CISC 371 Class 35

Summary of Nonlinear Data Analysis

In this course, we have touched on many technical matters related to data analysis. We assumed, as our intellectual foundation, that linear algebra and linear data analysis were familiar to us. We formulated our analysis methods by using an *objective function* that we sought to minimize, recognizing that a local non-global minimizer would most commonly be the result of our computations.

We began by understanding objective functions with a single scalar argument. Although this was a simple model, and relatively uncommon empirically, it helped us to understand key concepts such as:

- Relationships among derivatives
- Separation of search direction and stepsize
- Dynamic stepsize selection, especially by back-tracking
- The utility of linear models in an iterative framework

Before we proceeded to searching in a vector space, we needed some mathematical objects. The derivative of a scalar function with a vector argument was written as a *1-form*; this is a dual of a vector so, if a vector was written as a matrix with one column, then a 1-form was written as a matrix with one row. This notation clarified the use of a gradient and the distinction between a gradient and a vector. We wrote the direction of *steepest descent* for a vector argument as the negated transpose of the gradient.

Next, we studied *unconstrained optimization*. This model used an objective function with a single vector argument. The search process was to iteratively update an estimate of a minimizer argument, using a *search direction* and a stepsize. The basic method was steepest descent; alternative methods included Newton's method and scaling methods, which are descent methods that alter the direction of steepest descent by a transformation that is a symmetric positive semidefinite matrix. Unconstrained optimization was extended to nonlinear least squares, training neural networks, and Tikhonov regularization that altered to objective function with an additional term.

Our final exploration was an introduction to *constrained optimization*. We used Lagrange multipliers to transform a constrained problem into an unconstrained problem that could be solved using descent methods. In some applications, we found that the dual Lagrange formulation was a computationally superior expression of a constrained problem. We explored the support vector machine, or SVM, which incorporated kernel methods and soft margins in the dual formulation.

This was a first course in nonlinear data analysis. Complementary courses might include formulations other than optimization in a vector space, such as discrete optimization and decision formulations. Subsequent courses might include numerical methods for optimization, and other methods for transforming a constrained optimization problem such as penalty methods or barrier methods.

Nonlinear optimization is found abundantly in current machine learning, especially computational architectures of deep neural networks. Understanding how optimization works is an essential tool in understanding how to analyze data and how to learn from data.