Project Title

Multi Class Classification Predictive Model development for Iris flowers data (Python 3.x and Ubuntu 16.04)

Developed by: MSc. Shiva Agrawal

Place: Germany

Date: September 2018

Content

1. Introduction

1.1. Project description

The aim of the project is to develop the predictive model for the Iris flowers dataset. This dataset can be downloaded from [1]. This is multi-class classification type of Machine learning model. In this project, **python scikit-learn package** is used for the machine learning algorithms and data preprocessing and **Pandas** to import the dataset into python environment in form of **dataframe.**

The project is developed by referring the book Mastering Machine Learning using Python by Jason Brownlee [2], scikit-learn official website, Prof. Andrew Ng videos of machine learning and other online references.

1.2. Outline

- Chapter 1: Project description and outline of the project report
- Chapter 2: Python packages information and Iris Flowers dataset
- Chapter 3: Predictive model development
- Chapter 4: Results
- Chapter 5: Conclusion
- Chapter 6: References

2. Python packages and datasets

2.1. Python Packages

2.1.1. Scikit-learn

[3] This is open source package available for Machine Learning in Python. It is built on python other packages like numpy, scipy and matplotlib. It contains most of the required functions and tools to preprocess, analyze and develop the ML models.

2.1.2. Pandas

[4] It is an open source, BSD License library providing high performance, easy to use data structure and data analysis tool for the python programming language.

2.1.3. Numpy

[5] It is the fundamental package for scientific computing in python. It has powerful N dimension array object, sophisticated broadcasting functions and useful linear algebra, Fourier transform, and random number capabilities.

2.1.4. Matplotlib

[6] Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. It can generate plots, histograms, power spectra, bar charts, error charts, scatterplots, etc.

2.2. Iris Flowers dataset

It contains 150 samples, 4 features (all numeric), 1 output (type: string) (3 different classes): 150 x 5. This is for multiclass classification dataset

Features:

1. sepal length in cm
2. sepal width in cm
3. petal length in cm
4. petal width in cm

Outcome:

- class: Iris Setosa, Iris Versicolour, Iris Virginica

3. ML - Predictive Model Development (Classification)

The source code of the model is available in src folder of the repository. The source file contains well commented steps of the model development process. The model is developed by using the template available in the docs folder. Finally, the model is saved using the pickle package of python.

NOTE: Please first copy the required datasets into the corresponding folder from the src folder and then use.

4. Results

In this project, the model is developed without using object oriented or functional programming. Hence no separate test files are required. In order to run the code, just run the Multiclass Classification Model Development.py using python 3.x

The developed model is saved as LDA_model.sav inside results folder of the project repository. The same folder also contains the generated plots and result.txt file which contains the output from all the steps of the model development process.

5. Conclusion

Multiclass classification model for the Iris flowers dataset is developed and tested using different machine learning algorithms. Then best of all the algorithms is found and the model is saved for future use.

6. References

[1] "Iris dataset UCI archieve," [Online]. Available: https://archive.ics.uci.edu/ml/datasets/iris.

[2] J. Brownlee, Mastering Machine Learning using Python.

[3] "Home page," [Online]. Available: http://scikit-learn.org/stable/.

[4] "pandas," [Online]. Available: https://pandas.pydata.org/.

[5] "Numpy," [Online]. Available: http://www.numpy.org/.

[6] "Maplotlib," [Online]. Available: https://matplotlib.org/.

[7] "scipy," [Online]. Available: https://www.scipy.org/.