

Machine Learning

Dae-Ki Kang

A decorative graphic consisting of several horizontal lines of varying lengths and colors (teal, light blue, white) extending from the right side of the slide towards the center.

Machine Learning is

- a scientific discipline that is concerned with the design and development of algorithms that allow computers to evolve behaviors based on empirical data, such as from sensor data or databases.
- A major focus of machine learning research is to automatically learn to recognize complex patterns and make intelligent decisions based on data;
- the difficulty lies in the fact that the set of all possible behaviors given all possible inputs is too complex to describe generally in programming languages, so that in effect programs must automatically describe programs.

The goal of this course

- learn and research on the the design, development and analysis of algorithms that enable computers to acquire knowledge from empirical data ranging from database to sensor data.
- We will discuss algorithms including naive Bayes classifiers, decision tree classifiers, artificial neural networks, support vector machines, computational learning theory, ensemble classifiers, probabilistic graphical models, multiple instance learning, relational learning, expectation maximization.

The applications we will deal with

- Particularly in this semester, we will focus on
 - the methods for applying taxonomy for concise classifiers,
 - optimizing NN/SVM/DT ensemble,
 - heterogeneous ensemble,
 - messy genetic algorithm,
 - GA based decision tree construction,
 - intrusion detection,
 - social network services, etc.
- You will choose one of the above as a term project

Evaluation

- Presentation (Mid term & 8 times)
- Term project (Final term)
- You have to write a technical report, of which the format follows Elsevier journal formats.

Lab assignment

- The purpose of lab experiments is understand each algorithms and how the algorithm is applied to real world problems.
- The lab experiments include
 - naive Bayes classifiers,
 - decision tree classifiers,
 - artificial neural networks,
 - support vector machines,
 - and ensemble classifiers such as AdaBoost, Bagging, etc.

Lab experiments

- We will use WEKA for experiments on Windows based PC.
- For evaluation, each student needs to submit lab report.

Overall schedule

- (1) Introduction
- (2) naive Bayes classifiers
- (3) decision tree classifiers
- (4) artificial neural networks
- (5) support vector machines
- (6) computational learning theory
- (7) ensemble classifiers
- (8) probabilistic graphical models
- (9) multiple instance learning
- (10) relational learning
- (11) expectation maximization
- (12) applying taxonomy for concise classifiers
- (13) optimizing NN/SVM/DT ensemble
- (14) messy genetic algorithm
- (15) GA based decision tree construction
- (16) intrusion detection