

STAT 4012 Statistical Principles of Deep Learning with Business Application

Guidelines for course group project 2022-23 Term 2 (Spring)

The purpose of this group project is to provide a hand-on experience for students to try the techniques learned from this course on real dataset. Students are required to choose at least **two neural network models** of the following topics, but not limited to:

1. Comparison of various numerical optimization methods, such as SGD, Momentum, Adam, AdaGrad, and RMSProp.
2. Multi-layer Perceptron (MLP)
3. Convolutional Neural Network (CNN)
4. Recurrent Neural Network such as GRU and LSTM

There are several remarks on this project:

- Students need to demonstrate how the selected techniques are used to solve a real-life problem. The data has to be a real dataset of at least 1000 records. Students may use any dataset downloaded from the internet. (See e.g. <http://finance.yahoo.com/>, <https://archive.ics.uci.edu/ml/datasets.php>, and <https://www.kaggle.com/>, Python library built-in dataset) or Bloomberg Terminal available at the STAT and R labs in LSB. Students are **required** to cite the source of the data and explain the data cleaning process in the report.
- Students are suggested to compare your chosen model (if suitable) with several machine learning tools such as SVM, KNN, Naïve Bayes Classifier, Comonotone-Independence Bayes Classifier (CIBer), Regression learners, Decision Tree, Random Forest, and adopt several suitable model assessment schemes to evaluate the performance your chosen models.
- Students may use the same dataset or two separate datasets to demonstrate the selected techniques.
- The final report contains introduction to your problem of interest and description of your dataset; methods to use; findings and conclusion.
- Students are suggested to use Python **MAINLY** and **R** sometimes in their analysis.
- Students are expected to hand in their group report. The report should be around **20** pages of A4 size (tables and figures are exempted from this 20-page limit). Please also indicate the work (in %) involved by each member in the group; otherwise we assume the equally and evenly spread of efforts among members.
- **Each group of students should give a presentation of around 10-15 mins on Apr. 20th, 2023.**
- **Each group of students should submit a project outline and/or a mid-term report on or before Mar. 13th, 2023 (Monday).**
- Students can form their own group with 4 to 5 students. Choose **one** member to send your group members' Name, SID, and Major to TA on or before **Jan 30th** (Monday).
- Choose **one** member in your group to submit a soft copy of your report, dataset (csv or txt) and Python programs (py or ipynb), related **R** programs (R NOT .RMD).

This project consists of **40%** of your total mark in this course and to be submitted by sending a soft copy to pokhimcheng@link.cuhk.edu.hk on or before **2nd May, 2023** (Monday).

The University places very high importance on honesty in academic work submitted by students, and adopts a policy of zero tolerance on cheating and plagiarism. See the following link for details: <http://www.cuhk.edu.hk/policy/academichonesty/>