Machine Learning

(Học máy – IT3190E)

Khoat Than

School of Information and Communication Technology

Hanoi University of Science and Technology

2022

About the course

- Period: 16 weeks
  - Lectures: 12-13 weeks
  - Project report: 3-4 weeks
- Lecture directory:
  https://users.soict.hust.edu.vn/khoattq/lectures/ML-IT3190E-131679

  https://www.youtube.com/watch?v=jc1wo_8VA1w&list=PLaKukjQCR56ZRh2cAkweftiZCF2sTg11_&index=1
- Time & location:
  - 12:30-15:50 Wednesday
- Question + advice: khoattq@soict.hust.edu.vn
- Join and discuss somethings with us:
  http://www.facebook.com/groups/1578056932500777/

Contents

- Lecture 1: introduction to Machine Learning
- Lecture 2: linear regression
- Lecture 3: classification and kNN
- Lecture 4: random forest
- Lecture 5: neural networks
- Lecture 6: support vector machines
- Lecture 7: clustering with K-means
- Lecture 8: ensemble Learning
- Lecture 10: model assessment & selection
- Lecture 11-12: probabilistic models
- Lecture 13: reinforcement learning

Goals of the course

- Help students to have a good basic background on Machine Learning (ML).
- Identify the main **advantages** and **limitations** of the methods/models in ML.
- Be able to design & implement an ML-based system, and evaluate its performance.

Some technologies/libraries

python™

ANACONDA
Powered by Continuum Analytics®

scikit learn

APACHE Spark™

TensorFlow

PyTorch

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$

Evaluation (đánh giá)

- Attendance and activeness
- Midterm test: **Capstone Project**
- Final exam
  - Online test or Paper-based test
- Overall: Midterm test (40%) + Final exam (60%)

Capstone Project

- Students work in groups, each consists of 3-5 students.
- Each group choose a problem/topic to be solved, datasets to be used, algorithms in ML.
- Each proposal should be precisely described
  - The problem: short description, input, output, data type, future application, …
  - The algorithms or tools, planned to be used
  - Data sets to be used
- **Project registration: before 09/05/2022**
  - Via Google Form (TBA)

Capstone Project: requirements

- The result will be presented in the ending period of this subject.
  Every member is required to contribute to his/her project.
- Project report:
  - **Source code**: save your code into one zip file
  - **Readme.txt**: describes clearly how to setup, compile, and run your code
  - **Written report:**
    - Introduce the problem to be solved, the data sets were used
    - Details about the methods for analyzing data
    - Results of different evaluations, new conclusions/findings, …
    - The main components of your code
    - The difficulties in this project, and your proposed solution
    - …

Capstone Project: evaluation

- The evaluation of each project will be based on
  - The difficulty of the problem of interest
  - The appropriateness & quality of the chosen method/solution
  - The rigor of the empirical evaluation and assessment on the chosen method/solution
  - The quality of the presentation
  - The quality of the written report
- Each project will have 15' for slide presentation & demo
- **If you use some existing libraries/packages/codes, you have to clearly declare your usage in the written report and slide presentation**

Some references

- Lecture slides + Youtube
- Reference books:
  - T. M. Mitchell. *Machine Learning*. McGraw-Hill, 1997.
  - Trevor Hastie, Robert Tibshirani, Jerome Friedman. *The Elements of Statistical Learning*. Springer, 2017.
  - Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT press, 2016.
  - E. Alpaydin. *Introduction to Machine Learning*. The MIT press, 2020.
- Software:
  - Scikit-Learn (http://scikit-learn.org/)
- Data for experiments:
  - UCI repository: http://archive.ics.uci.edu/ml/