Stochastic Gradient Descent

STAT 672 Project

Tom Wallace

George Mason University

Spring 2018

Introduction

Stochastic gradient descent (SGD) is an optimization algorithm that is fundamental to many machine learning approaches. We will examine SGD from four perspectives.

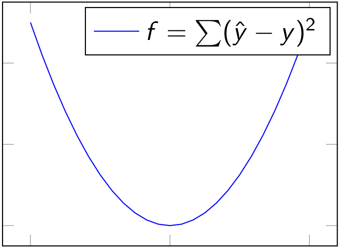
Table of Contents

Optimization is everywhere, and sometimes easy

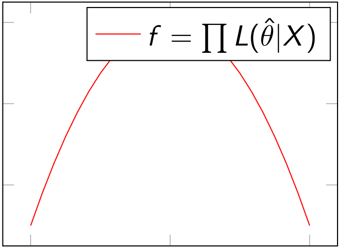
Many statistical procedures involve minimizing or maximizing some function applied to data

In **parametric** statistics, we often make assumptions about the distribution of the data that make this optimization "nice" (convex or concave)

OLS



MLE



Other times, optimization is not easy

In **non-parametric** settings, we conduct **empirical risk minimization**

Table of Contents

foo

Table of Contents

foo

Table of Contents

foo