

Solve Multi-Class Logistic Regression Problems using

-GD

-BCGD with randomized rule

-BCGD with GS rule

students ≤ 2 Implement GD and BCGD with GS rule

students > 2 Implement all algorithms

Deliverables

1. Every group needs to upload only one project on moodle;
- 2) the files to be included in the moodle homework page:
 - a pdf file with a detailed description of the hw solution (pdf file name should include last names of the group members);
 - all the files needed to run the code you developed;
 - a readme file with the details (first and last name, UniPD ID code) of all the group members.

Multi class regression

Consider

$$\min_{x \in \mathbb{R}^{d \times k}} \sum_{i=1}^m [-x_{b_i}^T a_i + \log(\sum_{c=1}^k \exp(x_c^T a_i))] = \min_{x \in \mathbb{R}^{d \times k}} \hat{t}(x)$$

The Likelihood for classes assigned to example i with features $a_i \in \mathbb{R}^d$ and label $b_i \in \{1, 2, \dots, k\}$ is

$$p(b_i | a_i, X) = \frac{\exp(x_{b_i}^T a_i)}{\sum_{c=1}^k \exp(x_c^T a_i)}$$

Where x_c is defined to be the column c of parameter matrix $X \in \mathbb{R}^{d \times k}$.

Note maximizing the likelihood over m i.i.d. training samples is the same as minimizing the negative log-likelihood.

$$\min_{x \in \mathbb{R}^{d \times k}} \hat{t}(x) = \max_{x \in \mathbb{R}^{d \times k}} p(b_i | a_i, X)$$

Free Tip

$$\frac{\partial f(x)}{\partial X_{jc}} = - \sum_{i=1}^m [I(b_i=c) - \frac{\exp x_c^T a_i}{\sum_{c'=1}^k \exp(x_c^T a_i)}]$$

Homework

Due 21 May

1. Randomly generate 1000 by 1000 matrix A . Take entries from $N(0, 1)$.
2. Generate $b_i \in \{1, 2, \dots, k\}$ with $k = 50$ to create the target classes. Create the initial output by calculating

$$b = AX + E$$

Where X, E also sampled from normal distribution

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

3. Solve the above optimization problem with
 1. Gradient Descent
 2. BCGD with randomized Rule
 3. BCGD with Gauss-Southwell Rule

Note blocks are $X_{jc'}, c' = \{1, 2, \dots, k\}$
 And each row of X_{jc} is one block
4. Chose a publicly available dataset and test methods
5. Analyze accuracy vs. CPU time.
6. Document work in PDF writeup
7. Submit