CS/ECE/ME 532 Sec. 004
Matrix Methods in Machine Learning

[Unit 5] Iterative Methods for Solving Systems of Linear Equations – Overview

Learning Objectives

At the end of this module, students will be able to:

- Derive and implement gradient descent for solving the least squares problem
- Identify the range of step size required for convergence of gradient descent
- Explain why ell-1 regularization tends to encourage sparse solutions
- Apply the proximal gradient method to obtain a gradient descent algorithm for solving regularized least-squares problems
- Identify the range of step size required for convergence of proximal gradient descent
- Define convexity and explain its importance for solving optimization problems
- Explain the term sub gradient and its role with nondifferentiable function optimization
- Explain how to perform stochastic gradient descent and why it is used
- Define hinge loss, explain why it is used, and calculate hinge loss for simple classifiers

Significance of Unit

Unit 5 is iterative methods. So far when we've looked at solving machine leaning problems, in general we've assumed that we could write down the solution or that we could compute a transpose inverse and that's what we needed to do is multiply that by a transpose times "d" and that would give us the solution to a problem. In Unit 5 we are going to talk about finding that solution in an iterative/sequential manner. The basic idea is that we are going to look at the loss function or the cost that we are trying to minimize. One example is squared air, it has a bowl-shaped surface and our solution is at the bottom of the bowl. So, we'll start at some point and if we want to go towards the solution all we have to do is take a step in the downhill direction, and from wherever we end up take another step, so we sequentially iterate towards the solution. And these iterative methods are really powerful for a number of reasons and they are critical for machine learning because a lot of the problems are really large, and we can't compute a transpose inverse because the computation would take way too much time. A lot of times we can't even store the entire matrix in memory at once because of the size of our training set. Being able to do this in an iterative manner allows us to solve those problems where we can't do the computation any other way. It also allows us to generalize to different types of cost functions, for example we are going to look at ways of finding solutions that are sparse, in other words our classifier for example only has a few non zero terms and when you set up that problem you can't solve for it exactly in closed form so we have to solve for that iteratively. And then we are also going to look at different classification problems where we solve for something called hinge loss which is more relevant to errors. The stuff we're doing in Unit 5 is very important and profound, it also sets the stage for training neuro which we are going to look at in Unit 6.

Key Topics

1. Gradient descent
2. Proximal gradient for regularized problems
3. Stochastic gradient descent
4. LASSO regularization
5. Hinge Loss

Learning Activities

- Instructional Unit 5.1
- Activity 16
- Instructional Unit 5.2

- Activity 17
- Instructional Units 5.3, 5.4
- Activity 18
- Assignment 8
- Instructional Units 5.5, 5.6
- Activity 19
- Instructional Units 5.7, 5.8
- Activity 20
- Unit 5 Overview Quiz

Recommended Reading

- None