MACHINE LEARNING

CONTINUOUS ASSESSMENT

Module Assessments

- Python Machine Learning Continuous Assessment
  - **Part 1 - Regression, Classification, Decision Trees**
  - **Part 2 – Neural Network**

Part 1's Requirements (1)

- Select datasets from the followings
  - https://data.gov.sg/
  - https://archive.ics.uci.edu/ml/index.php
  - https://www.kaggle.com/datasets
  - The source dataset must contain a minimum of 300 rows of data

Part 1's Requirements (2)

- Develop machine learning models (total 3 sets of models)

  a) Classification techniques

  Choose three of the followings and do a comparison of the outcome:

  - Logistic Regression
  - K-NN Classification
  - Decision Trees

  b) Perform data or feature engineering to evaluate the impact on the outcome

Part 1's Deliverables

- Dataset
- Python Codes
- Write-up
  - Problem statement
  - Simple Data Dictionary of training model (only the essential columns involved in training), including the source of dataset
  - Duration of training and prediction for each model
  - Sharing of the process, for example,
    - impact of data engineering, feature engineering on the outcome
  - Conclusion
  - What you have learnt

Part 2's Requirements (1)

- A **Fruits** dataset will be provided to you, and it consists of images of (1) apples, (2) oranges, (3) bananas and (4) a mix of fruits
- Implement a **CNN** to recognize the 4 classes. Use the images in the "train" folder to train your image classifier
- Test the accuracy of your image classifier using images from the "test" folder
- Document what you did to increase the accuracy of your image classifier (e.g. image augmentation)
- Use Matplotlib to plot the **loss** and **accuracy** of your **training** (pick any of your better runs)
- Specify the best/final **accuracy** of your image classifier on the **test data**

Part 2's Deliverables

- Python code for your CNN implementation
- A short and to-the-point report (not more than 2 pages)
  - Document the improvements you made to your CNN model that increases its accuracy
  - Plot(s) to show the **loss** and **accuracy** of your classifier on the **training data** (pick any of your better runs)
  - State the best accuracy of your classifier on the **test data**
  - A reflection of lessons learnt in building your network

Timeline

- Uploading of deliverables
  - Deliverables - 17th Nov (Tue), by 11pm

- Peer Evaluation
  - Scores will be adjusted according to peer evaluation