Fundamentals of Mathematical and Computing Sciences:
Applied Mathematical Science

Part III: Low rank matrix estimation (Lecture 5) Advanced topics

2015-7-24

Taiji Suzuki (Room W707, post W8-46)
e-mail: suzuki.t.ct@m.titech.ac.jp

The following topic will be explained in this document:

- Efficient randomized method for SVD of very large (near) low rank matrix

1 Preliminary

We want to compute the SVD of a very large matrix A:

$$A = \underbrace{U}_{M\times d} \times \underbrace{\Sigma}_{d\times d} \times \underbrace{V^\top}_{d\times N} \in \mathbb{R}^{M\times N}.$$

Suppose a situation where M, N are very large, say 10^6, but d is not large, say 10^3.

2 Algorithm

1. Draw a $N \times d$ Gaussian random matrix Ω ($\Omega_{i,j} \sim N(0,1)$, i.i.d.).
2. Compute $Y = A\Omega \in \mathbb{R}^{M\times d}$ (Y is much smaller than the original matrix A).
3. Compute an orthonormal matrix $Q \in \mathbb{R}^{M\times d}$ ($Q^\top Q = I$) such that columns of Q spans the image of Y.
4. Compute $B = Q^\top A$ $(= Q^\top U\Sigma V^\top) \in \mathbb{R}^{d\times N}$. Note that B is a small matrix.
5. Compute SVD of B; $B = U_B \Sigma_B V_B^\top$.
6. Obtain $U = QU_B$.

Note that $QB = QQ^\top A = A$. The third step can be executed in a standard way such as the Gram-Schmidt orthonormalization.
Verification:

- The columns of Y spans the image of A almost surely. Thus the columns of Q also spans the image of A a.s..
- Therefore Q can be written as $Q = US$ for some $S \in \mathbb{R}^{d\times d}$. Here $Q^\top Q = I$ implies $S^\top U^\top US = S^\top S = I$. That is, S is orthogonal.
- Thus $B = Q^\top A = S^\top U^\top A = S^\top \Sigma V^\top$. This yields $S^\top = U_B$ and $V = V_B$. In particular, $QU_B = QS^\top = USS^\top = U$.

3 Theory

Q: What happens if the rank of A is larger than d?

Theorem 1. *For any $k \geq 2$ and $p \geq 2$ such that $k+p=d \leq \min\{M,N\}$, we have that*

$$\mathrm{E}[\|A - QQ^\top A\|_F] \leq \left(1 + \sqrt{\frac{k}{p-1}}\right) \sigma_{k+1} + \frac{e\sqrt{d}}{p} \left(\sum_{j>k} \sigma_j^2\right)^{1/2}.$$

Moreover, we have the following deviation bound.

Theorem 2. *For any $k \geq 2$ and $p \geq 4$ such that $k+p=d \leq \min\{M,N\}$, we have that*

$$\|A - QQ^\top A\|_F \leq \left(1 + t\sqrt{\frac{12k}{p}}\right) \left(\sum_{j>k} \sigma_j^2\right)^{1/2} + ut\frac{e\sqrt{d}}{p+1}\sigma_{k+1},$$

with probability as least $1 - (5t^{-p} + 2e^{-u^2/2})$.

See [1] for the details and the proofs of these theorems.

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(2):217–288, 2011.