Summary of "Generalized Low Rank Approximations of Matrices"

Matt Dickenson `mcd31`
STA571/CS590.01 Due: 7 March, 2014

Ye (2004) introduces a novel approach for computing generalized low rank approximations of matrices (GLRAM). This approach is applicable when the data of interest is a sequence of matrices A, which each matrix $A_i \in \mathbb{R}^{r \times c}$ representing one datum (e.g. an image). The goal of the method is to compute a reduced representation of A using two matrices with orthonormal columns, $L \in \mathbb{R}^{r \times \ell_1}$ (the left-side transformation) and $R \in \mathbb{R}^{c \times \ell_2}$ (the right-side transformation), and n matrices $D_i \forall A_i \in A$. The researcher sets ℓ_1 and ℓ_2 to control the quality of the approximation and the data compression ratio $\frac{nrc}{r\ell_1 + c\ell_2 + n\ell_1\ell_2}$. The desired L, R, and D are solutions to the optimization problem:

$$\min \sum_{i=1}^{n} ||A_i - LD_iR^\intercal||_F^2$$

that is, to minimize the total squared reconstruction error over the sequence of matrices.

The approximations of L and U can be computed iteratively by an algorithm that is monotonically decreasing in the root mean squared reconstruction error (RMSRE), where

$$\text{RMSRE} \;=\; \sqrt{\frac{1}{n}||A_i - LD_iR^\intercal||_F^2}.$$

The total time complexity for computing this approximation is $O(I(r+c)^2 \max(\ell_1, \ell_2)n)$. As long as the entries A_i are "nearly square" so that $r \approx c$ and $n < N$, GLRAM will be faster than SVD with its complexity of $O(n^2N)$.

When and why might we employ dimensionality reduction methods such as GLRAM? Typically this becomes a concern when analyzing high-dimensional data (e.g. images, videos) pushes us up against the limits of extant computing power. When there is too much data to hold in memory for analysis, image reduction can compress the size while maintaining close approximations of the original data. More generally, data compression is of interest whenever there may be redundancy in our data (e.g. removing stop words and compressing large text files made up of a known vocabulary). These methods are not without their costs, however. For example, Ye does not theoretically motivate the selection of two key parameters, ℓ_1 and ℓ_2, despite the influence they have on the resulting approximation and how this choice may interact with the methods that are subsequently applied to the approximated data.