning Learning Activity Proposal:

"Analyzing Decision Tree and K-means Clustering using Iris dataset"

Agil Srinivasan(1NT19CS400), Guruprasad k(1NT19CS407)

Course Cordinator:- Dr. Vani V

Abstract

The decision tree is one of the most widely used machine learning algorithms due to its ease of interpretation. A decision tree consists of creating different rules by which we make the prediction The K-means algorithm, which is a very simple machine learning algorithm from the unsupervised learning method is used.

Introduction

As we all know from the nature, most of creatures have the ability to recognize the objects in order to identify food or danger. Human beings can also recognize the types and application of objects. An interesting phenomenon could be that machines could recognize objects just like us someday in the future.

This thesis mainly focuses on machine learning in pattern recognition applications. Machine learning is the core of Artificial Intelligence (AI) and pattern recognition is also an important branch of AI. In this thesis, the conception of machine learning and machine learning algorithms are introduced. Moreover, a typical and simple machine learning algorithm called K-means is introduced. A case study about Iris classification is introduced to show how the K-means works in pattern recognition.

Data Set

We will use the dataset from the Kaggle. The dataset is obtained from the following link –
https://www.kaggle.com/uciml/iris
The data set contains 3 classes with 50 instances each, and 150 instances in total, where each class refers to a type of iris plant.

Machine Learning Methods

Supervised Learning:
Supervised machine learning algorithms are trained to find patterns using a dataset. The process is simple, It takes what has been learned in the past and then applies that to the new data. Supervised learning uses labelled examples to predict future patterns and events.

Supervised learning is further divided into:
● Classification
● Regression

Decision Tree Classifier:
It tries to solve the problem, by using tree representation. Each internal node of the tree corresponds to an attribute, and each leaf node corresponds to a class label.

Unsupervised Learning:
Unsupervised learning is used against data without any historical labels. The system is not subjected to a pre-determined set of outputs, correlations between inputs and outputs or a "correct answer." The algorithm must figure out what it is viewing by itself, as it does not have any storage of reference points. The goal is to explore the data and find some sort of patterns or structures.

Unsupervised learning can be classified into:
● Clustering
● Association

K-means Clustering:
The goal of the K-means clustering algorithm is to find groups in the data, with the number of groups represented by the variable K. The algorithm works iteratively to assign each data point to one of the K groups based on the features that are provided.

Roles:

Agil Srinivasan- Supervised and unsupervised learning methods implementation.
Guruprasad K- K-means clustering and decision tree methods implementation.

Presentation and Visualization: -
Will be using graphs to show the presentation.

Schedule
The schedule is a table of dates and tasks that you plan to complete.

Date	Tasks to be Completed
17/01/21	Tasks completed by chosen date
20/01/22	Tasks to be completed by the final report/ presentation date

Bibliography
- https://www.kaggle.com/sixteenpython/machine-learning-with-iris-dataset
- https://towardsdatascience.com/exploring-classifiers-with-python-scikit-learn-iris-dataset-2bcb490d2e1b
- https://scikit-learn.org/stable/
- https://certes.co.uk/types-of-artificial-intelligence-a-detailed-guide/
- https://www.kaggle.com/uciml/iris