

Randomization as Regularization

Mack Yi
Duke University

Objectives

Randomized algorithms for dimension reduction are interesting not only for their computational performance, but also for their statistical performance. Following up on recent work showing regularization behavior of randomized Singular Value Decomposition (SVD), the objectives of this project were to:

- Empirically simulate/observe the regularization behavior of randomized SVD in Principal Component Analysis (PCA) and Principal Components Regression
- Explore the theoretical questions which arise surrounding this regularization behavior.

Introduction

In the regime of large data sets which have both a large number of datapoints as well as a large number of features, traditional machine learning algorithms can become infeasible due to computational constraints. One approach to mitigating this problem has been in the area of dimension reduction algorithms, such as Principal Component Analysis (PCA). When datasets are very massive, the dimension reduction algorithm can itself become computationally infeasible. This has inspired the development of randomized dimension reduction algorithms, which can provide even better computational performance.

However, not only do these randomized algorithms offer computational benefits, but they also appear to offer statistical benefits. Previous work by [Darnell et al.(2015)] and [Georgiev & Mukherjee(2012)] have observed regularization behavior of randomized Singular Value Decomposition in Linear Mixed Models and (Localized) Sliced Inverse Regression, respectively. Here, the goal is to explore the use of randomized SVD in Principal Components Regression (PCR) as regularization, with an eye towards establishing a theoretical framework for the guarantees which might exist for the generalization error of algorithms using randomized SVD.

Methods

The Randomized SVD algorithm is as follows. This is a shortened version of the Adaptive Randomized SVD described by [Darnell et al.(2015)].

1. Find an orthonormal basis for the range of \mathbf{X} :

- Set the number of working directions: $l = d_{max} + \Delta$
- Generate random matrix $\Omega \in \mathbb{R}^{n \times l}$ with $\Omega_{ij} \stackrel{iid}{\sim} N(0, 1)$
- Construct blocks $F^{(t')} = XX^T F^{(t'-1)}$, $F^{(0)} = \Omega$ for $t' \in 0, \dots, t$
- Estimate basis for block using QR decomposition $F = QR$

2. Project data onto the range basis and compute SVD:

- Project onto the basis: $B = X^T Q \in \mathbb{R}^{r \times l}$
- Factorize $B \stackrel{svd}{=} U \Sigma W^T$
- Take the rank d^* approximation using the first d^* columns of U , the first d^* rows and columns of Σ , and QxW_{d^*} , where W_{d^*} is the first d^* column of W .

Results

The regularization behavior of the randomized SVD algorithm is studied by observing the behavior of PCA and PCR. Plugging a randomized SVD into PCR is straightforward. The Principal Component Analysis step is replaced by PCA using randomized SVD, and this PCA can then be used in PCR.

Low rank data is simulated using the same method used by [Darnell et al.(2015)]. The data matrix $X \in \mathbb{R}^{n \times p}$ is generated as $X = USV^T + E$, where $U^T U = V^T V = I_{d^*}$. The columns of U and V are drawn uniformly at random from the unit sphere and the ("true") singular values are generated starting from a baseline value, which is a fraction (κ) of the maximum noise singular value, with exponential increments separating consecutive entries. The noise is iid Gaussian.

The relative error of the singular values is simulated for several values of κ and the regularization parameter t controlling the number of power iterations.

Conclusion

The simulations of Principal Component Analysis corroborate the intuition that the randomization SVD algorithm is providing regularization in the estimation of the singular values, especially in the case where there is a relatively low signal to noise ratio. This result lines up well with the previous observations which prompted this project.

However, the generalization bound guarantees which might exist for this randomized SVD process remain poorly understood, as well as the effect this might have when extended to PCR. [Rudi et al.(2015)] studies the regularization behavior of different randomized subsampling strategies, and suggests that the regularization behavior and randomization methods can be related in some more rigorous manner.

References

- [Darnell et al.(2015)] Darnell, G., Georgiev, S., Mukherjee, S., & Engelhardt, B. E 2015, arXiv:1504.03183
- [Georgiev & Mukherjee(2012)] Georgiev, S., & Mukherjee, S. 2012, arXiv:1211.1642
- [Rokhlin et al.(2008)] Rokhlin, V., Szlam, A., & Tygert, M. 2008, arXiv:0809.2274
- [Rudi et al.(2015)] Rudi, A., Camoriano, R., & Rosasco, L. 2015, arXiv:1507.04717

Important Result

Simulations show randomized SVD appears to provide regularization, parameterized by the number of power iterations, in the estimation of the singular values, especially in the case where there is a relatively low signal to noise ratio.

Table 1: Percent relative error of singular values relative to "true" simulation values

Parameters	RSVD	SVD
$\kappa = 0.5, t = 1$	8.981	21.457
$\kappa = 0.5, t = 2$	10.899	21.457
$\kappa = 0.5, t = 3$	16.179	21.457
$\kappa = 1.0, t = 1$	10.831	22.222
$\kappa = 1.0, t = 2$	11.950	22.222
$\kappa = 1.0, t = 3$	15.960	22.222
$\kappa = 2.0, t = 1$	8.865	5.522
$\kappa = 2.0, t = 2$	5.000	5.522
$\kappa = 2.0, t = 3$	5.237	5.522
$\kappa = 3.0, t = 1$	10.981	4.230
$\kappa = 3.0, t = 2$	4.091	4.230
$\kappa = 3.0, t = 3$	4.216	4.230

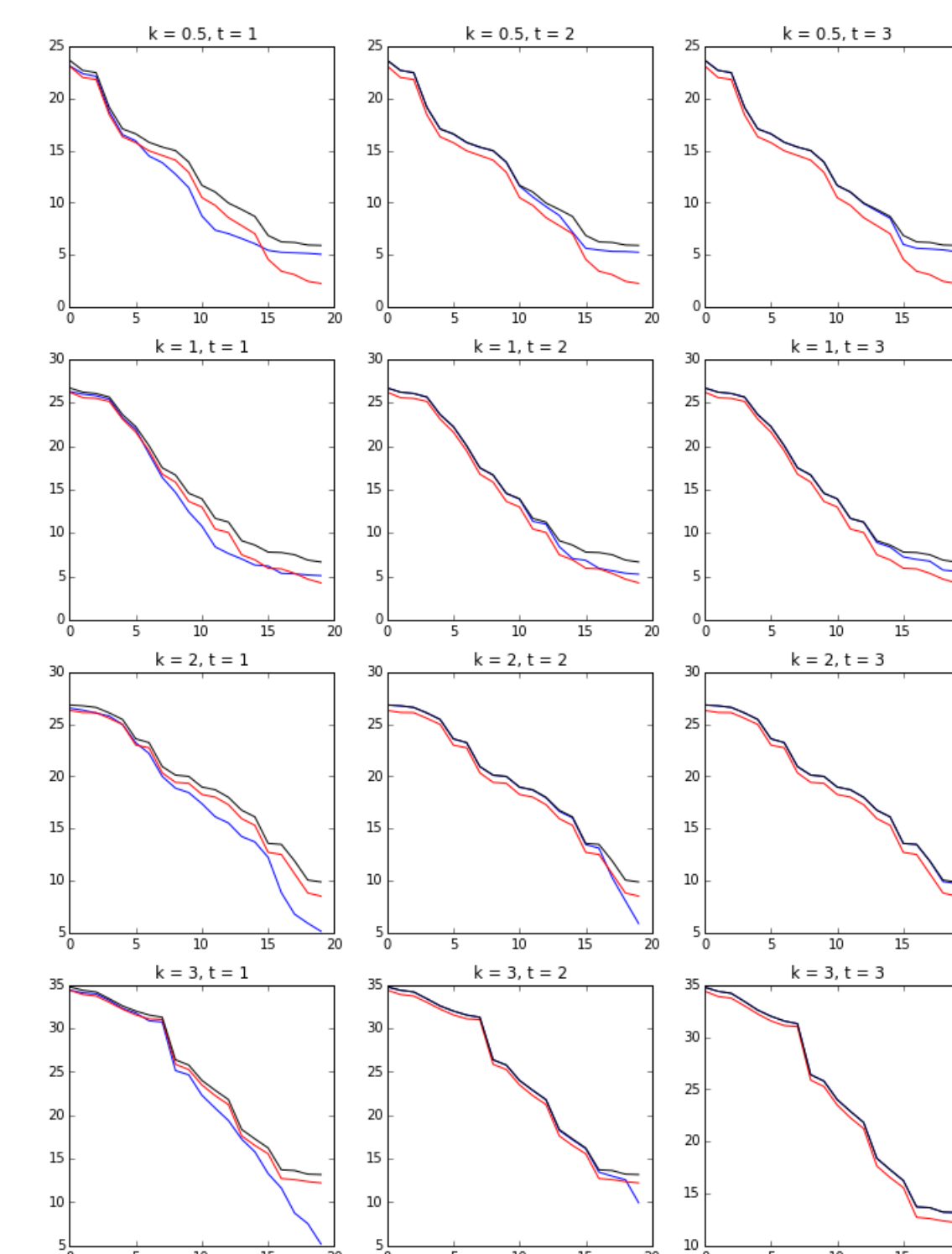


Figure 1: Singular values of the signal (red), unrandomized SVD (black), and randomized SVD (blue) for the simulated low rank data. The x axis indexes the singular values, while the y axis is the value of each singular value. X is 200×5000 with rank 20

Contact Information

- Email: mack.yi@duke.edu