Principle Component Analysis (PCA)

Dreycey Albin

November 2019

The key need for Principle Component Analysis (PCA) is to be able to linearly reduce the dimensions of data while preserving higher dimensional distances. This is important because other techniques would require looking at each of the features for all of the data separately, and this may not yield useful information. More specifically each axis could be plotted against one another, in a cartesian product like fashion, which could give the correlations between features, but this is inefficient [2].

The key need for random projections (RPs) was the time complexity associated with computing the eigen vectors and values using matrix decomposition in PCA. The random projection technique thereby lowers the computational complexity, increasing the speed for dimensional reduction [1].

They key object or solution for PCA is the idea of choosing orthonormal (orthogonal and unit vectors/normalized) basis vectors that have a direction in the area that has the most variance- which are the eigen vectors of the covariance matrix. The first PCA axis is the basis for the direction with the most variance, followed by PCA axis 2 with the second most and so on. The loads represented by the eigen values of the decomposition give the relative importance of each feature or axis for that particular set of data [2].

They key object or solution for random projections is the idea that most vectors of high dimensional vectors are close to orthognal. Therefore, random projections onto these dimensions should be nearly orthogonal. If the random projections are projected onto a suitably high dimension, then the relative distances are preserved [1].

Non-optimal results may be given based on the way PCA finds the orthogonal vectors. Essentially, orthogonal vectors are randomly chosen by having the PCA algorithm choose vectors orthogonal to the primary PCA vector. This problem arises when the data set is non-Gaussian or the second primary axis is not directly perpendicular/orthogonal. Lastly, the eigen value decomposition is computationally intensive to compute [2].

One limitation, stemming from the Johnson-Lindenstrauss lemma, seems to be the dimensionality in which the data can be mapped. For example, one can map data down to 2 dimensions using PCA, but it seems like the dimension must still be larger for random projections. To me, this seems like a limitation, especially if the goal is to be able to visualize the higher dimensional distances using 2 axis'. [1].

The connection between matrix decomposition and finding the Eigen vectors of the co-variance matrix is really intriguing. This treats the co-variance matrix as a linear transformation matrix. From here, there is a non-complex Eigen value matrix calculated - non-complex since the covariance matrix (or linear transformation matrix) is symmetric. The entire space is sheared and stretched, and from this, the vectors that don't move (i.e. eigen vectors) are pointed in the direction of the most variance - pretty neat! One interesting thought is the idea of a non-linear transformation matrix, and how deriving this could potentially ameliorate the negative effects/limitations noted in the paper [2].

Something I found interesting was the bounds suggested for the random projections (RPs). They calculated that 1600 dimensions would be required, but found that only 50 are truly needed. This can drastically reduce the number of dimensions used! This has me curious about how RP and PCA could potentially be used for regularization [1].

References

[1] Ella Bingham and Heikki Mannila. Random projection in dimensionality reduction: applications to image and text data. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 245–250. ACM, 2001.

[2] Jonathon Shlens. A tutorial on principal component analysis. *arXiv preprint arXiv:1404.1100*, 2014.