OPTIMIZATION PROJECT

Consider a multiclass logistic regression problem of the form:

$$\min_{x \in \mathbf{R}^{d \times k}} \sum_{i=1}^{m} \left[-x_{b_i}^T a_i + \log \left(\sum_{c=1}^{k} e^{x_c^T a_i} \right) \right]$$

Likelihood for a single training example i with features $a_i \in \mathbf{R}^d$ and label $b_i \in \{1, 2, \ldots, k\}$ is given by

$$P(b_i \mid a_i, X) = \frac{e^{x_{b_i}^T a_i}}{\sum_{c=1}^{k} e^{x_c^T a_i}}$$

where x_c is column c of the matrix parameter $X \in \mathbf{R}^{d \times k}$.

To maximize the likelihood over m independent and identically distributed (iid) training samples, use minimize negative log-likelihood:

$$f(x) = \sum_{i=1}^{m} \left[x_{b_i}^T a_i + \log \left(\sum_{c=1}^{k} e^{x_c^T a_i} \right) \right]$$

The partial derivative is:

$$\frac{df(x)}{dx_{jc}} = -\sum_{i=1}^{m} a_{ij} \left[\mathrm{I}\left(b_i = c\right) \frac{e^{x_c^T a_i}}{\sum_{c=1}^{k} e^{x_{c'}^T a_i}} \right]$$

with $\mathrm{I}\left(b_i = c\right)$ being indicator variants.

HOMEWORK:

1. Randomly generate a 1000x1000 matrix with entries from a N(0,1) distribution.

2. Generate $b_i \in \{1, 2, \ldots, k\}$ with $k = 50$ by loading

$$AX + E$$

with X, E sampled from a normal distribution

$$X \in \mathbf{R}^{d \times k}, E \in \mathbf{R}^{m \times k}$$

(consider max index in the row as class label!).

3. Solve the multiclass problem with:

A: Gradient Descent

B: BCGD with randomized rule

C: BCGD with Gauss-Southwell rule

Use blocks $X_{jc'}, c' \in \{1, \ldots, k\}$ (each row of $X_{jc'}$ is one block!!!)

4. Choose a publicly available dataset and test your methods on this.

5. Analyze accuracy vs. CPU time.

6. Describe what you did on a PDF file.

7. Submit project.