相关论文整理

1. **Machine Learning with Adversaries: Byzantine Tolerant Gradient Descent**

author：Peva Blanchard，El Mahdi El Mhamdi，Rachid Guerraoui，Julien Stainer

school：EPFL, Switzerland，8 March, 2017

link：<http://papers.nips.cc/paper/6617-machine-learning-with-adversaries-byzantine-tolerant-gradient-descent.pdf>

思想：compute a score for each candidate estimator gradient, rank the score, pick the minimum score.

算法：krum aggregation rule as follows.

1) For any , denote by the fact that belongs to the

closet vectors to .（f is the number of Byzantine worker）

2)For each worker ,compute score

3)Finally, , where refers to the worker minimizing

the score, .

1. **Distributed Statistical Machine Learning in Adversarial Settings:**

**Byzantine Gradient Descent**

Author: Yudong Chen(Cornell), Lili Su(UIUC), Jiaming Xu(Purdue)

Link: <http://delivery.acm.org/10.1145/3160000/3154503/pomacs44-chen.pdf?ip=114.214.166.227&id=3154503&acc=ACTIVE%20SERVICE&key=BF85BBA5741FDC6E%2EA4F9C023AC60E700%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&__acm__=1531300193_09c63afb5f7f68e6c67da747dca4bbdf>

思想：使用geometric median，GD-based算法

算法：geometric median of means

Formulation: 

1. **Securing Distributed Machine Learning in High Dimensions**

Author: Lili Su(MIT), Jiaming Xu(Duke University)

Link : https://arxiv.org/pdf/1804.10140.pdf

1. **Generalized Byzantine-tolerant SGD**

Author: Cong Xie(UIUC), Oluwasanmi Koyejo(UIUC), Indranil Gupta(UIUC)

Link: <https://arxiv.org/pdf/1802.10116.pdf>

思想：使用论文1中提出的一个性质，提出自己的算法满足这个性质，具体看论

文1中的 “Byzantine Resilience”

Formulation: 

算法：都是median based 的算法

1. Marginal Median

,

for any , the th dimension of is

is the th dimension of the vector , median(.) is the one-dimensional median.

1. Mean round median 在（1）的基础上，对离每一维的中位数最近的 个元素取平均，其中q为错误节点个数。

,

for any , the th dimension of is

, is the indices of the top- values lying in nearest to the median , is the th dimension of the vector .

1. **Zeno: Byzantine-suspicious stochastic gradient descent**

Author: Cong Xie(UIUC), Oluwasanmi Koyejo(UIUC), Indranil Gupta(UIUC)

Link: <https://arxiv.org/pdf/1805.10032.pdf>

思想： 设计了一个检测算法。对每个节点，server会根据准则计算其可靠性。本文算法的关键就是计算可靠性的函数。

算法：（Stochastic Descendant Score）定义 ，则计算可靠性的函数为：

Score越大，则可靠性越大；对每个节点计算得到的score做排序，选择其中分数最高的前 个节点对应的梯度作平均。

1. **Byzantine Stochastic Gradient Descent**

Author: Dan Alistarh(IST Austria), Zeyuan Allen-Zhu(Microsoft Research AI),

Jerry Li(MIT CSAIL)

Link: <https://arxiv.org/pdf/1803.08917.pdf>

Formulation: 

思想：设计了两个阈值，需要累积所有迭代次数的梯度，不是很实用

1. **《LIBMF\_ A Library for Parallel Matrix Factorization in Shared-memory Systems》**

（基于C/C++，Lib\_P\_MF）

做并行SGD矩阵分解的快速调用包。相当于简易说明书

包括：non-negative matrix factorization, binary matrix factorization(与实值矩阵相对), logistic matrix factorization, one-class matrix factorization

1. **《A Learning-rate Schedule for Stochastic Gradient Methods to Matrix Factorization》**

使用SGD做矩阵分解的时候如何快速准确地调参。

已有方法：

1.固定步长

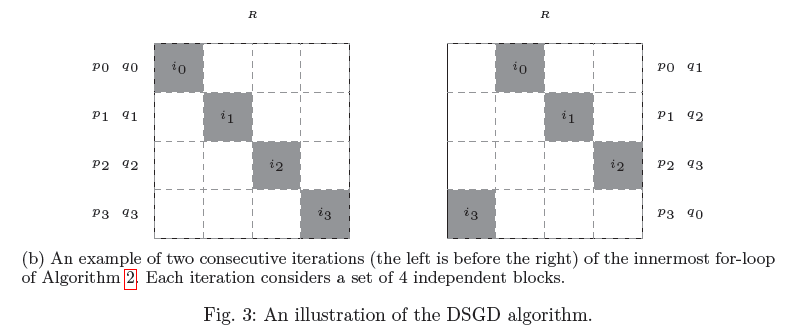
2.单调下降（Monotonically Decreasing Schedule, MDS）

3.Bold-driver Schedule, BDS

根据误差减小的正负，放缩地乘在系数上。这个可以考虑，但是涉及函数内修改全局变量的问题，是不是又要用manager。

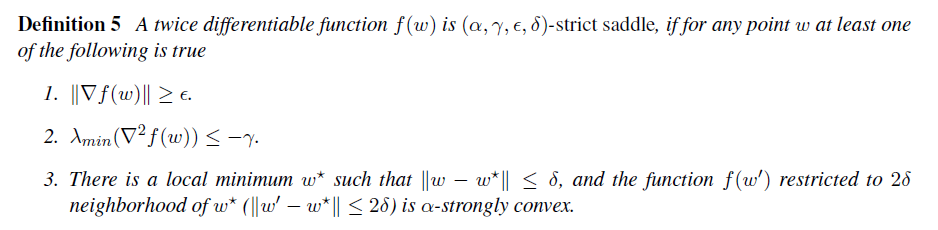
1. **A Fast Parallel Stochastic Gradient Method for Matrix Factorization in Shared Memory Systems**

将矩阵按行列互不相同取出不同的pattern，每次迭代将pattern中被选中的块分发给不同节点进行运算，避免死锁问题。

****

1. **Escaping From Saddle Points – Online Stochastic Gradient for Tensor Decomposition**

定义了严格鞍点，并分析SGD如何逃离鞍点到达局部最优

****

1. **HOGWILD! A Lock-Free Approach to Parallelizing Stochastic Gradient Descent**

Ntoe：将一些问题归结为一个稀疏超图上的优化，结点处变量1维。问题结构与我们关心的有差异。SGD的“随机”在于随机抽取超图的边。存在术语定义不清的问题，如locking、k(j)的定义、prop 4.1下三行processor的数量等。不建议作为常用reference。