25th International Congress on Modelling and Simulation, Darwin, NT, Australia, 9 to 14 July 2023
mssanz.org.au/modsim2023

A continuous optimization algorithm for column subset selection

A. Mathur [a], S. Moka [a] and Z. Botev [a]

[a]*School of Mathematics and Statistics, UNSW Sydney, New South Wales, 2052*

Email: anant.mathur@unsw.edu.au

Abstract: Recent advances in the technological ability to capture and collect data have meant that high-dimensional datasets are now ubiquitous in the fields of engineering, economics, finance, biology, and health sciences to name a few. In the case where the data collected is not labeled it is often desirable to obtain an accurate low-rank approximation for the data which is relatively low-cost to obtain and memory efficient. Such an approximation is useful to speed up downstream matrix computations that are often required in large-scale learning algorithms.

The Column Subset Selection Problem (CSSP) is a tool to generate low-rank approximations based on a subset of data instances or features from the dataset. The chosen subset of instances or features are commonly referred to as "landmark" points. The choice of landmark points determines how accurate the low-rank approximation is.

More specifically, the challenge in the CSSP is to select the best k columns of a data matrix $\mathbf{X} \in \mathbb{R}^{m\times n}$ that span its column space. That is, for any binary vector $s \in \{0,1\}^n$, compute

$$\operatorname*{argmin}_{\mathbf{s}\in\{0,1\}^n} \|\mathbf{X} - \mathbf{P}_s\mathbf{X}\|_F^2, \quad \text{subject to } \|s\|_0 \leq k,$$

where $\|\cdot\|_F$ is the Frobenius matrix norm, $\|s\|_0 = \sum_{j=1}^{n} I(s_j = 1)$ is the number of non-zero elements elements in s and $\mathbf{P}_s$ is the projection matrix onto span$\{x_j : s_j = 1, j = 1, \ldots, n\}$ (x_j being the j-th column of $\mathbf{X}$).

It is well known that choosing the best column subset of size k is a difficult combinatorial problem. In this work, we show how one can approximate the optimal solution by defining a penalized continuous loss function that is minimized via stochastic gradient descent. We show that the gradients of this loss function can be estimated efficiently using matrix-vector products with a data matrix $\mathbf{X}$. We provide numerical results for a number of real datasets showing that this continuous optimization is competitive against existing methods.

Keywords: *Low-rank approximation, dimensionality reduction, feature selection*