

ODS-24 sratch

Tanner Aaron Graves

May 2024

1 Introduction

Multi-class logistic classification is an extension of traditional logistic regression. The task goes from predicting a Bernoulli random variable to predicting one label to classify each example from a set of k mutually exclusive labels. We encode the target $b_i \in \mathbb{R}^k$ where k is the number of output classes to be a one hot vector, corresponding to the correct classification of example $a_i \in \mathbb{R}^d$. It assigns these labels by learning the conditional probability distribution that the correct label is b_i given example a_i and learned parameter matrix $X \in \mathbb{R}^{d \times k}$. Where d is the number of features in each example. The output distribution is computed with the softmax function in this context:

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

Multi-class logistic regression is a simple yet effective method with ubiquitous use. Its training does, however, require learning the parameter matrix X , which can be done using various methods like maximum likelihood or gradient descent based methods. This corresponds to solving the following optimization problem:

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

which is the cross-entropy of the predicted class distribution and the target label b_i . In this report we will analyze the relative performance of various gradient methods at solving this problem. Specifically, we are interested in comparing full gradient descent with block coordinate based methods (BCGD) which optimize one coordinate, or column in parameter matrix X , at a time.

2 Algorithms

test

Algorithm 1 Full Gradient Descent

```
Initialize  $X \in \mathbb{R}^{d \times k}$ 
for  $k = 1, \dots$  do
  If  $X \in X^*$  STOP (Optimality conditions)
  Set  $X_{k+1} = x_k - \alpha_k \nabla$ 
   $N \leftarrow \frac{N}{2}$  ▷ This is a comment
end for
```

2.1 Gradient Descent

2.2 Block coordinate Gradient Descent

Algorithm 2 An algorithm with caption

```
Require:  $n \geq 0$ 
Ensure:  $y = x^n$ 
 $y \leftarrow 1$ 
 $X \leftarrow x$ 
 $N \leftarrow n$ 
while  $N \neq 0$  do
  if  $N$  is even then
     $X \leftarrow X \times X$ 
     $N \leftarrow \frac{N}{2}$  ▷ This is a comment
  else if  $N$  is odd then
     $y \leftarrow y \times X$ 
     $N \leftarrow N - 1$ 
  end if
end while
```
