

Multivariate Optimization 3

Exercise 1: Stochastic Gradient Descent

Consider the ordinary linear regression (without intercept) where we want to minimize

$$\mathbb{E}_{\mathbf{x}, y} \|\boldsymbol{\theta}^\top \mathbf{x} - y\|_2^2$$

with $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{x}})$ and $y|\mathbf{x} \sim \mathcal{N}(\boldsymbol{\theta}^{*\top} \mathbf{x}, \sigma^2)$.

- (a) Show that $\mathbb{E}_{\mathbf{x}, y} \nabla_{\boldsymbol{\theta}} \|\boldsymbol{\theta}^\top \mathbf{x} - y\|_2^2 = \nabla_{\boldsymbol{\theta}} \mathbb{E}_{\mathbf{x}, y} \|\boldsymbol{\theta}^\top \mathbf{x} - y\|_2^2$
- (b) Interpret a) in terms of SGD.
- (c) Consider the univariate setting with $\boldsymbol{\Sigma}_{\mathbf{x}} = 1/4$, $\sigma = 1/10$, $\boldsymbol{\theta}^* = 1/2$ and a data set of size 10,000.
Write an R script which plots the "confusion", i.e., the variance of the gradients, for $\theta \in \{0, 0.05, 0.1, \dots, 0.95, 1.0\}$. For each θ , plot 200 gradient samples.
Perform two such simulation studies with random batches of size 100 and 1,000.
- (d) What do you observe in c) ?
- (e) Write an R script which solves the setting in c) with SGD with random batch sizes of 1 and $\alpha = 0.3$. Start with $\boldsymbol{\theta} = 0$ and perform 20 iterations. Repeat this process 200 times. Compare with GD.