Machine Learning

Applied Computational Intelligence, High Performance Computing
(sem. 1)

Prof. dr. Czibula Gabriela
Contact: gabriela.czibula[at]ubbcluj.ro

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 in order 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. ML activities

All activities require physical participation.

- The course materials (lecture notes, books, bibliographic material) will be available in **General/Files/Class materials** before the lecture hours.
- The lecture hours on Weeks 8-14 and lab hours on Weeks 9-12 will be allocated for research reports presentations, according to the planning available in **General/Files/Class materials/TimePlannigReport.docx** (face-to-face presentations).
- The lab assignments must be submitted on MSTeams (through the corresponding Assignments) by the end of the lab class.
- The research reports must be submitted on MSTeams (through the corresponding Assignments) at least 24h in advance of the presentation date.
- **Planning**

  Lectures

  Weeks 8-14 – student presentations (research reports)

  Labs

  Labs 2-3 (Weeks 3/4, 5/6) **-** individual discussions (assignments, research topics, etc)
  Lab 4 (Weeks 7/8) – 1st software project presentation (face-to-face)
  Lab 5, 6 (Weeks 9/10, 11/12) – research reports presentations (face-to-face) (**TimePlannigReport.docx**)
  Lab 7 (Weeks 13/14) - 2nd software project presentation (face-to-face)

V. Bibliography

1. Mitchell, T., *Machine Learning*, McGraw Hill, 1997 (available at www.cs.ubbcluj.ro/~gabis/ml/ml-books)
2. Nillson, N., *Introduction to Machine Learning*, Stanford University, 1996 (available at www.cs.ubbcluj.ro/~gabis/ml/ml-books)
3. Sutton, R.S., Barto, A.G., *Reinforcement learning*, The MIT Press Cambridge, Massachusetts, London, England, 1998 (http://incompleteideas.net/book/the-book.html)
4. Ian Goodfellow, Yoshua Bengio, Aaron Courville, *Deep Learning*, MIT Press, 2016 (online edition at http://www.deeplearningbook.org/)
5. Aston Zhang, Zachary C. Lipton, Mu Li, and Alexander J. Smola, *Dive into Deep Learning*, 2020 (http://d2l.ai/)

6. Li Deng and Dong Yu, *Deep Learning. Methods and Applications*, Foundations and Trends® in Signal Processing, Volume 7 Issues 3-4, 2014 (https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/DeepLearning-NowPublishing-Vol7-SIG-039.pdf)
7. Cristiani, N., *Support Vector and Kernel Machines*, BIOwulf Technologies, 2001

Journals

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

….

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.

Resources

- ML repository
  http://archive.ics.uci.edu/ml/
- RL repository
  http://www-all.cs.umass.edu/rlr/
- RL resources
  - https://aikorea.org/awesome-rl/
  - http://busoniu.net/repository.php
- SBSE repositories
  - http://promise.site.uottawa.ca/SERepository/
  - http://openscience.us/repo/
- Object recognition data sets
  - https://cvgl.stanford.edu/teaching/cs231a_winter1314/lectures/datasets.pdf
- Computer vision image datasets
  - http://www.cs.utexas.edu/~grauman/courses/spring2008/datasets.htm

- Other data sets
  https://vincentarelbundock.github.io/Rdatasets/datasets.html

VII. Activity

Each student should prepare and present the following:

(1) A theoretical **research report** on an ML-based topic, based on some recent research papers.

a) a written paper of 9-10 pages using the LaTeX template available in the folder Files/Class Materials/LaTeX template (research paper)

- Overleaf – online LaTeX editor.
- The use of the LaTeX template is compulsory.

b) an oral presentation

- a one-page outline of the presentation

Requirements

The research report will present a survey on some recent research results/research papers (about 5-10 titles) on the chosen topic and not an overview of the topic.

The grade of the research report is composed by considering the following:

- the papers should fulfill the requirements of a research paper:
  - suggestive title corresponding to the content;
  - 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;
  - discussion and SWOT analysis;
  - 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.

  E.g.,

  - Manevitz, L., Yousef, M., 2007. One-class document classification via neural networks. Neurocomputing 70, 1466–1481 (**@article** bibtex entry)
  - Le, Q., Mikolov, T., 2014. Distributed representations of sentences and documents, in: Proceedings of the 31st International Conference on Machine Learning, Beijing, China, 2014, pp. 1188–1196 (**@inproceedings** bibtex entry)
  - Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press (**@book** bibtex entry)

- Google. Google Code Jam Competition. https://codingcompetitions.withgoogle.com/codejam. Online; accessed 15 September 2021 (**@misc** bibtex entry)

- the one-page outline must correspond both to the written text and to the oral presentation, and should be self-explanatory;
- the oral presentation itself, including the discussions.

Deadline

Week 4 – a one page document containing:

- a brief description of the approached topic and its importance (relevance) (3-4 paragraphs);
- the title of the theoretical report;
- bibliographical references (at least 5);
- the presentation slot (Weeks 8-14 lecture hours or Weeks 9-12 lab hours);
- represents 5% from the grading of the research report.

– About 12 minutes will be allocated for the presentation of the research report + 3 minutes for questions and discussions.
– The written paper will have to be submitted at least 24h in advance of the presentation date.
  - The research reports will be checked against plagiarism.
  - The maximum allowed similarity score is 30%.
– The time planning for the research report and details about the oral presentations can be found in **General/Files/Class materials/TimePlannigReport.docx**.

(2) Two **software projects** must be completed for the lab activity. The projects must demonstrate the use of ML for some specific tasks.

Requirements

A. 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 task.

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:

(1) a description of the programming framework used, including used features (doc);

(2) a document (doc) containing:
- problem definition (the problem statement in natural language, **what** should be solved);
- problem specification (input/data and preconditions, output/results and postconditions);
- specification of the learning task (Task, Performance, Experience);
- target function to be learned.

(3) comments about the solutions (learning hypotheses, details about the ML models/architectures) and experimental results (doc).
- cross-validation must be used for performance evaluation;
- comparative analysis of the results.

(4) Project demonstration (during the lab hours)

Deadlines for the 1st software project

Lab 2 – installation of ML software + component (1) project 1
Lab 3 – component (2) project 1
Lab 4 – components (3) + (4) project 1

B. **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:

(1) problem definition and used ML technique (doc);
- problem definition;
- problem specification (input/data and preconditions, output/results and postconditions);
- specification of the learning task (Task, Performance, Experience);
- brief description of the ML technique.

(2) comments about the solution (problem analysis) (doc);
- description and analysis of the used data;
- description and analysis of the features used in learning
  - correlation, independence;
- description and analysis of the proposed solution.

(3) a design documentation (doc);
- the design of the learning task (target function to be learned, representation of the learned function, learning algorithm, learning hypothesis);
- the design of the application.

(4) experimental results and their analysis (doc);
- for evaluating the performance of the ML model
  - cross-validation must be used for performance evaluation;

- a statistical analysis of the results must be provided (including confidence intervals).
- evaluation measures
  - **for classification**: accuracy, precision, recall, sensitivity, f-measure, AUC (Area under the ROC curve), AUPRC;
  - **for regression**: MAE (Mean of Absolute Errors), RMSE (Root Mean Squared Error), NRMSE (Normalized Root Mean Squared Error).

(5) the electronic version of the source code (including the data sets used for demonstration and the executable file)

(6) project demonstration (during the lab hours)

Deadlines for the 2nd software project

Lab 5 – component (1) project 2
Lab 6 – components (2) + (3) project 2
Lab 7 – components (4) + (5) project 2

VIII. Grading

- Identical projects will NOT be considered.
- The grading of the research report 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.
- The grade for a software project is computed as the average of the grades received for the required written documentations.
- The delay of one week in providing the document with the theoretical report topic, title and presentation date will be penalized with **1 point**.
- The delay of one week in completing a lab assignment will be penalized with **1 point**.
- The failure to deliver the presentation at the due date will lead to a grade penalty of **1 point** for each delayed week.

The final grade is computed as follows:

50% Theoretical research report (written and presented)
15% Software project 1 (documented and demonstrated)
35% Software project 2 (implemented, documented, and demonstrated)

The activities (components of the final grade) **are not compulsory**.

If **components (4) + (5) + (6)** for the 2nd software project will not be delivered at **Lab7**, they may be delivered in the **examination session**, with a penalty of **2 points** for the delay**.**

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

IX. Semester and final grades

The grades received during the semester, as well as the final grades for the **Machine learning** course are available at the following link

https://docs.google.com/spreadsheets/d/e/2PACX-1vRmd2IxLscx6SD9XXbgqvpQiqOS-BoW5ZCT07TrdowgGUuAm7SHJ6ch7Uq2WWTFgPUY0vC_EYVqK6Ff/pubhtml?gid=1675021479&single=true

The code used for displaying the grades is the **unique identification code** (from Academic Info).

X. Retake session

- The activities (theoretical report/software projects) can be graded in the retake session, excepting the oral presentations, but **ONLY** if at least one activity (theoretical report, 1st software project, 2nd software project) has been completed during the semester.
  - The maximum grade for a report/project in the retake session is **6**.
- No report or software project can be resubmitted for grade increase in the retake session.

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

The topics below are suggestions only. Please feel free to choose other ML related topics.

1. Artificial neural networks
2. Recurrent neural networks
3. Time delay neural networks
4. Long-short term memory networks
5. Self organizing maps
6. Hebbian learning
7. Semi-supervised learning
8. Radial Basis Function networks
9. Decision Trees (fuzzy, lazy, etc)
10. Bayesian learning
11. Machine learning in bioinformatics
12. Machine learning in software engineering (*search-based software engineering*)
13. Instance based learning
14. Case based reasoning (learning)
15. Inductive logic programming
16. Boosting algorithms (Adaboost, Epsilon Boost, Gradient boosting, etc)
17. Bagging algorithms
18. (Deep) Q-learning
19. Adaptive clustering
20. Hierarchical clustering

21. Partitional clustering
22. (Deep) Support vector machines
23. Kernel methods in machine learning
24. Association rule mining
25. Hidden Markov Models
26. Convolutional neural networks
27. One-shot learning
28. Deep Reinforcement Learning
29. Autoencoders
30. Belief network learning
31. Generative models
32. Siamese networks
33. Contrastive learning
34. Representation learning

a.s.o