CLL788 - Process Data Analytics

Assignment 2

Deadline: 20th March 11:55 PM

1. A university conducts 2 exams – Aptitude & Verbal as its entrance test to a 2- year program. Based on the scores of these 2 papers, admission is given to students. University has not mentioned the exact criteria of selection. Based on historical data, you need to predict whether a student will get admission based on his/her scores in the 2 exams. Data is provided in q1train.csv & q1test.csv. Train.csv contains training data. First column contains the score of the Aptitude exam, 2nd column contains the score of the verbal exam and 3rd column indicates whether that student got admission or not. 0 indicates not selected whereas 1 means selected. q1test.csv contains test data.

    A) Apply perceptron on training data with the first two columns as input data and the third column as output. Use any suitable learning rate. Test your resultant model on test data*. Write your code from scratch. No use of any toolkits like scikit-learn, PyTorch, TensorFlow etc.*
    B) Build an MLP (Multi-Layer Perceptron) with one hidden layer on the training data with first two columns as input data and the third column as output. Make suitable assumptions for parameters like learning rate etc. when needed. Test your resultant model on test data. *Write your code from scratch. No use of any toolkits like scikit-learn, PyTorch, TensorFlow etc.* Once your model is trained, using any toolkit, implement MLP and compare results of the toolkit model with your model.
    C) Train a Logistic Regression model on the training data (code from Assignment 1 can be used). Make suitable assumptions for parameters when needed. Test the model on the test data.
    D) Compare Logistic Regression model results with your models built in sub questions (A) and (B)
    E) You can use ChatGPT to generate your responses/code. However, code generated by ChatGPT might not be syntactically correct. Do ensure that it's correct and working fine. Report the bugs you find in generated responses from ChatGPT. Do you see a pattern? Report your findings and solutions.

We strongly suggest use visualization methods like plots, confusion matrices etc. to better demonstrate and compare your results.

Submission Details

1. Submit two files, first is a pdf report describing your work with all the graphs and analysis included, second is a zip file containing codes on Moodle.
2. Name the files <EntryNumber>_report_assgn2.pdf and < EntryNumber >_codes_assgn2.zip respectively. Only MATLAB(.m) & python (.py or .ipynb) are allowed for this assignment.
3. The deadline for the submission is 20th March 11:55 PM.