University of Jaffna, Sri Lanka

Bachelor of Science Degree Examination in Computer Science, Level 3S – 2022

Bachelor of Science Honours Degree Examination, Level 3M – 2022

In-Course Assessment 2 (Practical)

CSC311S3/CSC306M3: Machine Learning

Answer All Questions Time Allowed: Two Hours

You are required to develop a simple end-to-end machine learning model in MATLAB to classify given dataset. Your implementation should cover the following aspects:

(i). **Data Loading:** Load the Iris dataset provided in the folder named 'ML-ICA-2' into MATLAB. Ensure that you understand the structure of the data, including the features and target labels.

(ii). **Data Preprocessing:** Handle any missing values if present, normalise or standardise the features as necessary, and split the dataset into training (e.g., 70%) and testing (e.g., 30%) subsets.

(iii). **Model Selection:** Implement the following classification algorithms

a) k-Nearest Neighbours (k-NN)

b) Linear One-versus-All (OVA) Support Vector Machines (SVMs) using the SVM-Light built-in functions: `svmlopt`, `svmlwrite`, `svmlread`, `svm_learn`, and `svm_classify`

(iv). **Model Training**: Train your selected models on the training subset and adjust any relevant hyperparameters such as k-NN with k = 1, 3, 5 and SVM with *C* ranging from 2^{-5} to 2^{4} in powers of 2.

(v). **Model Evaluation:** Evaluate your model's performance on the testing subset and report the classification accuracy.

Submit your MATLAB code, ensuring it is well-commented to explain your approach, and provide a brief report summarising your findings, including comparisons of model performance.

Note: Change the name of the 'ML-ICA-2' directory on your computer to match your student registration number, such as "2020csc200" or "2020sp200." Place all your source code and the table comparing classification rates into this newly renamed folder. Then, compress the folder into a zip file before submitting it through the LMS course page.