UECS3213 / UECS3453 Data Mining

SESSION: January 2019

Lab 10: Ensemble Learning in Python

Introduction

Scikit-learn is a free software machine learning library for the Python programming language.[3] It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

Ensemble learning uses multiple machine learning models to try to make better predictions on a dataset. An ensemble model works by training different models on a dataset and having each model make predictions individually. The predictions of these models are then combined in the ensemble model to make a final prediction. Every model has its strengths and weaknesses. Ensemble models can be beneficial by combining individual models to help hide the weaknesses of an individual model.

In this tutorial, you'll learn what ensemble is and how it improves the performance of a machine learning model.

Objectives

At the end of this lab, you are expected to acquire the following:

a) What is Ensemble learning?
b) How it improves the performance of a machine learning model?
c) Different Ensemble learning methods
d) Pitfalls of Ensembles
e) A Pythonic implementation of different Ensemble learning methods with a real test dataset
f) Further studies on Ensemble learning

Instruction

1. Visit the "Ensemble Learning in Python" at the following link: https://www.datacamp.com/community/tutorials/ensemble-learning-python

2. Follow the step-by-step instructions in the tutorial.

Other Related References

- https://scikit-learn.org/stable/modules/ensemble.html
- https://towardsdatascience.com/ensemble-learning-using-scikit-learn-85c4531ff86a
- https://www.analyticsvidhya.com/blog/2018/06/comprehensive-guide-for-ensemble-models/

The End