PG Certificate in Software Engineering for Data Science
Week 1 Assignment 2.2

In this assignment you will get familiar with data loading, preprocessing, training and testing phases involved in machine learning workflow. You can make use of the boiler plate code to begin with.

Programming Language: Python

Introduction

In this assignment, we will aim to determine the species of a flower based on the physical parameters of the flower.
This data set consists of the physical parameters of three species of flower — Versicolor, Setosa and Virginica. The numeric parameters which the dataset contains are Sepal width, Sepal length, Petal width and Petal length. In this data we will be predicting the classes of the flowers based on these parameters.

Dataset

- There are 150 observations with 4 features each (sepal length, sepal width, petal length, petal width).
- There are no null values, so we don't have to worry about that.
- There are 50 observations of each species (setosa, versicolor, virginica).

You can obtain the dataset from
https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data   or   sklearn's   inbuilt dataset function.

Tasks

To get started, there is a iris_starter.ipynb notebook given with the assignment, which can be used to do all the tasks.

1. Load the data and display the first few rows of the data.
2. Explore the dataset and plot the frequency of each label as a histogram.
3. Implement KNN classifier(using sklearn) and find the right number of neighbours that gives best performance for the credit scoring dataset.
4. Try implementing two other classifiers of your choice and report the performance.
5. Plot confusion matrix to better analyze the results.

Bonus Tasks

6. Use k-fold validation while training inplace of normal random split in the starter code.

7. Implement any one classifier (without sklearn) used in this experiment.

Environment

We will be expecting that you run the notebook in python3.8 environment with the following packages:
- numpy
- scikit-learn
- pandas
- matplotlib
- seaborn

Submission

Make changes in the iris_starter.ipynb notebook and submit the same on the portal.