Distributed Box-Constrained Quadratic Optimization for Dual Linear SVM

Distributed environment

First method with global linear convergence ($O(\log(1/\epsilon))$ iterations) for distributed dual linear L1-loss SVM.

Key: cheap line search for dual variables’ update, reducing the number of communication rounds needed.

Training speed significantly better than other methods.

Linear SVM (Boser et al., 1992; Vapnik, 1995)

Given $\{(x_i, y_i)\}_{i=1}^n \in \mathbb{R}^d \times \{-1, 1\}$, $C > 0$

$$\min_{w, \alpha} \frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i$$

subject to $y_i (w^T x_i + b) \geq 1 - \xi_i$, $\xi_i \geq 0$

$support vectors (s.v.)$ $\mathcal{S} = \{i \mid \xi_i > 0\}$

$$w = \sum_{j \in \mathcal{S}} \alpha_j y_j x_j$$

$$b = \frac{1}{|\mathcal{S}|} \sum_{j \in \mathcal{S}} y_j$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

$$\|w\|_2^2 = \frac{1}{n} \sum_{i=1}^n (y_i (w^T x_i + b) - 1)^2 + \frac{1}{n} \sum_{i=1}^n \xi_i^2$$

Compute $\Delta \alpha^t$

All machines synchronize the same

$$\Delta \alpha^{t+1} = \bar{\alpha}^{t+1} - \bar{\alpha}^t + \eta \Delta \bar{\alpha}^t$$

Compute $\eta$ (i): Armijo backing line search

given $\beta, \sigma \in (0, 1)$, $\eta = \max_{\beta \geq 0} \beta^k$ such that

$$f(\alpha^t + \beta \Delta \alpha^t) - f(\alpha^t) \leq \beta^k \sigma \nabla f(\alpha^t) \Delta \alpha^t$$

Closed-form solution by using (2), (3) and taking logarithm

$$\eta = \min(\beta \geq 0 \mid \beta^k \sigma \nabla f(\alpha^t) \Delta \alpha^t)$$

$$\hat{k} \equiv \left\lfloor \frac{1}{\log \beta} \left[ \frac{2}{\Delta f(\alpha^t) \Delta \alpha^t} + \log((\sigma - 1) \nabla f(\alpha^t) \Delta \alpha^t) \right] \right\rfloor$$

Compute $\eta$ (ii): Exact Solution

$$\frac{\partial f(\alpha + \eta \Delta \alpha)}{\partial \eta} = 0$$

Feasible solution: $\eta = \min \{\eta > 0 \mid \alpha^t + \eta \Delta \alpha \leq Ue\}$

Use allreduce to communicate across machines to get

$$\Delta w^{t+1} \equiv \sum_{k=1}^K (Y_k x_k)^T \Delta \alpha_k^t$$

Conclusions

An efficient algorithm for distributedly training linear SVM in the dual.

Superior to state of the art both theoretically and practically.
