instructor: Dr. Sandra Zilles, zilles@cs.uregina.ca

***TENTATIVE*** TOPIC OUTLINE:

(0) Organizational issues, introduction to machine learning, math background (approx. 1 week)

(1) PAC learning (approx. 3 weeks)
  1.1 Formal learning models in general
  1.2 Introduction to PAC learning
  1.3 PAC learning in the realizable case
  1.4 Agnostic PAC learning
  1.5 Characterization of PAC-learnability

(2) Ensemble methods (approx. 1 week)
  2.1 Ensemble methods in classification
  2.2 AdaBoost and variants
  2.3 Ensemble methods in clustering

(3) Regression (approx. 1 week)
  3.1 Optimization approach to linear regression
  3.2 Probabilistic approach to linear regression
  3.3 Non-linear regression
  3.4 Regularized regression

(4) Support vector machines (approx. 1 week)
  4.1 Classification with linear SVMs
  4.2 Classification with kernel SVMs

(5) Learning decision trees (approx. 2 weeks)
  5.1 Decision trees as classifiers
  5.2 Top-down induction of decision trees
  5.3 Application example from bioinformatics

(6) Measuring classifier performance (approx. 1 week)

(7) Student presentations (approx. 1 week)

first class: January 15
last class: April 09

GRADING:

20% assignments (due dates as shown on the assignment sheet; there will be 3–5 assignments)
20% oral presentation of project methodology and outcomes (Apr 09)
20% oral exams (April 12-13)
40% project report (literature background and proposal in the form of an intermediate report due Feb 12, full report due Apr 16)

Each student will work on the same research project, in the form of a Data Science challenge posed by Farm Credit Canada. It is recommended to begin background reading for the project as early as possible.

DETAILS ON DELIVERABLES:

(i) Assignments will be handed out via UR Courses and announced in class. Solutions have to be handed in in UR Courses on or before the due date given on the assignment sheet.
   - NOTE: No-Plagiarism rules apply. The students must use only their own solutions, ideas, figures, examples, words, etc wherever possible. If a concept or an idea presented is not originally the students' contribution, sources must be clearly referenced. Other sources must not be quoted word for word, not even for a sequence of three words, unless absolutely necessary. It is also considered plagiarism if only minor changes to the original wording are made. If a source is quoted word for word, then quotation marks must be placed around the quote, followed by a reference to the source, including the original page number. (The same applies to figures, tables, program code, etc …)
   - NOTE: Any case of plagiarism will be immediately reported to the Dean of Graduate Studies and Research who will investigate.

(ii) An intermediate report on the project (due Feb 12)
This requires:
   -- a title
   -- an introduction which (a) motivates the problem the student is been working on and puts it into the context of existing literature, (b) clearly formulates the problem, (c) briefly summarizes the student's proposed approach,
   -- a section introducing relevant background and/or notation, including the required background knowledge as far as it has not been covered in CS 830 lectures
   -- a section explaining the student's proposed research approach
   -- references
   -- many examples, where helpful
   -- figures and tables, where helpful
   -- subsections (i.e., more structure in the report), where helpful
   - NOTE: Students will receive feedback on this intermediate report and can then revise the intermediate report and extend it for final submission of a full report at the end of the term (see item (iv) below)
   - NOTE: No-Plagiarism rules apply. The students must use only their own figures, examples, words, etc wherever possible. If a concept or an idea presented is not originally the students' contribution, sources must be clearly referenced. Other sources must not be quoted
word for word, not even for a sequence of three words, unless absolutely necessary. It is also considered plagiarism if only minor changes to the original wording are made. If a source is quoted word for word, then quotation marks must be placed around the quote, followed by a reference to the source, including the original page number. (The same applies to figures, tables, program code, etc ...)  
- NOTE: Any case of plagiarism will be immediately reported to the Dean of Graduate Studies and Research who will investigate.

(iii) An oral presentation (expected duration will be announced later) of the performed research towards the research project
- The presentation should motivate and explain the proposed research problem, present the related work, and briefly address the proposed approach. It is not expected that research results can be presented at this stage.
- The guidelines for preparation (and thus the criteria for evaluation) can be found in the document "Presentation Guidelines" posted on UR Courses.
- In the above document, Section 1.2 about outlines can be ignored. The instructor does not insist on outline slides at all. If, however, the student chooses to have an outline slide, then the corresponding section in the guideline document above applies.
- The time for your presentation must not exceed the time allocated by the instructor. The speaker will be cut off after the allocated time. After the presentation, the student will have to answer questions by the audience, which consists of the instructor and other graduate students.
- NOTE: No-Plagiarism rules apply. The students must use only their own figures, examples, words, etc wherever possible. If a concept or an idea presented is not originally the students' contribution, sources must be clearly referenced. Other sources must not be quoted word for word, not even for a sequence of three words, unless absolutely necessary. It is also considered plagiarism if only minor changes to the original wording are made. If a source is quoted word for word, then quotation marks must be placed around the quote, followed by a reference to the source, including the original page number. (The same applies to figures, tables, program code, etc ...)  
- NOTE: Any case of plagiarism will be immediately reported to the Dean of Graduate Studies and Research who will investigate.
- NOTE: A pdf version of the presentation slides, as used for the presentation, must be handed in by email to the instructor, no later than 24 hours before the presentation.

(iv) A final report on the project (about 15 pages long in single-spaced letter format) which revises and extends the intermediate report (cf. (ii))
- The final report should look like a publishable research article as much as possible (even if the results are not publishable). This requires:
  -- a title
-- an abstract summarizing the contributions of the student's work (it briefly states the context and says what the main contribution of the project is)
-- an introduction which (a) motivates the problem the student is been working on and puts it into the context of existing literature, (b) clearly formulates the problem, (c) briefly summarizes the student's proposed approach and main results*,
-- a section introducing relevant background and/or notations, including the required background knowledge as far as it has not been covered in CS 830 lectures
-- a section explaining the student's approach
-- a section (or maybe, if needed, more sections) showing and explaining all results in detail (for experimental work: a clear description of the experimental setup is required, so that the reader can reproduce the results)
-- a section interpreting the results
-- a conclusion section summarizing the report, discussing the limitations of the work, and suggesting future work
-- references
-- many examples, where helpful
-- figures and tables, where helpful
-- subsections (i.e., more structure in the report), where helpful
- NOTE: No-Plagiarism rules apply. The students must use only their own figures, examples, words, etc wherever possible. If a concept or an idea presented is not originally the students' contribution, sources must be clearly referenced. Other sources must not be quoted word for word, not even for a sequence of three words, unless absolutely necessary. It is also considered plagiarism if only minor changes to the original wording are made. If a source is quoted word for word, then quotation marks must be placed around the quote, followed by a reference to the source, including the original page number. (The same applies to figures, tables, program code, etc ...)  
- NOTE: Any case of plagiarism will be immediately reported to the Dean of Graduate Studies and Research who will investigate.

(v) Oral exams. The length of oral exams will be decided later in the term. Oral exams will be closed-book (no auxiliary material allowed) and cover the regular lecture material across the whole term, with the exception of designated special lecture material/tutorials specifically designed to aid students with their research project. The current plan is to have every exam audio-recorded. Students will have the opportunity to use paper and pencil during the exam to help explain their ideas.

FURTHER INFORMATION:

- Every document that is handed in late will be graded with 0%. Exceptions to this rule may be made at the instructor's discretion, on a case by case basis (e.g., in case of sickness).
- Office hours will usually be at the following times, starting after
the first lecture:
Monday from 9.15 am to 11.00 am via zoom;
Wednesday from 8.30 pm to 9.30 pm via zoom;
or by appointment via zoom.
Changes to this schedule will be announced by email through UR Courses.
- Lecture notes will usually be posted on UR Courses after class.
- Additional material of help for the research project will be posted on UR Courses. This additional lecture material will be provided in the form of notes and/or recorded video lectures and will be clearly designated as "RESEARCH PROJECT TUTORIAL". This material will NOT be tested on the oral exam.