Stat 991: Multivariate Analysis, Dimensionality Reduction, and Spectral Methods **Lecture: 23**

Matrix Concentration

Instructor: Sham Kakade

1. Hoeffding review
2. say we have a random matrix
3. two natural norms

1 Hoeffding Bound

Let $X_1, \ldots X_n$ be i.i.d. real valued random variables bounded in $[0, M]$, almost surely. Then with probability greater than $1 - \delta$,

$$|\frac{1}{n}\sum_i X_i - \mathbb{E}[X]| \leq \frac{M}{\sqrt{n}}\sqrt{\log(2/\delta)}$$

2 Matrix Concentration

2.1 Norms

Recall that the Frobenius norm of a matrix, $\|M\|_{\mathrm{F}}$, is the square root of the sum of squares of the elements of the matrix. The spectral norm of a matrix, $\|M\|_2$ is it's maximal singular value.

Note that:

$$\|M\|_2 \leq \|M\|_{\mathrm{F}}$$

2.2 Concentration

Let $X \in \mathbb{R}^{m \times n}$ be a random matrix. In many settings, we are interested in the behavior of either:

$$\left\| \frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X] \right\|_{\mathrm{F}} \leq ?, \quad \left\| \frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X] \right\|_2 \leq ?$$

where each X_i is sampled i.i.d. from some distribution. Here, $\| \cdot \|_2$ denotes the spectral norm (the largest singular value) and $\| \cdot \|_{\mathrm{F}}$ denotes the Frobenius norm.

The following theorem provides a high probability bound on these quantities.

Theorem 2.1. *Assume that $X_i \in \mathbb{R}^{d_1 \times d_2}$ are sampled i.i.d. Let $d = \min\{d_1, d_2\}$.*

- *(Frobenius Norm) Suppose $||X||_{\mathrm{F}} \leq M$ almost surely. Then with probability greater than $1 - \delta$,*

$$\left\| \frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X] \right\|_{\mathrm{F}} \leq \frac{6M}{\sqrt{n}}\left(1 + \sqrt{\log\frac{1}{\delta}}\right).$$

- *(Spectral Norm) Suppose $||X||_2 \leq M$ almost surely. Then with probability greater than $1 - \delta$,*

$$\left\| \frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X] \right\|_2 \leq \frac{6M}{\sqrt{n}}\left(\sqrt{\log d} + \sqrt{\log\frac{1}{\delta}}\right).$$

2.3 Examples

Two special case of interest, are when:

1. The samples X_t are of the form $xx^\top$ where x is a vector. Here if the Euclidean norm $\|x\|_2 \leq 1$ then $\mathcal{X}_i\|_\mathrm{F} \leq 1$.
2. Another case may be where X_t only has one entry which is 1. For example, we are estimating a probability matrix $\Pr(x_1 = i, x_2 = j)$, and each X_t is a sample (where the i, j entry being one corresponds to the event i, j occurring). Again, the Frobenius norm is bounded by one.

Instead, it might be the case that random matrix X_t has large Frobenius norm. Here, we might hope that it's spectral norm is small, in which case the latter concentration result is more appropriate.

3 Accuracy of Projections

Let us assume that $\mathbb{E}[X]$ is "low rank", say rank k. The question we ask is how accurate our projections are onto the left (or right) singular subspace using the sample matrix $\widehat{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Let the SVDs be $\mathbb{E}[X] = UDV^\top$ and $\widehat{X} = \widetilde{U}\widetilde{D}\widetilde{V}^\top$. Let

Let $\widetilde{U}$ correspond to the top k singular vectors (so it is of size $d_1 \times k$). Let $\widetilde{U}_\perp$ be the matrix whose columns are orthonormal and perpendicular to $\widetilde{U}$.

Let λ_k be the smallest (non-zero) singular value of $\mathbb{E}[X]$. Following from Stewart and Sun (theorem 4.1 and theorem 4.4, pages 260 and 264), we have that:

$$\|\sin(\text{angles between } U \text{ and } \widehat{U})\|_\mathrm{F} = \|\widehat{U}_\perp^\top U\|_\mathrm{F} \leq \frac{\|\widehat{X} - \mathbb{E}[X]\|_\mathrm{F}}{\lambda_k}$$

and

$$\|\sin(\text{angles between } U \text{ and } \widehat{U})\|_2 = \|\widehat{U}_\perp^\top U\|_2 \leq \frac{\|\widehat{X} - \mathbb{E}[X]\|_2}{\lambda_k}$$