OPTIMIZATION. HOMEWORK 12

OSCAR DALMAU

(1) Consider the optimization problem

$$F(\theta) = \frac{1}{N}\sum_{i=1}^{N}(h_\theta(\boldsymbol{x}_i) - y_i)^2 \tag{1}$$

where $(\boldsymbol{x}_i, y_i)$, $\boldsymbol{x}_i \in \mathbb{R}^n$, $y_i \in \{0,1\}$, $i = 1, 2, \cdots, N$ are given and

$$h_\theta(\boldsymbol{x}) = f_{\boldsymbol{a},b}(g_{\boldsymbol{c},\boldsymbol{d}}(\boldsymbol{x}))$$

$$\begin{aligned} g_{\boldsymbol{c},\boldsymbol{d}} &: \mathbb{R}^n \to \mathbb{R}^m \\ f_{\boldsymbol{a},b} &: \mathbb{R}^m \to \mathbb{R} \end{aligned}$$

$m \in \mathbb{N}$ is known,

$$\begin{aligned} g_{\boldsymbol{c},\boldsymbol{d}}(\boldsymbol{x}) &= [\sigma(\boldsymbol{c}_j^T\boldsymbol{x} + d_j)]_{j=1}^m \\ f_{\boldsymbol{a},b}(\boldsymbol{z}) &= \sigma(\boldsymbol{a}^T\boldsymbol{z} + b) \\ \sigma(t) &= \frac{1}{1+e^{-t}},\ t \in \mathbb{R} \end{aligned}$$

and θ corresponds to the set of parameters $\boldsymbol{a}, b, \boldsymbol{c}, \boldsymbol{d}$.

Implement the following algorithms:

- Gradient descent with fixed step size and gradient approximation.
- BFGS with gradient approximation.

Compare the previous algorithms with respect to number of iteration and computational time using the mnist dataset and selecting only two digits.

Note: You can use any library to load mnist dataset. However, you can use dataset mnist.pkl.gz

- train_set[0] is a matrix of size (50000, 784) where $n = 50000$ is the number of observations and each row represents an observation $\boldsymbol{x}_i \in \mathbb{R}^{784}$

- train_set[1] is a vector of size (50000) where each entry $y_i \in \{0, 1, \cdots, 9\}$

Select from train_set[0] and train_set[1] the set of observations $\mathcal{S} = \{(\boldsymbol{x}_i, y_i)\}$ with $\boldsymbol{x}_i \in \mathbb{R}^{784}$ and $y_i \in \{0, 1\}$ and estimate the parameters $\hat{\theta}$ that minimizes the function $F(\theta)$ in equation (1).

Select from test_set[0] and test_set[1] the set $\mathcal{T} = \{(\boldsymbol{x}_i, y_i)\}$ such that $\boldsymbol{x}_i \in \mathbb{R}^{784}$ and $y_i \in \{0, 1\}$ and compute the error

$$error \quad = \quad \frac{1}{|\mathcal{T}|} \sum_{(\boldsymbol{x}_i, y_i) \in \mathcal{T}} |\mathbf{1}_{h_{\hat{\theta}}(x_i) > 0.5}(\boldsymbol{x}_i) - y_i|$$

where $|\mathcal{T}|$ represents the number of elements of the set $\mathcal{T}$.