= dimension, = # of triplets, = data density, = row density, = # of SGD projections.

Data5 boosting7

1Boyd et al. *Convex Optimization*, 2014, 2Jacob et al. *Group lasso with overlap and graph lasso*, 2009, 3Obozinski et al. *Group lasso with overlaps: the latent group lasso approach*, 2011,

Evaluated COMET with three datasets: Object recognition (Caltech256, trained with 135k triplets), text classification (RCV1, 4 classes subset, trained with 100k triplets), bio-informatics (Protein-LIBSVM, trained with 20k triplets).

Using an overlapping decomposition [Jacob 2009, Obozinski 2011] of into group components: is a diagonal matrix, each is a symmetric matrix of non-zero values only on the row and column, with an all-zeros diagonal,

At each coordinate step, we solve the following proximal problem, which admits a closed form solution.

\begin{equation}\label{eq:prox}

V\_k^{new} = \argmin\_{\mat{V} \in \mathcal{V}\_k} \left\langle \frac{\partial{L (V\_k)}}{\partial V\_k}, \mat{V} \right\rangle + \frac{1}{2\theta}\|V - V\_k\|\_F^2 + \lambda \|V\|\_F,

\end{equation}

and set $\W^{new} = \W + V\_k^{new} - V\_k$, where $\theta$ corresponds to the step size of the proximal update.

We add a group-sparse norm penalty to the loss to encourage solutions with fewer features and obtain the following objective:

\begin{equation}

L(\Vg) =

\min\_{\Vg} \sum\_{\trip \in \cal{T}} l\_{\W}(\qt, \pt^+, \pt^-) - \alpha \log \det(\sum\_{k=0}^{d}{\Vk}) + \lambda \sum\_{k=1}^d \|V\_k\|\_F \quad.

We use an *overlapping decomposition* of into group components: The matrix is a diagonal matrix, and each matrix is a symmetric matrix of non-zero values only on the row and column, with an all-zeros diagonal.

is the sum

$\C^\* = \Wscalar^{new} \in \R$ (a scalar), $\B^\* = \Wvec^{new} \in \R^{d-1}$ (a column vector) and $A^\* = \newW\_{2:d,2:d} = \W\_{2:d,2:d}\in \R^{(d-1)}$.

(a scalar), and

We aim to optimize the following regularized objective:

Z

\min\_{\W} \sum\_{\trip \in \cal{T}} l\_{\W}(\qt, \pt^+, \pt^-) - \alpha \log \det(\W) + \frac{\beta}{2} \frobsq{\W},

We learn a metric over a set of entities such as images or text documents, based on their relative pairwise similarities. We measure the similarity of two samples using a bilinear form parametrized by a model , .

The similarity between two data points and through the matrix , is equivalent to an Euclidean inner product in the transformed space .

. zzzz