Abstract

We present a new algorithm for finding a near optimal low-rank approximation of a matrix A in $O(nnz(A))$ time. Our method is based on a recursive sampling scheme for computing a representative subset of A's columns, which is then used to find a low-rank approximation. This approach differs substantially from prior $O(nnz(A))$ time algorithms, which are all based on fast Johnson-Lindenstrauss random projections. It matches the guarantees of these methods while offering a number of advantages. Not only are sampling algorithms faster for sparse and structured data, but they can also be applied in settings where random projections cannot. For example, we give new single-pass streaming algorithms for the column subset selection and projection-cost preserving sample problems. Our method has also been used to give the fastest algorithms for provably approximating kernel matrices.