MEM T380 – Applied Machine Learning in Mechanical Engineering

Case Studies Assignment 1

Dimitrios Fafalis, PhD

`Dimitrios.Fafalis@drexel.edu`

due Tuesday, April 25, 2023

Students' names & ID:

1. _____   _____

2. _____   _____

Submit your files (`.ipynb`) and a report (`.pdf` or `.html`) on Blackboard by due date.

Exploring datasets!

The very first steps in developing a `Machine Learning` model are to load, explore, and pre-process the data. The goal of this assignment is to familiarize ourselves with basic operations such as importing the data, cleaning the data, exploring the data.

CASE STUDY 1. points 25 – **Iris Dataset**

This case study works on the *Iris* data-set available to load within the `scikit-learn` package. The raw data and a description is available in the website The Iris Dataset.

Your task is to create a Jupyter notebook, load the dataset and explore it thoroughly, in a similar way we did in class for the *Auto-MPG* data-set.

To get started, import the necessary python packages as follows:

```
1    import numpy as np
2    import pandas as pd
3    import matplotlib.pyplot as plt
4    import seaborn as sns
5    %matplotlib inline
```

Then, use the following command to load the dataset into your notebook:

```
1    # import some data to play with
2    iris = datasets.load_iris()
```

Explore the information available in the loaded dataset, and organize them into `numpy` arrays and `pandas DataFrames`. Look for missing data, rename the columns with convenient naming, identify the features and the output variables, identify numerical and categorical variables, etc. Create all varieties of plots to visualize the data in the dataset and explain what you see. Let the data tell you their stories and document them in `markdown` cells.

For every code cell you write in your `Jupyter` notebook, explain what you attempt to do. You can either write this as a comment in the dame coding cell after the symbol #, or you can add `markdown` cells before and/or after the coding cell.

CASE STUDY 2.   points 25 – **3D Printer Dataset**

This mini case study contains data from a 3D printer. The datsaset was downloaded from 3D Printer Dataset for Mechanical Engineers. The dataset is available to you in the **Case Studies** folder for HW-1, with the file name `data_3D_printer.csv`. When you create your own Jupyter notebook, make sure you place it in the folder you will be using the notebook.

Use the following command to load the dataset into a `pandas DataFrame`:

```
1           data = pd.read_csv('data_3D_printer.csv')
```

Refer to pandas.read_csv documentation for examples and additional options.

Your task is to explore the data in a similar way you did for case study 1.

CASE STUDY 3.   points 50 – **Tensile Properties of Austenitic Stainless Steel**

This mini case study contains data related to tensile properties of austenitic stainless steel, and how they are affected by various parameters, such as composition, manufacturing processes, etc. The dataset was downloaded from Materials Algorithms Project Program Library. The dataset is available to you in the **Case Studies** folder for HW-1, with the file name `STMECH_AUS_SS.xls`. When you create your own Jupyter notebook, make sure you place it in the folder you will be using the notebook. Notice that the data are stored in an excel file, and to read it directly to a `pandas DataFrame` you should use the following command:

```
1           data = pd.read_excel('STMECH_AUS_SS.xls')
```

Refer to pandas.read_excel documentation for examples and additional options.

Your task is to explore the data in a similar way you did for case study 1.