Machine Learning

Won Kim

Enrollment in this Course Is Limited.

- to Students Majoring in AI Major or SW Major (within AI-SW학부)
- And to Students Second-Majoring in AI or SW (within AI-SW학부) and who can program in Python.
- All other students who have registered for this course, please drop the course right away.
  - There are too many students, and we cannot take any additional students.

Course Objectives

- Learn many of the widely-used machine learning algorithms, and basic math behind them.
- Learn to apply some in the data science process.
- Learn to use Python and Python-based platforms to carry out a data science project.
- Prepare to take the deep learning course.

Topics for the Course (1/4)
* (new algorithms to learn: in red letters)

- Classification
  - decision trees
  - k-nearest neighbors
  - decision trees (closer look)
  - logistic regression
  - support vector machine
- Clustering
  - k-means
  - HAC – max, mean
  - clustering quality measures
  - K-Medoids
  - EM
  - DBSCAN

Topics for the Course (2/4)

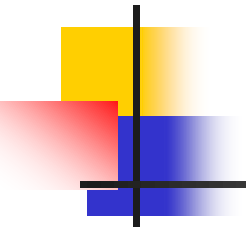
- Correlation and Regression
  - linear, polynomial, multiple regression
  - loss function and errors
  - decision tree regression
- optimization
  - regularization
  - gradient descent
- Other Algorithms
  - Naïve Bayesian
  - Markov chain
  - association rules discovery
  - reinforcement learning

Topics for the Course (3/4)

- Neural Network for Deep Learning (brief introduction)
  - perceptron
  - multi-layer perceptron
  - CNN (not to be covered)
  - RNN, LSTM (not to be covered)
- Text Processing
  - text preprocessing
  - vector space model
  - n-gram
  - word embedding

Topics for the Course (4/4)

- Evaluation and Tuning
  - cross validation
  - bagging
  - confusion matrix
  - ROC
- Ensemble Learning
  - bagging
  - random forest
  - boosting
- Inferential Statistics
  - chi-square test
  - t-test
  - central limit theorem



Course Elements

- 10+ Lectures
- 4 Labs
- 2 Exams
- 2 Quizzes
- 4 Programming Homework
- 4-6 Written Homework
- Term Project and Presentation

Course Grading Policy (in points)

- Exams
  - mid-term: 200
  - final: 300
- Term Project: 100
- Labs: 100
- Programming Homework: 100
- Homework: 100
- Attendance 100

About the Course PPT

- Created by studying over 100 PPT/PDF files and blogs/writings on the Internet.
- Those that were deemed especially worthy have been included in Acknowledgments.
- Some of them was used with minor changes.
- Many others were used to a small extent.

Note (1/2)

- From now, our School of Computing (SW Department + AI Department) will adopt self-directed learning methods in all computer-science courses (except 1st year courses).
- Professors will become more of coaches than lecturers.
- We are doing this because life is a continuous learning process, and we believe students will find, after graduation, training in school on self-directed learning very valuable.
- There are two key elements to the self-directed learning.
  - MOOC and Active Learning

MOOC (massively online open courseware)

- At least 3 weeks' classes will be conducted using MOOCs.
- Students can take the classes from anywhere any time (within one week of the posting of the MOOC).
- There will be homework for each MOOC class (due in one week).
- Students can ask questions online and receive answers within 24 hours.
- This method has been used very successfully for 5 semesters for all SW ELITE courses and SW Basic courses for all non-Computer Science students in Gachon University.

Attendance Check for MOOC Classes

- We do not want to bother checking if you attended the MOOC class.
- We will simply take “not submitting homework solution in time” as “absence” from the class.

Active Learning

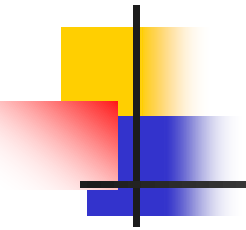
- At least 2 weeks' classes will be conducted using an active learning method.
- The instructor will give a brief introduction to one or more topics, and the students will learn them in depth on their own, and then submit a report and present using PPT in the next class.
- ** This learning model has been used successfully for several years in the "Software Industry Seminar" course.

SLAM (Software Learning and Assessment Mini-Courses) Platform

- Our School of Computing will build a self-directed and adaptive learning platform (named SLAM).
  - (The plan is to have a working platform by February 2022.)
- Students will be asked to submit good learning materials found in their active learning, and selected materials will be uploaded to SLAM for other students to use later.
- We envision SLAM to be a symbol of Gachon University's education excellence, and strongly encourage all SW/AI major students to contribute to its building.
- The School will reward strong contributions to the building of SLAM (both the software and contents).

Machine Learning MOOC Topics

- Topics
  - Classification algorithms: logistic regression, support vector machines
  - Optimization (math): regularization, gradient descent
  - Inferential statistics: chi-square test, t-test, central limit theorem
  - Text processing: vector model, TF-IDF, Word Embedding
  - Clustering: quality measures, DBSCAN
  - Big Data platforms
  - Introduction to artificial neural networks



End of Overview