IEEE MAG 1: Introduction

Tuesday, April 2, 2024 5:34 PM

- Distributed consensus-based gradient methods for convex problems.
- Static graphs (the topology of the computational graph is static during execution.

- A standard problem:

Minimizing the Average of the Sum of Functions:

$$\min_{x \in \mathbb{R}^n} \frac{1}{p} \sum_{i=1}^{p} f_i(x)$$

$f_i(x)$: loss associated w/ a data point x

Data points: $\{(z_i, y_i),\ i = 1, \cdots, p\}$ (z_i, y_i) is an instance of x

$z \in \mathbb{R}^n$ feature vector

$y_i \in$ is the corresponding label $\{+1, -1\}$, $\{0, 1\}$, etc.

Non-linear classification problem: Minimizing the problem:

$$\min_{x \in \mathbb{R}^n} \Big(c \underbrace{\rho}_{\text{greek } \rho}(x) + \frac{1}{\underbrace{p}_{\text{latin } p}} \sum_{i=1}^{p} \ell(x : z_i, y_i) \Big)$$

p = number of data points

$c > 0$: regularization parameter

x : input vector

$\rho(x)$: regularizing function, strongly convex

$\ell(x : z_i, y_i)$: loss function, usually convex, sometimes non-differentiable

↳ Logistic Regression Loss, Hinge Loss, etc.

$$\underbrace{f_i(x)} = c\rho(x) + \ell(x : z_i, y_i)$$ associated w/ data point (z_i, y_i)

- For datasets with few data points p,

↳ Use iterative gradient method:

$$x^{k+1} = x^k - \frac{\alpha_k}{p} \sum_{i=1}^{p} \underbrace{\nabla f_i(x^k)}$$

↳ Converges if $\sum_{k=0}^{\infty} a_k = \infty$ and $\sum_{k=0}^{\infty} a_k^2 < \infty$