Project

Due Date: **Proposal Submission by March 15, 11:30 pm**
Report Submission by April 20, 11:30 pm
Assessment: 25% **of the total course mark**

Description:

The aim of this project is for you to explore in more depth a machine learning topic of your interest. You will have to study the related literature, then write a report and make a presentation on the chosen topic. It is allowed for more students to choose the same topic. In this case, the presentations should not have overlapped material, but the reports may have overlaps.

Following are some suggested topics along with some suggested study material.

1. Handling class imbalance [1].
2. Methods for feature extraction and feature selection [2, Chap. 6], [3, 4].
3. Methods to integrate prior knowledge into learning [5].
4. Challenges and risks of using machine learning (data privacy and security, biased data and bias in making decisions, transparency, explainable ML) [6, Chap. 7], [7, 8, 9, 10, 11].
5. Dealing with label noise [12].
6. Adversarial Examples [13].
7. Optimization methods for machine learning [14].
8. Multiobjective reinforcement learning [15].
9. Optimal control using reinforcement learning [16].
10. Adaptive dynamic programming for control [17].
11. Nearest neighbours classification with different numbers of nearest neighbours [18].
12. Scalable nearest neighbours algorithms for high dimensional data [19].

Topics 1 and 2 are important ML topics. I would like to have two people covering each of them.

Instead of exploring a general ML topic, you could pick a particular ML problem and explore the solution methods proposed in the literature. Some examples of problems are provided next.

- ◇ Machine learning for clinical outcome prediction [20].
- ◇ Machine learning for predicting seizures using EEG signals [21].
- ◇ Fake news detection [22].
- ◇ Deep facial expression recognition [23].
- ◇ Deep learning for human affect recognition [24].

◇ Chronic kidney disease diagnosis [25].

You may select a different topic or problem than those shown above. One method to find interesting problems is to look for (popular) papers published in the IEEE Transactions related to your area of interest or in other journals of high reputation in your area.

How many papers to study? Some guidelines:

◇ One survey paper of about 20 pages (double column) and at least one other paper on a particular method included in the survey.

◇ One survey paper of about 10-14 pages (double column) and at least two other papers on a particular method included in the survey.

◇ Three or four non-survey papers.

PROPOSAL:

- Due date: March 15, 11:30 pm.
- The proposal should be submitted **as a pdf file** in the designated drop box.
- It should contain a brief description of the problem/topic and a list of papers you intend to study.
- Before submitting the proposal double-check with me to see if I approve (send me a message with the name of the topic in the Teams chat or via email). If too many people choose the same topic I might not approve this choice for all of them. The first who announces me will have priority. In addition, if I approve the same topic for more students, they have to coordinate their work to make sure that their presentations will cover different material.

REPORT:

- Due date: April 20, 11:30 pm.
- The report should be submitted **as a pdf file** in the designated drop box.
- No longer than 20 pages, single-column, double-spaced, font size 11.

PRESENTATION:

- The presentation should be no longer than 20 minutes and it will be followed by approximately 10 minutes of questions from the audience.
- If you present a survey, you will not be able to address all works/techniques in detail, but you should pick at least one to discuss in more detail.
- The presentations will be delivered in four sessions, each including six presentations. They could be spread over four days or only over two days.
- The tentative dates and times for the presentations are April 20, 21, from 9:30 am to 1:00 pm and from 2:00 to 5:30 pm. The final dates will be established during a class discussion.
- Each student must attend at least half of the presentations and ask at least three questions (in three different presentations).

- Participation by asking good questions will be rewarded. Each of the first three questions will be rewarded with 1% of the course mark. **Bonus:** each of the following four good questions will be rewarded with 0.5% of the course mark.
- Attendance will also be rewarded as follows: a) 0.2% of the course mark for each presentation from the first 12 presentations you attend, b) **bonus** 0.2% of the course mark for attending each additional presentation.
- The delivery format will be in person if there will be rooms available.

Bonus:

Up to 5% of the course mark may be awarded if you perform simulations and/or experimental comparisons between the discussed techniques. The amount of the bonus awarded depends on how extensive your experiments are or how involved the implementation is.

All the papers in the reference list are available freely online for McMaster users. The links will be posted on avenue.

References

[1] H. He, and E. A. Garcia, "Learning from Imbalanced Data", *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, Sept. 2009, pp. 1263-1284.

[2] E. Alpaydin, *Introduction to Machine Learning*, 4th Ed., The MIT Press, 2020 (available at McMaster library).

[3] J. C. Ang, A. Mirzal, H. Haron, and Haza Nuzly Abdull Hamed, "Supervised, Unsupervised, and Semi-Supervised Feature Selection: A Review on Gene Selection", *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 13, no. 5, Sept./Oct. 2016, pp. 971-989.

[4] R. Jenke, A. Peer, and M. Buss, "Feature Extraction and Selection for Emotion Recognition from EEG", *IEEE Transactions on Affective Computing*, vol. 5, no. 3, July-Sept. 2014, pp. 327-339.

[5] L. von Rueden, S. Mayer, K.a Beckh, B. Georgiev, S. Giesselbach, R. Heese, B. Kirsch, J. Pfrommer, A. Pick, R. Ramamurthy, M. Walczak, J. Garcke, C. Bauckhage, and J. Schuecker, "Informed Machine Learning – A Taxonomy and Survey of Integrating Prior Knowledge into Learning Systems", to appear in *IEEE Transactions on Knowledge and Data Engineering*.

[6] E. Alpaydin, *Machine Learning*, e-book, The MIT Press, 2021, DOI: https://doi-org.libaccess.lib.mcmaster.ca/10.7551/mitpress/13811.001.0001 (free for McMaster users).

[7] J. Zhou and L. Schiebinger, "Design AI so that It's Fair", *Nature*, vol. 559, pp. 324-326, July 2018.

[8] M. Hardt, E. Price, and N. Srebro, "Equality of Opportunity in Supervised Learning", 30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

[9] A. van Wynsberghe, "Sustainable AI: AI for sustainability and the sustainability of AI", *AI and Ethics* (2021) 1:213–218, https://doi.org/10.1007/s43681-021-00043-6.

[10] E. Tjoa and C. Guan, "A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI", *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no 11, Nov. 2021, pp. 4793-4813.

[11] X.-H. Li, C. C. Cao, Y. Shi, W. Bai, H. Gao, L. Qiu, C. Wang, Y.n Gao, S. Zhang, X. Xue, and L. Chen, "A Survey of Data-Driven and Knowledge-Aware eXplainable AI", *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 1, Jan. 2022.

[12] B. Frénay and M. Verleysen, "Classification in the Presence of Label Noise: a Survey", *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 5, May 2014, pp. 845-869.

[13] X. Yuan, P. He, Q. Zhu, and X. Li, "Adversarial Examples: Attacks and Defenses for Deep Learning", *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 9, Sept. 2019, pp. 2805-2824.

[14] S. Sun, Z. Cao, H. Zhu, and J. Zhao, "A Survey of Optimization Methods From a Machine Learning Perspective", *IEEE Transactions on Cybernetics*, vol. 50, no. 8, Aug. 2020, pp. 3668-3681.

[15] C. Liu, X. Xu, and D. Hu, "Multiobjective Reinforcement Learning: A Comprehensive Overview", *IEEE Transactions on Systems, Man, and Cybernatic: Systems*, vol. 45, no. 3, Mar. 2015, pp. 385-398.

[16] B. Kiumarsi, K. G. Vamvoudakis, H. Modares, and F. L. Lewis, "Optimal and Autonomous Control Using Reinforcement Learning: A Survey", *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 6, June 2018, pp. 2042-2062.

[17] D. Liu, S. Xue, B. Zhao, B. Luo, and Q. Wei, "Adaptive Dynamic Programming for Control: A Survey and Recent Advances", *IEEE Transactions on Systems, Man, and Cybernatic: Systems*, vol. 51, no. 1, Jan. 2021, pp. 142-160.

[18] S. Zhang, X. Li, M. Zong, X. Zhu, and R. Wang, "Efficient kNN Classification With Different Numbers of Nearest Neighbors", *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 5, May 2018, pp. 1774-1785.

[19] M. Muja and D. G. Lowe, "Scalable Nearest Neighbor Algorithms for High Dimensional Data", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 11, Nov. 2014, pp. 2227- 2240.

[20] F. Shamout, T. Zhu, and D. A. Clifton, "Machine Learning for Clinical Outcome Prediction", *IEEE Reviews in Biomedical Engineering*, vol. 14, 2021, pp. 116-126.

[21] K. Rasheed, A. Qayyum, J. Qadir, S. Sivathamboo, P. Kwan, L. Kuhlmann, T. O'Brien, and A. Razi, "Machine Learning for Predicting Epileptic Seizures Using EEG Signals: A Review", *IEEE Reviews in Biomedical Engineering*, vol. 14, 2021, pp. 139-155.

[22] P. K. Verma, P. Agrawal, I. Amorim, and R. Prodan, "WELFake: Word Embedding Over Linguistic Features for Fake News Detection", *IEEE Transactions on Computational Social Systems*, vol. 8, no. 4, Aug. 2021, pp. 881-893.

[23] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey", to appear in *IEEE Transactions on Affective Computing*.

[24] P. V. Rouast, M. T. P. Adam, and R. Chiong, "Deep Learning for Human Affect Recognition: Insights and New Developments", *IEEE Transactions on Affective Computing*, vol. 12, no. 2, Apr.-June 2021, pp. 524-543.

[25] A. Ogunleye and Q.-G. Wang, "XGBoost Model for Chronic Kidney Disease Diagnosis", *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 17, no. 6, Nov./Dec. 2020, pp. 2131-2140.