OULU UNIVERSITY OF APPLIED SCIENCES

General Machine Learning Practices Using Python

Nibesh Khadka
Degree Programme in Information Technology,
Bachelor's Thesis, 15 credits

Objective

The thesis aims to introduce:

1. Machine Learning(ML) and its phases in theory.

2. Phases of ML in practice using Python programming language.

Introduction

ML is a process of teaching algorithms to learn. Algorithms try to find an underlying pattern between data points which can be used to predict future instances.

Figure 1 shows the categories in which ML can be divided into.

FIGURE 1. Categories of ml[1]

A typical ML model development process can be divided into the phases demonstrated in figure 2.

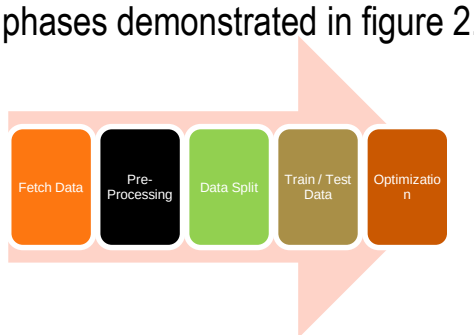

FIGURE 2: Phases in ml

Methods

Figure 3 demonstrates the proceeding and intentions of chapters in the thesis.

FIGURE 3. Chapter proceedings in thesis

Datasets

Three datasets were used in the thesis.

1. Datasets on the salary of employees on fake Company for preprocessing.
2. Load Boston Data for Linear Regression Algorithm (Supervised Learning).
3. Load Iris Data for K-Nearest Neighbor (KNN, Supervised Learning).

1. Preprocessing

Process of cooking data, an algorithm's food.

Steps:

1. Deal With Missing Values (Imputer / Pandas FillNa)
2. Deal with categorical values(Label + One Hot Encoder / Pandas Get_Dummies)
3. Normalize Data (Sklearn's Scaler)
4. Split Data (Sklearn's Train_Test_Split)

2. Train/ Test Data

Train and Test Model with train and test data respectively.

1. Linear Regression Model (Sklearn's Linear Model)
2. K-Nearest Neighbor (Sklearn's KNN)

3. Optimization

Process of Tuning ML model

1. Root Mean Square Error(RMSE) for regression evaluation, feature engineering for optimization.
2. Confusion Matrix and Classification Report for Evaluation, Elbow Method and GridSearchCV for optimization in KNN.

Result

1. Pre-processed IT salary data
2. Linear Regression Model for Boston Data
3. KNN model for Iris Data

Conclusion

1. Preprocess theory and codes can be used for any data.
2. Training and Testing model, Optimization process are similar for any Supervised Learning.

Future Steps

Learn mathematics and Intuition of different algorithms, practice with more algorithms.

References

[1] Categories of ML, Date of Retrieval 21.05.2019, https://qph.fs.quoracdn.net/main-qimg-dc432c347586a8c052b87bd3aad3b937

Date of Publication: 25.05.2019
Instructor: Kari Jyrkkä