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 $1 - \delta$,

$$\frac{1}{n} \sum_{i=1}^{n} X_i - \mathbb{E}[X] \leq \frac{M}{\sqrt{n}} \sqrt{\log(2/\delta)}$$

2 Matrix Concentration

2.1 Norms

Recall that the Frobenius norm of a matrix, $\|M\|_F$, is the square root of the sum of squares of the elements of the matrix. The spectral norm of a matrix, $\|M\|_2$ is its maximal singular value.

Note that:

$$\|M\|_2 \leq \|M\|_F$$

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:

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

where each $X_i$ is sampled i.i.d. from some distribution. Here, $\| \cdot \|_2$ denotes the spectral norm (the largest singular value) and $\| \cdot \|_F$ 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 $\|X\|_F \leq M$ almost surely. Then with probability greater than $1 - \delta$,

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

- (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 cases of interest are when:

1. The samples $X_t$ are of the form $xx^T$ where $x$ is a vector. Here if the Euclidean norm $\|x\|_2 \leq 1$ then $X_i \|F \leq 1$.

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 $E[X]$ 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 $\Hat{X} = \frac{1}{n} \sum_{i=1}^n X_i$. Let the SVDs be $E[X] = UDV^T$ and $\Hat{X} = \Hat{U}\Hat{D}\Hat{V}^T$. Let $\Hat{U}$ correspond to the top $k$ singular vectors (so it is of size $d_1 \times k$). Let $\Hat{U}_\perp$ be the matrix whose columns are orthonormal and perpendicular to $\Hat{U}$.

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

$$\| \sin(\text{angles between } U \text{ and } \Hat{U}) \|_F = \|\Hat{U}_\perp U\|_F \leq \frac{\|\Hat{X} - E[X]\|_F}{\lambda_k}$$

and

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