CS406/531 Course Project: Parallelisation of Optimal Decision Tree Construction

ConTree builds optimal binary classification trees on continuous features using dynamic programming + branch-and-bound, with options such as max_depth, time_limit, and a max_gap tolerance for near-optimal trees.

Your project is on parallelising ConTree on multicore CPUs (OpenMP) and many-core GPUs (CUDA). Using MPI is not a must for this project. However, if you think that you are done with the single-machine parallelisation, it will be a bonus (your solution can be simulated on a single machine, but I will try to provide cluster access once you let me know).

Your solution should not deviate from the DP formulation in the paper. Yet, you can modify the implementation and apply techniques such as fusing levels or perform extra pruning to, respectively, increase the amount of work within a task or reduce the total amount of work. The main focus for the evaluation will be on the **parallelisation of the branch-and-bound method** used by the authors for the optimal decision tree construction.

The repo is here: https://github.com/ConSol-Lab/contree. It ships with C++ code, CMake build files, and example datasets that you can reuse in this project.

The corresponding paper is Briţa, Cătălin E., Jacobus G. M. van der Linden, and Emir Demirović. "Optimal Classification Trees for Continuous Feature Data Using Dynamic Programming with Branch-and-Bound." In *Proceedings of AAAI-25* (2025). pdf

You will need to read the paper to understand the code.

All AI usage, ChatGPT, CoPilot, Gemini, Claude, etc., is allowed. However:

1. You must provide your chat histories at the end in the appendix (including the reasoning history within the LLM).
2. It is you who will answer my questions on the things you have done in the project (why did you do this, what was your motivation, etc.)
3. It will be you who will answer the questions during the presentation. As a side note, for the same reason, the grades for the students within the same team may differ.

Next steps:

1. **Form your groups** - (until November 19, 2025, through here: https://docs.google.com/spreadsheets/d/1gtxSWOsSD4W8WkysgWKlYYzyrHw3IFeU4E0_F1HOlYg/edit?usp=sharing)
2. **Progress report**, including the roadmap, your ideas, preliminary results, etc. - (until December 8, 2025, through SUcourse).
3. **Presentations** will be done in January (will be scheduled later on January 12-14, 2026).
4. **The final report, codes**, etc., submission is due by January 14, 2026.

Have fun.