

Online Algorithms for Principal Component Analysis

Dan Garber, Technion

Principal Component Analysis is a widely used pre-processing technique in machine learning for reducing the dimension of the data and cleaning noise.

In standard PCA, the input to the problem is a set of d dimensional vectors x_1, \dots, x_n and a target dimension $k < d$; the output is a set of k dimensional vectors y_1, \dots, y_n that best capture the top singular directions of the original vectors.

In the online setting, the vectors x_t are presented to the algorithm one by one, and for every presented x_t the algorithm must output a low-dimensional vector y_t before receiving x_{t+1} . This setting is interesting for instance in case that the PCA is part of an online learning pipeline and the low-dimensional vectors y_t are fed to an online learning algorithm.

We present the first approximation algorithms for this setting of online PCA. Our algorithm produces vectors of dimension $k \cdot \text{poly}(1/\epsilon)$ whose quality admit an additive ϵ approximation to the optimal offline solution allowed to use k dimensions.