CS7052-Machine Learning

Workshop 2

A First Application, Classifying Iris species

Python, Jupyter notebook, more on Pandas library

You will learn:

- To run your first application
- To learn more about data analytics with python, pandas

Task 1- A First Application, Classifying Iris Species

Open Muller & Guido's book from the reading list. Open your Jupyter notebook.

Follow the instructions on pages 13-23 in chapter 1.

Muller and Guido's book comes with accompanying code, which you can find on https://github.com/amueller/introduction_to_ml_with_python.

You can download the code and then open 01-introduction.ipynb in your Jupyter Notebook.

Make sure you understand the meaning of each line of code, make some changes to improve your understanding.

Answer the following questions:

W2.1 How many samples are in Iris dataset?

W2.2. How many features are in Iris dataset?

W2.3. List features and target variable of the sample number 99.

W2.4. What is figure 1.3 is telling you about Iris data set?

Show the output to your tutor when you are done and answer Q1-4 of the test under week 2 folder.

Task 2- Python libraries, more on Pandas

Open Nelli's book, refer to '**Chapter** 5: pandas: Reading and Writing Data' to learn how to upload a csv file in Jupyter notebook. Refer to www.Kaggle.com and navigate to datasets.

Open UK inflation data 1989-2022 in your Jupyter notebook.

W2.5. Which year and month has the highest inflation in UK inflation data 1989-2022?

Refer to '**Chapter 6**: pandas in Depth, data manipulation', read the section related to string manipulation.

W2.6. Write a snippet of code using pandas in Jupyter notebook to count the number of times ChatGPT is repeated in this text:

"ChatGPT is an artificial intelligence (AI) chatbot that uses natural language processing to create humanlike conversational dialogue. The language model can respond to questions and compose various written content, including articles, social media posts, essays, code and emails. ChatGPT, which stands for Chat Generative Pre-trained Transformer, is a large language model-based chatbot developed by OpenAI and launched on November 30, 2022, which enables users to refine and steer a conversation towards a desired length, format, style, level of detail, and language"

Show your code to your tutor and answer Q5 and 6 of the test under week 2 folder.