Robust and Effective Factorization Machines

Anonymous

Optimization Algorithm

The original objective for classsification task takes the form:

$$\min_{\mathbf{w}\in\mathbf{R}^d,\mathbf{Z}\in\mathbf{S}_+^{d\times d}} \sum_{i=1}^{n} e_i(\max\{\max(y_i(\mathbf{w}^\top\mathbf{x}_i + \langle\mathbf{Z},\mathbf{x}_i\mathbf{x}_i^\top\rangle),0) - \epsilon_1, 0\})^2 + \frac{\alpha}{2}\|\mathbf{w}\|^2 + \sum_s \min\{\lambda_s^2, \epsilon_3\}, \tag{1}$$

$$e_i = \begin{cases} \frac{1}{2error}, & 0 < error \le \epsilon_2; \\ 0, & otherwise \end{cases}$$

where $error = \max(y_i(\mathbf{w}^\top\mathbf{x}_i + \langle\mathbf{Z},\mathbf{x}_i\mathbf{x}_i^\top\rangle),0) - \epsilon_1$
The subgradient with respect to $\mathbf{Z}$ is

$$\bigtriangledown_{\mathbf{Z},I} = \sum_{i=1}^{b}\mathbf{x}_i\mathbf{x}_i^\top + \beta\mathbf{P}_M\mathbf{P}_M^\top\mathbf{Z} \tag{2}$$

To incrementally calculate the SVD of $\mathbf{Z} - \eta\bigtriangledown_{\mathbf{Z},I}$. Let the symmetric and low rank matrix $\mathbf{Z}$ has rank k and its economy SVD is $\mathbf{Z} = \mathbf{P}_k\Sigma_k\mathbf{P}_k^\top$. As matrix $\bigtriangledown_{\mathbf{Z},I}$ is symmetric and low rank, we can represent it as $\mathbf{A}\mathbf{A}^\top$.

$$\begin{aligned} \bigtriangledown_{\mathbf{Z},I} &= \mathbf{X}\mathbf{X}^\top + \beta\mathbf{P}_M\mathbf{P}_M^\top\mathbf{Z} \\ &= \mathbf{X}\mathbf{X}^\top + \beta \end{aligned} \tag{3}$$

Experimental Results

Dataset	#Training	#Test	#Feature	#class
Magic04	12680	6340	10	2
w8a	49749	14951	300	2
IJCNN	49990	91701	22	2
Covtype	387342	193670	54	2
MNIST	60000	10000	784	10
epsilon	400000	100000	2000	2

Table 1: Summary of datasets used in our experiments.