Homework 2

Due: Sunday, 11:59PM, **October 10, 2021**

1. Programming Assignment

The linear regression is formulated as

$$\min_{\boldsymbol{\theta}} \quad \mathcal{L}(\boldsymbol{\theta}) := \frac{1}{N} \sum_{n=1}^{N} \left(y_n - \mathbf{x}_n^\top \boldsymbol{\theta} \right)^2 \tag{1}$$

where $\{\mathbf{x}_n, y_n\}_{n=1}^N$ are given training data.

1.1 Dataset

Air Quality dataset collects $9,358$ instances of hourly averaged responses from five chemical sensors located in an area of Italy. The averaged sensor response $\mathbf{x}_n \in \mathbb{R}^{13}$ contains the hourly concentrations of e.g., CO, Non Metanic Hydrocarbons, and Nitrogen Dioxide (NO2), where the goal is to predict the concentration of polluting chemicals $y_n \in \mathbb{R}$ in the air.

Download: https://archive.ics.uci.edu/ml/datasets/Air+quality

1.2 Algorithm

1. Gradient descent algorithm.
2. Stochastic gradient descent algorithm (one sample per iteration).
3. Mini-batch stochastic gradient descent algorithm (50 samples per iteration).

1.3 Submitted results

1. Well documented MATLAB/Python codes.

2a. Plots of iteration index vs objective error for a fixed stepsize.

2b. Plots of CPU time vs objective error for a fixed stepsize.

3. Brief discussions of your findings relative to the convergence theory.

2. Programming Assignment

The ℓ_2 regularized logistic regression is formulated as

$$\min_{\boldsymbol{\theta}} \quad \mathcal{L}(\boldsymbol{\theta}) := \frac{1}{N} \sum_{n=1}^{N} \log \left(1 + \exp(-y_n \mathbf{x}_n^\top \boldsymbol{\theta}) \right) + \frac{\lambda}{2} \|\boldsymbol{\theta}\|^2 \tag{2}$$

where $\{\mathbf{x}_n, y_n\}_{n=1}^N$ are given training data, and $\lambda > 0$ is the regularization constant.

2.1 Dataset

Ionosphere dataset is to predict whether it is a "good" radar return or not – it is "good" if the features in $\mathbf{x}_n$ show evidence of some type of structure in the ionosphere.

Download: https://archive.ics.uci.edu/ml/datasets/ionosphere

2.2 Algorithm

1. Gradient descent algorithm.
2. Stochastic gradient descent algorithm (one sample per iteration).
3. Mini-batch stochastic gradient descent algorithm (50 samples per iteration).

2.3 Submitted results

1. Well documented MATLAB/Python codes.

2a. For $\lambda = 0$, plots of iteration index vs objective error for a fixed stepsize.

2b. For $\lambda = 0$, plots of CPU time vs objective error for a fixed stepsize.

3a. For $\lambda = 0.01$, plots of iteration index vs objective error for a fixed stepsize.

3b. For $\lambda = 0.01$, plots of CPU time vs objective error for a fixed stepsize.

4. Brief discussions of your findings relative to the convergence theory.