$\mathcal{U}$-Bootstrap percolation : critical probability, exponential decay and applications, by Ivailo HARTARSKY

Leo Davy Martin Gjorgjevski Aleksandr Pak

ENS Lyon
M2 Advanced Mathematics

March 2022

Principal Component Analysis (PCA)

PCA

The goal is to find the linear subspace P_k that best fits the d-dimensional data $\{x_i\}_{i=1}^N$ in the LS sense, i.e., find an orthogonal family of k vectors $\{u_l\}_{l=1}^k$ that maximizes

$$\sum_{l=1}^k \sum_{i=1}^N |u_l^T x_i|^2.$$

A solution is the k-principal eigenvectors of the empirical autocorrelation matrix

$$\hat{R} = \frac{1}{N}\sum_{i=1}^N x_i x_i^T =: \frac{1}{N}\sum_{i=1}^N \Phi(x_i).$$

$\hat{R}$ is a *sketch* of our data (of dim d^2).

The sketch

$$\hat{R} = \frac{1}{N}\sum_{i=1}^{N} x_i x_i^T =: \frac{1}{N}\sum_{i=1}^{N} \Phi(x_i) \in \mathbb{R}^{d^2}$$

is a very compressed version of the data $\{x_i\}_{i=1}^N$, *but*, it still contains the geometry of the data.

CS inspired idea

Take m random measurements[a] of each sample and use the sketch defined by $\Phi(x) = \mathcal{M}(xx^T)$. Provided $m > kd$, the principal eigenvectors can be recovered.

[a] $\mathcal{M} : \mathbb{R}^{d \times d} \to \mathbb{R}^m$ satisfying RIP on matrices of rank at most $2k$.