Machine Learning
(IT3190E)

Quang Nhat NGUYEN

quang.nguyennhat@hust.edu.vn

Hanoi University of Science and Technology
School of Information and Communication Technology

Academic year 2020-2021

The course's structure

- Number of weeks: 15
  - Lectures: The first 12 weeks
  - Course projects' presentations: The last 3 weeks
- Time and Location
  - Thursday, 08:25-11:45, Building B1, Room 402
- The course's web page

  http://soict.hust.edu.vn/~quangnn/ml/ml_ay2021_s2.htm
- Student meeting
  - **Prior arrangement via e-mail**
  - Department of Information Systems, School of Information and Communication Technology (Building B1, Room 702)

The course's goal

- Knowledge on core concepts in machine learning
- Knowledge on traditional and popular algorithms in machine learning, their advantages and disadvantages, and when to use them
- Introduction of very useful software frameworks and tools
- Experience on design, implementation and performance evaluation of machine learning system
  - Through the team work of the course project

The course's content

- Introduction of Machine learning, Software frameworks and tools
- Performance evaluation of machine learning system
- Supervised learning (Linear regression, Nearest neighbor learning, Probabilistic learning, Decision tree, Artificial neural network, Support vector machine)
- Unsupervised learning (Partition-based clustering, Hierarchical clustering)
- Ensemble learning
- Reinforcement learning

The course's grade

- Course project (**P**): Maximum 10 points
  - Each course project is implemented by a group of **2-3** students
  - Freely select and apply a machine learning approach to solve a real application problem
  - Implement a machine learning system and evaluate the system performance with appropriate datasets
- Final written exam (**E**): Maximum 10 points
- Course grade (**G**)
  - **G** = 0.3 x **P** + 0.7 x **E**

Course project – Proposal

- Freely select a real application problem, machine learning algorithm(s) to solve the problem, and dataset(s) used
- The course project proposal (stored in a .pdf file) must explain **in details**
  - Length of **1 or 2 pages**
  - **Description of the real application problem** to be solved (goal, requirements, use scenario, etc.)
  - Clearly specify **machine learning algorithm(s)** used to solve the problem
  - Clearly explain **the input** and **the output** of the implemented machine learning system, and **the representation of learning examples**
  - Clearly specify **the datasets** to be used
- Each project group must send the proposal to *quang.nguyennhat@hust.edu.vn*/ quangnn@soict.hust.edu.vn **not later than March 21, 2021**
  - **The project proposal** of the group
  - The information of each team member: **Full name, Student code, Email**

Course project – Requirements

- The results of the course project work must be presented

  All the project members must participate in the implementation and presentation

- The deliverables of the course project work include:
  - **The source codes**: stored in a .zip/.rar file
  - **Readme file** that describes in details how to install/compile/run the system (and also the other used software libraries/packages)
  - **The final report** (stored in a .pdf file):
    - Describes the real application problem to be solved
    - The details of the machine learning algorithm(s) and the used dataset(s)
    - The experimental results of the machine learning system's performance on the used dataset(s)
    - The system's main functions, and how to use them
    - The issues/difficulties occurs in the course project's implementation, and how you resolved them
    - The discussions/findings/conclusions, and the improvement proposals for the system in future

Course project – Assessment

- The course project work is assessed by the following criteria:
  - **The complexity/difficulty of the real application problem to be solved**
  - **The quality (correctness and adequacy) of the machine learning approach used to solve the problem**
  - **The quality of the experimental results, and the adequacy of the discussions (opinions, arguments) on the experimental results**
  - The quality of the presentation of the course project work
  - The quality of the report documents
  - The implemented software system (functions, usability, etc.)
- The slides presentation should be inline with your conducted work and the report documents
- **If any source codes/software packages/software tools are used/exploited, then they must be explicitly and fully cited in the report documents and in the presentation slides**

Course materials

- The lecture slides
- Reference books:
  - T. M. Mitchell. *Machine Learning*. McGraw-Hill, 1997.
  - T. Hastie, R. Tibshirani, J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd Edition)*. Springer, 2009.
- Software framework, libraries and tools for machine learning
- Datasets:
  - Kaggle
  - UCI
  - WEKA