HUST

ĐẠI HỌC BÁCH KHOA HÀ NỘI

HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

ONE LOVE. ONE FUTURE.

ĐẠI HỌC
BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Introduction to
Machine Learning and Data Mining

IT3190

Lecture: Course overview

ONE LOVE. ONE FUTURE.

About the course

- Period: 15 weeks
  - Lectures: 11-13 weeks
  - Project report: 2-3 weeks
- Lecture directory:
- Time & location:

- Question + advice:
  - Building B1
- Join and discuss somethings with us:

Contents

- Lecture 1: Introduction to Machine Learning & Data Mining
- Lecture 2: Data crawling and pre-processing
- Lecture 3: Linear regression
- Lecture 4+5: Clustering
- Lecture 6: Decision tree and Random forest
- Lecture 7: Neural networks
- Lecture 8: Support vector machines
- Lecture 9: Performance evaluation
- Lecture 10: Probabilistic models
- Lecture 11: Basics of data mining
- Lecture 12: Association rule mining
- Lecture 13: Regularization and advanced topics

Goals of the course

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

Some technologies/libraries

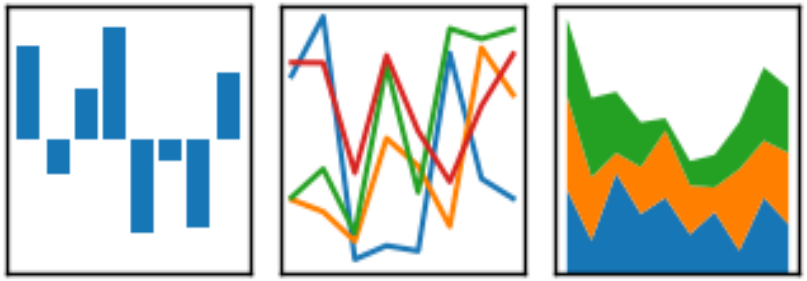
python™

ANACONDA
Powered by Continuum Analytics®

TensorFlow

pandas

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



Evaluation (đánh giá)

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

Capstone Project

- Students work in groups.
- Each group choose a problem/topic to be solved, datasets to be used, algorithms in ML or DM.
- 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:**
  - 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
- 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.
  - Jiawei Han, Micheline Kamber, Jian Pei. *Data Mining: Concepts and Techniques* (3rd Edition). Morgan Kaufmann, 2011.
- Software:
  - WEKA (http://www.cs.waikato.ac.nz/ml/weka/)
  - Scikit-Learn (http://scikit-learn.org/)
- Data for experiments: UCI repository: http://archive.ics.uci.edu/ml/

HUST

THANK YOU !