EECS 598 Computational Data Science

An in-depth introduction to computational methods in data science for identifying, fitting, extracting and identifying patterns in large data sets.

Lecture will typically begin with the introduction of a core data science method, followed by the student programming the method computationally with a computer assisting the student by certifying when the program is correct, interleaved with 'just-in-time' theory that will expose the student to the mathematics that underpin the methodology. Once the method has been correctly implemented, the students will be given a real world example or 'success story' to work with that illustrates when the algorithm 'works' as expected, followed by an instructor guided computational exploration of the various subtleties of the algorithm and its weakness.

The idea of these explorations is to *'lead with computation'* as a way to bring into sharper focus the underlying theory and to teach the learner `computational forensics' – in other words, a methodology for 'computational thinking' about when algorithm is working as it 'should' and how and when it might not and, most importantly, how the data might be re-processed for the algorithm to again work. Mathematical theory plays an important role in these explorations – in helping understand the failure modes thus discovered and when reasoning about if it is possible to mitigate their effect, and when possible, how.

Exposure to, and familiarity in, such forensics will facilitate higher-level computational reasoning about complex algorithms in data science that use a multitude of such tools we will learn about in class to accomplish a real-world task (e.g. handwriting recognition). This will give students the ability to reason about the behavior of the algorithm in real-world scenarios and identify how and when the algorithms and/or the data format have to be tweaked to realize the expected performance and/or infer hidden/suspected patterns in the dataset.

Homework sets including programming exercises illustrating concepts and algorithms on real-world data sets highlighting their applicability.

Example methods including linear and non-linear regression, neural networks, deep networks, convolutional neural networks, factor, tensorial and independent component analysis, sparsity inducing linear and non-linear regressions. Example real-world success stories including foreground and background subtraction in videos, image denoising, filling in (or imputing missing entries in an image, unmixing images and sounds, handwriting recognition, re-painting images in the style of your favorite artist and much more.

Prerequisites: Prior experience in programming in either MATLAB, C, C++, Python or R– this is critical since every lab-lecture involves substantial programming by the student. Graduate standing or permission of instructor. Students who have taken EECS 551 cannot take EECS 598 and vice-versa.

Proposed time: Fridays 1 – 4 pm @ TBD location on UM Central Campus
Office hours and GSI discussion: @ Weiser Hall (MIDAS facilities)
Instructor: Prof. Raj Rao Nadakuditi