Reviews of stochastic gradient descent variants

Introduction

We firstly give up the least square method of linear regression, and then address linear regression problem based on a general machine learning method. In my opinion, Machine learning algorithm consists of **model, strategy** and **optimization**.

$$\text{Learning algorithm} = \text{Model} + \text{Strategy} + \text{Optimization};$$

The first thing we should do is to determine what decision function or conditional probability distribution we need to learn and get. For instance, in linear regression we want to learn $h_\theta(x) = \theta x$. This decision function is the model we want to use.

We have the model, but we do not know parameters in it, and next thing we should do is parameter estimation, which means how to construct loss function, and we call this step strategy.

The last step is optimization which means how to solve the loss function.

Notation

Gradient Descent

Stoastic Gradient Descent

Stoastic Average Gradient Descent(SAG)

Stoastic Gradient Descent with Predictive Variance Reduction(SVRG)