Machine Learning

Applied Computational Intelligence, sem. 1

www.cs.ubbcluj.ro/~gabis/ml

I. Aims of the activity

1. To introduce the fundamental principles, techniques, and applications of Machine Learning.
2. To cover the principles, design and implementation of learning programs which improve their performance on some set of tasks by experience.
3. To offer a broad understanding of fundamental machine learning algorithms and their use in data-driven knowledge discovery.
4. To offer an understanding of the current state of the art in machine learning to conduct original research in machine learning.

II. Specific competencies acquired

Professional competencies

1. Understanding the concepts, methods and models used in Machine Learning.
2. Understanding the principles, design, implementation and validation of learning systems.
3. Learning to conduct incipient original research in machine learning.

Transversal competencies

1. The ability to apply machine learning techniques in solving real world problems.
2. Responsible execution of lab assignments, research and practical reports.
3. Application of efficient and rigorous working rules.
4. Manifest responsible attitudes toward the scientific and didactic fields.
5. Respecting the professional and ethical principles.

III. Course content

1.Introduction in Machine Learning. Statistical Foundations.
2. Decision Tree learning.
3. Artificial Neural Networks.
4. Support Vector Machines.
5. Bayesian Learning.
6. Instance based learning.
7. Unsupervised learning.
8. Reinforcement Learning.

IV. Bibliography

Mandatory bibliography – available at www.cs.ubbcluj.ro/~gabis/ml/ml-books

1. Mitchell, T., Machine Learning, McGraw Hill, 1997
2. Nillson, N., Introduction to Machine Learning, Stanford University, 1996

Optional bibliography

1. Deep learning textbook – available at **http://www.deeplearningbook.org/**
2. Sutton, R.S., Barto, A.G., Reinforcement learning, The MIT Press Cambridge, Massachusetts, London, England, 1998
3. Cristiani, N., Support Vector and Kernel Machines, BIOwulf Technologies, 2001
3. Russell, J.S, Norvig, P., Artificial Intelligence- A Modern Approach, Prentice- Hall, Inc., New Jersey, 1995

Journals

1. *Journal of Machine Learning Research* is a freely available ISI journal - http://jmlr.csail.mit.edu/
2. *Machine Learning* - http://www.springer.com/computer/artificial/journal/10994
3. *Neural Computation* - http://www.mitpressjournals.org/loi/neco
4. *Neural networks* - http://www.elsevier.com/wps/find/journaldescription.cws_home/841/description#description
5. *IEEE Transactions on Pattern Analysis and Machine Intelligence* - *http://www2.computer.org/portal/web/tpami/*
6. *Bioinformatics* is a journal focusing on analysing biological data - http://bioinformatics.oxfordjournals.org/
7. others

….

Papers

Ross Quinlan's papers - http://www.rulequest.com/Personal/

Conferences

International Conference on Machine Learning, Neural Information Processing Systems, Conference on Computational Learning, International Conference on Knowledge Discovery and Data Mining, European Conference on Machine Learning, and others.

ML repository

http://archive.ics.uci.edu/ml/

SBSE repositories

http://promise.site.uottawa.ca/SERepository/

http://openscience.us/repo/

Other data sets

https://vincentarelbundock.github.io/Rdatasets/datasets.html

V. Activity and grading

For the lab hours, the group will be divided in **2 subgroups – S1** (e.g. the first half of the group)**, S2** (e.g. the second half of the group)

Each student should prepare and present the following:

(1) A theoretical research report on a learning technique, based on some recent research papers.

a) a written paper of about 10 pages

b) an oral presentation

- a one-page outline of the presentation

You should present a survey on some recent research results on the considered topic.

The paper will contain theoretical considerations on the selected topic and compulsory sections regarding:

1. existing or possible applications of the selected topic
2. advantages and disadvantages of the selected approach (here you can present your own opinions)

(2) Two practical projects (software) must be completed for the lab activity.

Requirements

- The first project will be developed using an open source ML software. The project will have to comparatively demonstrate the use of two ML techniques for some specific tasks.

1. Python libraries (Scikit-learn, Keras, etc)
2. WEKA http://www.cs.waikato.ac.nz/ml/weka/
3. RapidMiner http://rapid-i.com
4. Orange http://www.ailab.si/orange/
5. ROCKIT http://xray.bsd.uchicago.edu/krl/KRL_ROC/software_index.htm
6. SVM software http://www.support-vector-machines.org/SVM_soft.html

7. MATLAB
8. others

The first project will include:

(a). a description of the programming software used, including used features (doc);
(b). problem definition (doc);
(c). comments about the solution (doc).

Deadlines

- **For subgroup S1**

Lab 2 (Week 3) – installation of ML software + components (a)+(b) project 1
Lab 4 (Week 7) – component (c) project 1 + project 1 demonstration

- **For subgroup S2**

Lab 3 (Week 5) – installation of ML software + components (a)+(b) project 1
Lab 5 (Week 9) – component (c) project 1 + project 1 demonstration

- The second project will be fully implemented, without using existing ML environments. The project will have to demonstrate the use of a ML technique for a specific task.

The second project will include:

(a). problem definition (doc);
(b). comments about the solution (problem analysis) (doc);
(c). a short design documentation (doc);
(d). the electronic version of the source code, test files and any other files required to test the project.

Deadlines

- **For subgroup S1**

Lab 4 (Week 7) – components (a) + (b) project 2
Lab 6 (Week 11) –components (c) + (d) project 2

- **For subgroup S2**

Lab 5 (Week 9) – components (a) + (b) project 2
Lab 7 (Week 13) –components (c) + (d) project 2 + project demonstration

For both subgroups, the demonstration of the second project is due on the day of the written exam (examinations session).

(3) The final exam, to take place in the examinations session.

Identical projects will NOT be considered.

The delay of one lab in completing an assignment will be penalized with 1 point.

The final grade will be computed as follows:

10% Class attendance
20% Theoretical research report (written and presented)
20% Practical project 1 (documented and demonstrated)
20% Practical project 2 (implemented, documented and demonstrated)
30% Final exam (written paper in exams session)

A minimal final grade of 5 is required to pass the discipline.

Remarks

- The title of the theoretical research report and the presentation date must be chosen by end of **week 5**.
- The time planning for the research report can be consulted at www.cs.ubbcluj.ro/~gabis/ml.
- The grade at the theoretical report is composed by considering the following:
    - the papers should fulfill the requirements of a research paper:
        - suggestive title corresponding to the contents;
        - about 10 lines abstract;
        - introductory section, detailing the purpose of the paper;
        - a section integrating the topic of the paper in the general field;
        - a few main sections, according to your topic;
        - concluding remarks and further work section;
        - bibliography of 5 to 10 titles; the bibliography entries should be written correctly and completely; all the bibliography items must be cited in the text;
    - the one-page outline must correspond both to the written text and to the oral presentation, and should be self-explanatory for the intended audience;
    - the oral presentation itself, including the discussions.

- The grading is done as follows:
    - 1.0 p = abstract and introduction;
    - 3.0 p = structure of main sections;
    - 1.0 p = concluding remarks;
    - 1.0 p = bibliography accuracy;
    - 1.0 p = one-page outline;
    - 3.0 p = quality of oral presentation + discussions.

- All the written materials will have to be provided by e-mail at least one day in advance of the presentation date.

VI. Time planning

- During the lecture hours on weeks 1-11 you can attend lectures on topics from Section III.
- The theoretical report will be allocated about 15 minutes (questions and discussions included), and will be presented during the **lab hours on weeks 11,13 and lecture hours on weeks 12, 13, 14.**

VII. Retake session

In the retake session, all activities can be reconsidered, excepting the oral presentations (the maximum grade for a report is 7).

VIII. Possible topics for the Theoretical Report (not an exhaustive list)

The topics below are suggestions. You may choose other machine learning related topics.

1. Artificial neural networks
2. Recurrent neural networks
3. Time delay neural networks
4. Long-short term memory networks
5. Deep learning
6. Convolutional neural networks
7. Self organizing maps
8. Hebbian learning
9. Unsupervised learning in recurrent neural networks
10. Semi-supervised learning
11. Radial Basis Function networks
12. Decision Trees
13. Bayesian learning
14. Machine learning in bioinformatics
15. Machine learning in software engineering
16. Belief network learning
17. Instance based learning
18. Case based reasoning (learning)
19. Inductive logic programming
20. Boosting algorithms
21. Deep Reinforcement Learning
22. Temporal difference learning
23. Q-learning
24. Autoencoders
25. Adaptive clustering
26. Hierarchical clustering
27. Partitional clustering
28. Support vector machines
29. Kernel methods in machine learning
30. Association rule mining
31. Hidden Markov Models

etc