https://slds-lmu.github.io/website_optimization/

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

- (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 the univariate setting with $\Sigma_{\mathbf{x}} = (0.5^2), \sigma = 0.1, \boldsymbol{\theta}^* = (0.5)$. Write an R script which plots the "confusion", i.e., the variance of the gradients, for $\theta \in \{0, 0.05, 0.1, \dots 0.95, 1.0\}$ (each setting repeated 200 times). Do two such simulation studies, one where the random batches are of size 100 and one with random batch sizes of 1000.
- (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 $\theta = 0$ and do 20 iterations. (Repeat 200 times). Compare with GD.