RANDOM PROJECTIONS AND DIMENSION REDUCTION

SUMMER @ ICERM 2020

1. Dimension reduction

Suppose we are given a (possibly large) set of points $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$. We are typically interested in the cases when both n and d are very large. Data analysis typically performs algebraic operations on these points; such analysis might involve regression, clustering, and/or classification. The large values of n and d can make this task very difficult. For example, some standard linear algebra operations on a matrix whose columns are defined by each point would require complexity on the order of $nd \min(n, d)$, which is not tractable if $n, d \gg 1$. The main strategy to overcome these computational limitations is to make n or d smaller.

The case when n is large corresponds to having a lot of data; while we could consider summarizing this data into fewer points, we will adopt the point of view here that more data is better and summarization is not an attractive approach.

Alternatively, we can try to reduce the ambient dimension d of the data points. For example, if we considered only the first m coordinates in each d-dimensional vector, this would be one way to reinterpret each point as an m-dimensional vector. However, reducing the dimension of these points in arbitrary ways is sure to affect quantitative analysis and interpretability of the lower-dimensional data. How can we construct dimension reduction procedures that respect the original geometric nature of the data?

2. The Johnson-Lindenstrauss Lemma

In order to preserve the geometric structure of the point set X, we might wonder if there is a procedure that embeds our n d-dimensional points into an m-dimensional space, with $m \ll d$, that preserves pointwise distances. Ideally, it would preserve distances exactly, but this is probably too much to hope for. (To keep the discussion simple, let us assume "distance" refers to Euclidean distance.) Instead, we might be willing to allow a small deviation in the distances.

More precisely, let $f : \mathbb{R}^d \to \mathbb{R}^m$ be our (unspecified) dimension reduction procedure. Then each original data point x_j is mapped to a new data point $y_j = f(x_j)$ that lies in $\mathbb{R}^m$. In order to nearly preserve distances, we suppose that for some $\epsilon \in (0, 1)$ we might be able to achieve

$$(1 - \epsilon)\|x_i - x_j\| \le \|y_i - y_j\| \le (1 + \epsilon)\|x_i - x_j\|, \qquad i, j = 1, \ldots, n. \tag{1}$$

Thus, we allow small, ϵ inflations or deflations of the projected distance relative to the original distance. One expects that this is not possible for m and ϵ both arbitrarily small, and instead that some balance must be struck. This balance is precisely what the Johnson-Lindenstrauss Lemma delivers.

Lemma 2.1 ((Johnson-Lindenstrauss [3])). *Let X, a set of n points in $\mathbb{R}^d$, be given. Let $\epsilon \in (0,1)$ and m satisfy*

$$m\epsilon^2 > 8 \log n.$$

Then there exists a linear *map $f : \mathbb{R}^d \to \mathbb{R}^m$ such that* (1) *holds.*

This Lemma is one mathematical cornerstone of dimension reduction methods, providing a quantitative description of what can be achieved with near-isometric dimension reduction.

3. Random projections

The statement of the Johnson-Lindenstrauss Lemma is quite a positive one, but does not seem to give guidance on how to construct the provided linear function f. However, since f is linear, then it must be representable as a matrix, i.e.,

$$y = f(x) = Ax, \qquad\qquad A \in \mathbb{R}^{m \times d}.$$

Then the question is how such a matrix can be constructed. One of the proof techniques of the Johnson-Lindenstrauss Lemma is actually constructive: A can be constructed by generating the matrix *randomly*, i.e., by using a matrix A whose entries are certain types of independent and identically-distributed draws of a random variable. See, for example, [2]. Since A is random, this randomized approach does not succeed always, but does succeed *with high probability*.

There is an attractive geometric interpretation of $f(x) = Ax$ where A is a random matrix: each component of the output Ax is equal to the inner product between x and a random vector (rows of A). Thus, Ax essentially contains coordinates of a vector that correspond to the projection of x onto a *random* subspace. Thus, this procedure actually projects the full data set X onto a random subspace. It may seem bizarre that this strategy succeeds in doing anything, but the proof of the Johnson-Lindenstrauss Lemma reveals that the randomness here exploits a deep result known as *concentration of measure*, which is what ensures that this random approach works.

After the data set X is reduced to $Y = \{y_1, \ldots, y_n\}$, one typically proceeds with standard algorithmic approaches for, e.g., clustering and regression, in order to analyze and manipulate the data set.

4. Project outlook

This project involves the following investigations:

- analysis of the Johnson-Lindenstrauss lemma, and investigation of its proof
- numerical implementation and investigation of randomized projections for reducing data
- application of randomized projection methods for data analysis and comparison against, for example, principal component analysis
- investigation of randomized linear algebra methods for large-scale matrix problems
- numerical investigation of alternative randomized methods for nonlinear dimension reduction

Students will program in either Matlab or Python to focus the numerical investigation. Some good starting points for learning about randomized projection methods for dimension reduction are [5, 4, 1]

References

1. N. Halko, P. G. Martinsson, and J. A. Tropp, *Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions*, SIAM Review **53** (2011), no. 2, 217–288.
2. Nick Harvey, *Cpsc 536n: Randomized algorithms, lecture 6*, https://www.cs.ubc.ca/ nickhar/W12/Lecture6Notes.pdf.
3. William Johnson and J. Lindenstrauss, *Extensions of Lipschitz mappings into a Hilbert space*, Conference in Modern Analysis and Probability **26** (1982), 189–206.
4. Michael W. Mahoney, *Randomized Algorithms for Matrices and Data*, Foundations and Trends® in Machine Learning **3** (2011), no. 2, 123–224, arXiv: 1104.5557.
5. Haozhe Xie, Jie Li, and Hanqing Xue, *A survey of dimensionality reduction techniques based on random projection*, arXiv:1706.04371 [cs] (2018), arXiv: 1706.04371.