EPFL

Prof. Dr. D. Kressner
D. Persson

Leverage score sampling

▶ Setting

The goal of this project is to explore the leverage score sampling for low-rank approximation. Let $A \in \mathbb{R}^{m\times n}$ and denote by $a_1, \dots, a_n$ its columns. The *leverage scores* are defined as

$$\ell_i(A) := a_i^T (AA^T)^\dagger a_i, \tag{1}$$

where $\dagger$ denotes the Moore-Penrose inverse of a matrix. The *ridge leverage scores* are defined as

$$\ell_{i,\lambda}(A) := a_i^T (AA^T + \lambda^2 I)^{-1} a_i \tag{2}$$

for a suitable regularization parameter $\lambda > 0$.

Sampling columns of A proportionally to the ridge leverage scores gives good approximation results in theory and in practice; however, computing them is expensive, so algorithms to approximate them have been developed, see e.g. [1].

▶ Tasks

1. Prove the following properties of the leverage scores (1):

   a. $\ell_i(A) \le 1$;

   b. $\sum_{i=1}^n \ell_i(A) = \text{rank}(A)$;

   c. If a column a_i is orthogonal to all other columns then $\ell_i(A) = 1$;

   d. If A has rank k then considering a thin SVD of $A = U_k \Sigma_k V_k^T$, the leverage scores are the squared norms of the rows of V_k (recall that sampling with respect to squared norms of the rows of V_k was analyzed in Lecture 7).

2. From what you proved in point 1, it follows that if A has full column rank then all the leverage scores are equal, which corresponds to uniform column sampling; however, in many situations this is not a good idea, so in practice it is useful to consider the ridge leverage scores (2). Work out an expression of $\ell_{i,\lambda}(A)$ in function of the singular values of A and the matrix V of the right singular vectors of A.

   To get an intuition of why ridge leverage scores make sense, consider the following situation. An $m \times n$ matrix A (with $m \ge n$) has a large gap between the kth and $(k+1)$th singular values, and we choose λ such that

   $$\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_k \gg \lambda \gg \sigma_{k+1} \ge \cdots \ge \sigma_n > 0.$$

   Argue that the leverage scores are all ones, but the ridge leverage scores are *close* to the squared norms of the rows of the matrix V_k from the truncated SVD of A.

3. Implement the following sampling strategies:

   a. Ridge leverage score sampling with $\lambda = 10^{-4}$;

   b. Uniform sampling;

   c. Sampling proportionally to the squares of the norms of columns of A.

   Consider the 100×100 Hilbert matrix A. For ranks $k = 1, \dots, 50$, for each of the strategies above, sample k columns C (independently, with replacement) and plot the errors $\|A - CC^\dagger A\|$. Compare these errors with the $(k+1)$th singular value of A. As these strategies are randomized, plot the average error that you obtain running each sampling strategy for 20 times.

▶ References

[1] Cohen, M. B., Musco, C., and Musco, C. Input sparsity time low-rank approximation via ridge leverage score sampling. In Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 1758–1777 (2017)