UECS3213 / UECS3453 Data Mining

SESSION: January, 2019

Lab 6 : Python Machine Learning - Introduction to scikit-learn

Introduction

Scikit-learn is a free software machine learning library for the Python programming language. Scikit-learn provides a range of supervised and unsupervised learning algorithms via a consistent interface in Python. It is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy. The library is focused on modeling data. It is not focused on loading, manipulating and summarizing data. For these features, refer to NumPy and Pandas.

Some popular groups of models provided by scikit-learn include:

- **Clustering**: for grouping unlabeled data such as KMeans.
- **Cross Validation**: for estimating the performance of supervised models on unseen data.
- **Datasets**: for test datasets and for generating datasets with specific properties for investigating model behavior.
- **Dimensionality Reduction**: for reducing the number of attributes in data for summarization, visualization and feature selection such as Principal component analysis.
- **Ensemble methods**: for combining the predictions of multiple supervised models.
- **Feature extraction**: for defining attributes in image and text data.
- **Feature selection**: for identifying meaningful attributes from which to create supervised models.
- **Parameter Tuning**: for getting the most out of supervised models.
- **Manifold Learning**: For summarizing and depicting complex multi-dimensional data.
- **Supervised Models**: a vast array not limited to generalized linear models, discriminate analysis, naive bayes, lazy methods, neural networks, support vector machines and decision trees.

Objectives

At the end of this lab, you are expected to learn the basics of Python machine learning:

- Loading the dataset.
- Exploring the dataset.
- Visualizing the dataset.
- Preprocessing the dataset.
- Constructing models and making some predictions.

Instruction

1. Go through the "Python Machine Learning: Scikit-Learn Tutorial" at https://www.datacamp.com/community/tutorials/machine-learning-python

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

Other Related References

- https://scikit-learn.org/stable/tutorial/basic/tutorial.html
- https://scikit-learn.org/stable/supervised_learning.html#supervised-learning
- https://machinelearningmastery.com/a-gentle-introduction-to-scikit-learn-a-python-machine-learning-library/
- https://www.datacamp.com/courses/supervised-learning-with-scikit-learn
- https://machinelearningmastery.com/machine-learning-in-python-step-by-step/

The End