IT3071 – Machine Learning and Optimization Methods 2024

Practical Assignment 1: Building a deep learning model with TensorFlow

Objective: The objective of this practical assignment is to create a deep learning model using TensorFlow to perform classification on a tabular dataset. By the end of this assignment, students should have a better understanding of how to design, train, and evaluate neural networks for tabular data using TensorFlow.

Dataset: The dataset consists of 480 student records and 17 features. The features are classified into three major categories: (1) Demographic features such as gender and nationality. (2) Academic background features such as educational stage, grade Level and section. (3) Behavioral features such as raised hand on class, opening resources, answering survey by parents, and school satisfaction.

Attributes

- Gender - student's gender ('Male' or 'Female’)
- Nationality- student's nationality (’ Kuwait’,’ Lebanon’,’ Egypt’,’ SaudiArabia’,’ USA’,’ Jordan’,’
- Venezuela’,’ Iran’,’ Tunis’,’ Morocco’,’ Syria’,’ Palestine’,’ Iraq’,’ Lybia’)
- Place of birth- student's Place of birth (’ Kuwait’,’ Lebanon’,’ Egypt’,’ SaudiArabia’,’ USA’,’ Jordan’,’ Venezuela’,’ Iran’,’ Tunis’,’ Morocco’,’ Syria’,’ Palestine’,’ Iraq’,’ Lybia’)
- StageID- educational level student belongs (‘ lowerlevel’,’MiddleSchool’,’HighSchool’)
- Grade ID (‘G-01’, ‘G-02’, ‘G-03’, ‘G-04’, ‘G-05’, ‘G-06’, ‘G-07’, ‘G-08’, ‘G-09’, ‘G-10’, ‘G-11’, ‘G-12 ‘)
- Section ID- classroom student belongs (’A’,’B’,’C’)
- Topic- course topic (’ English’,’ Spanish’, ‘French’,’ Arabic’,’ IT’,’ Math’,’ Chemistry’, ‘Biology’, ‘Science’,’ History’,’ Quran’,’ Geology’)
- Semester- school year semester (’ First’,’ Second’)
- Parent responsible for student (’mom’,’father’)
- Raised hand- how many times the student raises his/her hand on classroom (0-100)
- Visited resources- how many times the student visits a course content(0-100)
- Viewing announcements-how many times the student checks the new announcements(0-100)
- Discussion groups- how many times the student participate on discussion groups (0-100)
- Parent Answering Survey- parent answered the surveys which are provided from school or not (’Yes’,’No’)
- Parent School Satisfaction- the Degree of parent satisfaction from school(’Yes’,’No’)
- Student Absence Days-the number of absence days for each student (above-7, under-7)
- The students are classified into three numerical intervals based on their total grade/mark:
    - Low-Level: interval includes values from 0 to 69,
    - Middle-Level: interval includes values from 70 to 89,
    - High-Level: interval includes values from 90-100.

Link to Access dataset:

https://github.com/lakminia/academic_data

Your task is to predict the Class (H/M/L)

Task Description:

1. **Data Preparation (20 points):**
    - Load the tabular dataset.
    - Preprocess the data, including handling missing values, encoding categorical variables (if any), and scaling numerical features.
    - Split the dataset into training and validation sets.
2. **Model Design (25 points):**
    - Create a suitable deep learning model architecture using TensorFlow's Keras API for the classification task.
    - Experiment with different architectures, including varying the number of layers, neurons per layer, activation functions, and dropout layers (if necessary).
    - Choose an appropriate loss function and optimizer.
3. **Training (20 points):**
    - Train your model on the training data.
    - Monitor training progress by logging training loss and accuracy.
    - Implement early stopping to prevent overfitting.
4. **Evaluation (15 points):**
    - Evaluate the trained model on the validation set.
    - Calculate and report classification metrics such as accuracy, loss, precision, recall, F1-score(if applicable).
5. **Hyperparameter Tuning (5 points):**
    - Experiment with hyperparameters (e.g., learning rate, batch size) to improve model performance.
    - Report the best set of hyperparameters and their impact on accuracy in the report.
6. **Presentation and Documentation (15 points):**
    - Include code comments and explanations for clarity in the Jupyter Notebook.
    - Prepare a brief report (word document) summarizing your work and mention your findings, challenges faced, and lessons learned.

Submission:

Create a folder. The folder name should have your SLIIT registration number.

e.g. **ITXXXXXXX**

Rename Jupyter notebook as

ITXXXXXXX_model.ipynb

Rename your brief report (word document) as **ITXXXXXXX_summary**

Place the report (word document) and Jupyter notebook inside the created folder

Finally zip the main folder and upload it to the given link given in the Courseweb'.

Note: You are allowed to upload only 1 folder.

Ensure that your code is well-commented, organized, and the report is clear and concise.

Plagiarism Rules:

If your code/summary matches (more than 50%) with another student's code, all those students whose codes/reports match will be awarded with zero marks without any evaluation.