Multi-core implementation of stochastic variance reduced algorithms

LI Zhirong 1155092195

Prof. James Cheng, Phd. Yan Xiao

The Chinese University of Hong Kong

Objectives

- Stochastic variance reduced algorithms
- Multi-core implementation on Logistic Regression and SVM
- Experiment on classification

Introduction

Stochastic Gradient Descent (SGD) is a popular algorithm proved to be well suited to data-intensive machine learning tasks but has the limitation of low convergence due to the inherent variance. Stochastic variance reduced gradient(SVRG) is a better improvement for smooth and strongly convex functions. In this research, we combine the idea of lazy update[1] and SVRG[2] and implement them in multi-core logistic regression and support vector machines(SVMs). In addition, our program will perform on dataset RCV1 to do binary classification. As a result, with cores increasing, convergence becomes faster and algorithms perform better on logistic regression than SVM.

Method

In **logistic regression**, loss function is:

$$E = -\sum_t r^t \log y^t + (1 - r^t)\log(1 - y^t)$$

Stochastic Gradient in logistic regression is:

$$\Delta w_j = \eta(r^t - y^t)x_j^t, j = 1, ..., d$$

In **SVM**, loss function is:

$$f(w) = \frac{\lambda}{2}\|w\|^2 + \sum_{i=1}^{N} \max(0, 1 - y_i(w^T X_i + b))$$

Stochastic Gradient in logistic regression is:

$$\nabla f_i(w) = \lambda w + \Phi(1 - y_i(w^T X_i + b) > 0)(-y_i x_i)$$

SVRG update rule:

$$w^{(t)} = w^{(t-1)} - \eta_t(\nabla\psi_{i_t}(w^{(t-1)}) - \nabla\psi_{i_t}(\tilde{w}) + \tilde{\mu})$$

with

$$\tilde{\mu} = \frac{1}{n}\sum_{i=1}^{n} \psi_i(\tilde{w})$$

Lazy update in SVRG:

Algorithm 1 Lazy Stochastic Updates for SVRG

```
Input: x; μ; T
Initialize ρ(j) = 0, j = 0, ..., d
for t=1:T do
    Randomly sample i ∈ {1, ..., n}
    x_{S_i}=read coordinates S_i from x
    for j ∈ S_I do
        τ_j=t-ρ(j)-1
        w_j ← w_j − ητ_jμ_j
        w_j ← w_j − η(∇ψ_{i_t}(w^{(t−1)}) - ∇ψ_{i_t}(w̃))
        ρ(j) ←t.
```

Tools

- Programming in C++
- Using OpenMP for parallel computing
- CSE servers with multi-core system

Results

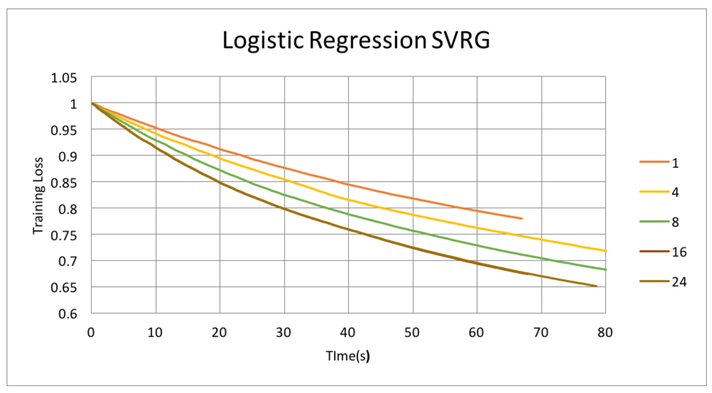

Figure 1: With cores increasing it converges faster

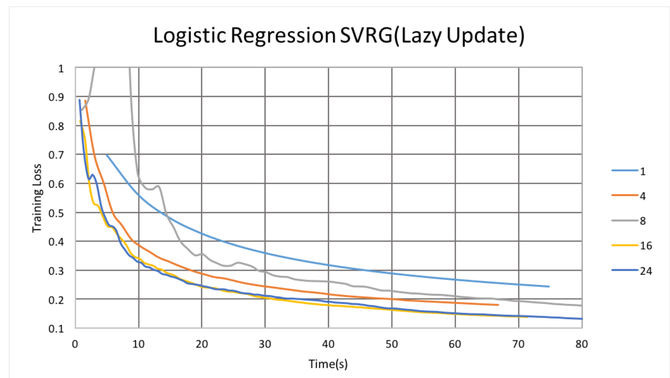

Figure 2: With cores increasing it converges faster but is a little unstable

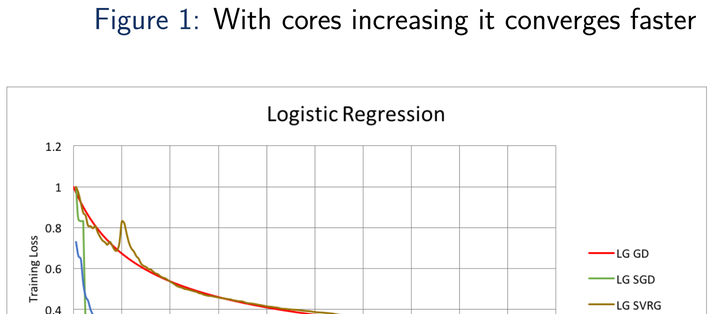

Figure 3: With lazy update, SVRG is more stable and converging faster

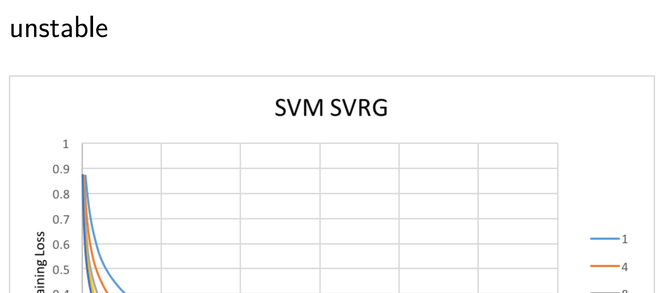

Figure 4: With cores increasing it converges faster

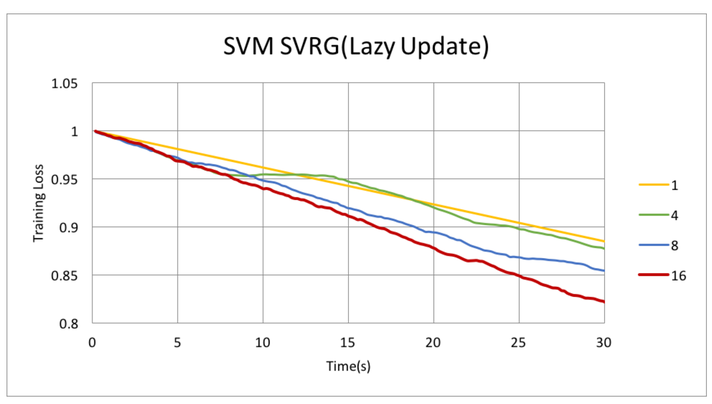

Figure 5: With cores increasing it converges faster but is a little unstable

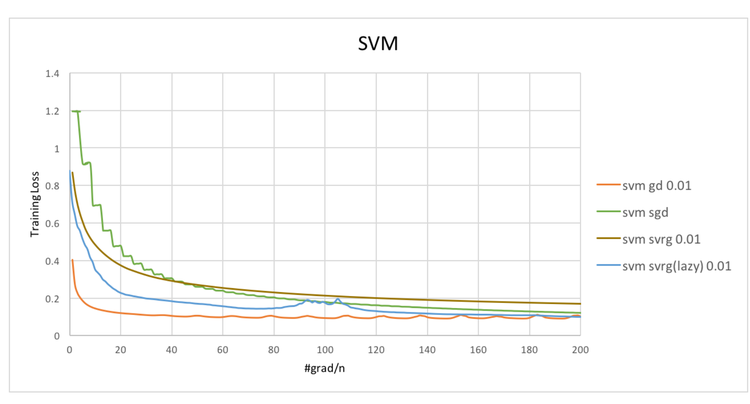

Figure 6: SVRG and lazy updated SVRG worked better than original SGD

Conclusion

Firstly, we performed multi-core binary class logistic regression on sparse dataset RCV1 with learning rate η=0.01(Fig1,2). When number of core increases, convergence was faster in SVRG(Fig.1). When lazy update was applied, convergence was also faster but it was not so smooth(Fig.2). Lazy updated SVRG performed better than others with 4 cores(Fig.3). In stochastic gradient descent, the learning rate decay in such a decrease scheduling: $\frac{\alpha}{\#iter*\beta+0.5} + 0.01$, where α and β was adjusted according to experiment(Here α=8 and β=1/50).

Secondly, we performed SVM on RCV1 with learning rate η=0.01 and regularization parameter λ=1e-4(Fig.4,5). As a result, when number of core increases, convergence was a little faster but it is not apparent(Fig.4). In addition, the lazy update also helped in convergence(Fig.5). In stochastic gradient descent, the decrease scheduling was the same as logistic regression, where α=50 and β=200. When the number of core is 4, We found that SVRG and lazy updated SVRG worked better than original SGD in the beginning, especially lazy updated SVRG(Fig.6).

References

[1] S. Tu D. Papailiopoulos C. Zhang M.I. Jordan K. Ramchandran C. Re B. Recht. X. Pan, M. Lam. Cyclades:conflict-free asynchronous machine learning. *arXiv:1605.09721*, 2016.

[2] T. Zhang. R. Johnson. Accelerating stochastic gradient descent using predictive variance reduction. *Advances in Neural Information Processing Systems*, 26, 2013.

Contact Information

- Email: stella.lee.lzr@gmail.com
- Phone: +852 67428565
- Address: Chung Chi Chollege, The Chinese University of Hong Kong , Shatin, N.T. Hong Kong