

周报

刘精昌

2016 年 12 月 8 日

本周工作

1. 阅读《Stochastic Alternating Direction Method of Multipliers》ICML13，主要是看了数学证明。文章的一个具有启发式的 idea 是，ADMM 迭代中，为了求解需要，使用泰勒近似，这种思想来自于 2011 年 Nips 上林宙辰的一篇，该文的引用好几百次。
2. 看了 MTL 相关文章，看的都比较粗略：
 - 《Multi-Task Feature Learning》Nips06，关注于任务间特征稀疏问题，比较古老以及经典的文章。
 - 《Parallel Multi-Task Learning》ICDM15 和《Distributed Multi-Task Learning》AISTATS16，看的不很明白，感觉做的有点儿杂乱。
3. 考虑了分布式多任务学习的问题。对于 $\sum_i^T \text{loss}(w_i, X_i, Y_i) + \lambda \text{Reg}(W)$ 形式的多任务学习优化问题。为了应用 ADMM 以进行分布式处理，需将 w_i 和 W 之间做到分离，把原优化问题转化为下面的优化问题：

$$\begin{cases} \min \sum_i^T \text{loss}(w_i, X_i, Y_i) + \lambda \text{Reg}(Q) \\ QI'_i = w_i, i = 1, 2, \dots, T \end{cases}$$

其中， $I_i = (0, \dots, 0, 1, 0, \dots, 0)^T$ ，第 i 个元素是 1。该问题的增广拉格朗日函数为：

$$L_\rho(w_1, \dots, w_T, \mu_1, \dots, \mu_T, Q) = \sum_i^T \text{loss}(w_i, X_i, Y_i) + \lambda \text{Reg}Q + \sum_i^T \mu_i (QI_i^T - w_i) + \sum_i^T \frac{\rho}{2} \|QI_i^T - w_i\|_2^2$$

利用 ADMM，对于 w_1, w_2, \dots, w_T ，表示每个 task，可以由各个 workers 来 update：

$$w_i^{k+1} = \arg \min_{w_i} L_\rho(w_i, Q^k, \mu_i^k)$$

由 master 来更新 Q ，表示 master 处理任务之间的联系：

$$Q^{k+1} = \arg \min_Q L_\rho(w_1^{k+1}, w_2^{k+1}, \dots, w_T^{k+1}, Q, \mu_1^k, \mu_2^k, \dots, \mu_T^k)$$

最后是 $\mu_i, i = 1, 2, \dots, T$ 的 update。

上面是同步的，可以借助《Asynchronous Distributed ADMM for Consensus Optimization》这篇文章的思想来做异步。