University of Washington

CSEP 521 Final Project

Making PCA Robust

Jonathan Donovan & Si Latt

February 25, 2017

1 Abstract

In this paper we analyze the performance of PCA and robust PCA on datasets that are artificially made noisy. In particular, we use a method[1] which attempts to decompose the corrupted matrix into a low rank and a sparse component. The low rank component can then be used to recover the principle components for the data being analyzed. We attempt this on facial data (to recover eigenfaces) and on video data (as background recovery) and examine the performance on each of these methods with varying degree of corruption.

2 Introduction

2.1 PCA

$$PCA(M,k) = \arg\min_{M'}(||M - M'||_F) \mid rank(M') \leq k$$

Principle Component Analysis[2] (or PCA) is the attempt to approximate a matrix M with a nearby (in the Frobenius norm) low rank matrix M'. As a reminder, the rank of a matrix M is the number of independent vectors that are needed to recreate the columns of the matrix, and the Frobenius norm of a matrix is a generalization of the euclidean norm for vectors, where $||A||_F = \sqrt{\sum_{a_{i,j} \in A} |a_{i,j}|^2}$.

When the columns of the matrix we are performing PCA on represent a series of measurements (usually with the average subtracted out), with PCA we are essentially trying to decompose the measurements into linear independent 'hidden variables'. If we have reason to believe that our measurements are the result of a smaller subset of independent variables, PCA gives us the tools needed to find the best set of these variables.

PCA would not be useful if it were costly to compute, but thankfully, the problem formulation listed above is intimately related to a common operation in the analysis of matrices, that is: the Singular Value Decomposition. The Singular Value Decomposition (SVD) problem is a problem with the goal of decomposing a matrix into $M = U\Sigma V^*$, where U and V are unitary matrices ($UU^* = I$ and $V^*V = I$) and Σ is a real diagonal matrix with the entries sorted. While the factorization may not be unique, the diagonal matrix is, and it can be shown that the best k-rank approximation of a matrix M can be found as $M' = U\Sigma' V^*$ where Σ' is the diagonal matrix Σ with all entries $\sigma_{i,i} = 0$ for $i > k$. It suffices here to say that the SVD problem is well studied, and that solutions can be found quite quickly in $O(mn^2)$ time for an $m \times n$ matrix. Probabilstic methods exist which can perform even faster when $k << m, n$.

2.2 Robust PCA

Robust PCA[1] is an extension of PCA to the domain of corrupted data. PCA can quite easily recognize data where redundancy in the information of the data points exists, but it does not properly account for the fact that the collection of data itself may be lossy. One circumstance in which PCA fails is where significant outliers exist in the data, where the inclusion of a single dramatic outlier may have a dramatic impact on the estimated principle components. In a world where massive amounts of data exist to be sample, but where the quality of the data may very significantly, a variation on PCA which incorporates robustness to these problems adds significant utility when analyzing data with the intent of discovering low rank approximations.

In the paper "Robust principal component analysis?"[1] such a method is outlined, called Principal Component Pursuit (PCP). PCP breaks down the finding of the solution into decomposing a given matrix M into two components L and S via the equation:

$$L, S = \arg\min_{L,S}(||L||_* + \lambda||S||_1) \mid L + S = M$$

where $||A||_*$ is the nuclear norm, defined as the sum of the singular values of M, or $\sum_{\sigma \in SVD(M)} \sigma$, and $||A||_1$ is the sum of the absolute values of the entries of A, or $\sum_{a_{i,j} \in A} |a_{i,j}|$. Amazingly, as part of the analysis the authors determine that there is a choice of parameter $\lambda = \frac{1}{\sqrt{n}}$ that succeeds with fairly high probability. To be specific this algorithm succeeds to capture the to perfectly capture the original matrices L and S with $P \geq 1 - \frac{c}{n^1 0}$ provided the conditions that (for an $n \times n$ matrix)

$$rank(L) \leq \frac{p_r n}{\mu(logn)^2} \text{ and } m \leq p_s n^2$$

Here p_r and p_s are positive constants, μ is a measure of incoherence[3] (with small μ it is required that the significant values are not sparse), and m is the support set of the matrix S_0 (or equivalently, the number of nonzero entries in S_0). Interestingly, this shows that the number of corrupted entries can scale linearly with the size of the matrix, and that with a fixed μ, the rank of the low rank component L_0 can grow as $\frac{n}{log(n)^2}$. Another interesting point to note is that while we have constraints on the number of corrupted entries (the support set of S_0) we have no such constraints on how the data is corrupted (that is, it could be corrupted with arbitrarily large absolute value errors). When in fact our goal in many circumstances is to recover a fixed low rank approximation from a large amount of (probabilistically corrupted) data, these constraints work perfectly!

2.3 Metrics

The remainder of our paper will be spent analyzing the performance of Robust PCA relative to PCA on various data sets. In order to do so we would like to define a few metrics that would be useful when discussing the accuracy of Robust PCA in reproducing the original low rank matrix.

2.4 Frobenius error

One useful metric, and the one used by the original paper, was the metric $\frac{||L_0 - L||_F}{||L_0||_F}$. This is a reasonable metric of the error of the approximation, as it is the sum of square errors of the approximation over the sum of squares of the values of the original matrix, giving a natural sense of 'how far' the resulting approximation lies from the original matrix.

2.5

References

[1] Candès, Emmanuel J., et al. "Robust principal component analysis?." Journal of the ACM (JACM) 58.3 (2011): 11.

[2] C. Eckart and G. Young. The approximation of one matrix by another of lower rank. Psychometrika, 1(3):211–218, 1936.

[3] E. J. Cand'es and B. Recht. Exact matrix completion via convex optimzation. Found. of Comput. Math., 9:717–772, 2009.