Spectral Gap Results & the Analysis of Sign Rank versus VC Dimension

Alexander Berenbeim

2016-11-22 Tuesday

Outline

Key Definitions Key Summary of the Paper

Key Definition: Sign Rank

For any $M \in (\mathbb{R}^{\times})^{n \times m}$, the sign matrix $\mathrm{sgn}(M)$ is defined,

$$(\mathrm{sgn}(M))_{i,j} := \mathrm{sgn}(M_{i,j})$$

for all $i \in [n], j \in [m]$. The sign rank of S is defined

$$srk(S) = \min\{rk(M) : \mathrm{sgn}(M) = S\}$$

where the rank is over the real numbers.

The Geometry of Sign Rank

By identifying columns of a matrix with points, and rows with half spaces through the origin, the sign rank captures the minimum dimension of a real space in which the matrix can be embedded using half-spaces.

Explicitly, a d-dimensional embedding of M using half spaces consists of a pair of maps (e_n, e_m) satisfying the following property:

for every row $r \in [n]$ and column $c \in [m]$, we have $e_n(r) \in \mathbb{R}^d$ and $e_m(c)$ is a half-space inside $\mathbb{R}^d$ such that $M_{r,c} = 1 \iff e_n(r) \in e_m(c)$.

Key Definition: VC Dimension (for Matrices)

Given a sign matrix S, and $C \subset \mathrm{col}(S)$, we say C is shattered

if each of the $2^{|C|}$ possible sign patterns appears in the restriction of S to the columns C.

The VC-dimension of a sign matrix S is

the maximum size of a shattered subset of columns.

Key Definition: Spectral Gaps and Regularity

A boolean matrix B is Δ-regular

if every row and column in it has exactly Δ ones.

A sign matrix S is Δ-regular if

its corresponding Boolean matrix is.

The singular values of a real-valued N$\times$ N matrix M are

non-negative $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_N \geq 0$ such that for each i there are pairs of vectors u_i, v_i such that

$$Mu_i = \sigma_i v_i \qquad M^T v_i = \sigma_i u_i$$

M has a spectral gap if

$$\frac{\sigma_2(M)}{\sigma_1(M)} = \frac{\sigma_2(M)}{\|M\|} < 1.$$

Quick summary of (most of) the paper

- The paper studies the bounds of the sign rank of square matrices given a VC dimension (and the complexity properties/results thereof).
- Whereas VC dimension relates to learnability, sign rank relates to learnability by kernel-based methods.
- The authors interpret their complexity results as describing both the universality of kernel-based classifiers, and the deficiencies of kernel-based classifiers that are inherent to real algebraic topology.

Some Motivation

- The question is a matter of how large can the sign rank be for a given VC dimension, e.g. with f denoting the maximum possible sign rank of an $N \times N$ matrix with VC dimension d by $f(N, d)$, how does $f(N, d)$ behave?
- Spectral gaps can entail pseudo-random properties (e.g. random sign matrices exhibit high sign rank);
- The existence of a maximum class of VC dimension 2 and sign rank $\tilde{\Theta}(\sqrt{N})$ follows from a general connection between sign rank and spectral gaps;
- This connection also yields a lower bound.

Forster's Theorem

(Forster's Lemma) For any $X \subset_{fin} \mathbb{R}^k$ in general position, there exists a matrix B such that

$$\sum_{x} \frac{1}{\|Bx\|^2} Bx \otimes Bx = \frac{|X|}{k} I_k.$$

(Forster's Theorem) Let S be an $N \times N$ sign matrix. Then

$$srk(S) \geq \frac{N}{\|S\|^*} \geq \frac{N}{\|S\|}$$

Limitations of the spectral norm establishing a lower bound

- Consider a $\frac{N}{3} \geq \Delta$-regular S and let x be an all 1

vector. Then $Sx = (2\Delta - N)x$, whence $\|S\| \geq N/3$. Consequently, Forster's theorem would indicate that the sign rank of S is $\Omega(1)$.

- The authors use a variant of the spectral norm they call the star norm,

$$\|S\|^* := \min\{\|M\| | \forall i \in [N] \forall j \in [N] M_{ij} S_{ij} \geq 1\}$$

which improves the bound using the spectral gap result.

Spectral gaps tighten the bounds

Let S be a Δ regular sign matrix with $\Delta \leq \frac{N}{2}$, and B is its Boolean version. Then $\|S\|^2 \leq \dfrac{N \cdot \sigma_2(B)}{\Delta}$.

- Set $M = \frac{N}{\Delta}B - I$. Then

$$N \geq 2\Delta \Rightarrow \bigwedge_i \bigwedge_j M_{ij}S_{ij} \geq 1 \therefore \|S\|^* \leq \|M\|.$$

- Since B is regular and singular at $y = 1_N$ with singular value Δ, we have $My = 0$, and decompose all $x = x_1 + x_2$ with x_1 the projection on y, and x_2 the orthogonal complement.
-

$$\langle Mx, Mx \rangle = \langle Mx_2, Mx_2 \rangle = \frac{N^2}{\Delta^2}\langle Bx_2, Bx_2 \rangle$$

Proof (Cont'D)

- Thus $\|B\| = \Delta$, and then by regularity, there are Δ permutation matrices $B^{(i)}$ such that they sum to B, and each has spectral norm 1.
- The desired bound follows by the triangle equality.
- Finally, by orthogonality,

$$\|Bx_2\| \leq \sigma_2(B) \cdot \|x_2\| \leq \sigma_2(B) \cdot \|x\|$$

whence

$$\|M\| \leq \frac{N \cdot \sigma_2(B)}{\Delta}.$$

Forster's Lemma's Proof

The idea is to use the compactness of the space to iteratively construct $B_1, B_2, B_3, \ldots$ such that for any X in general position, each B_i makes $B_{i-1}X$ closer to being equidistant.

.

We see $D \leq N$ by Cauchy-Schwarz

$$\begin{aligned} D &= \sum_i^k \sum_X \sum_Y M_{x,y} x_i y_i \\ &\leq \sum_i^k \|M\| \sqrt{\sum_X x_i^2} \sqrt{\sum_Y y_i^2} \\ &\leq \|M\| \sqrt{\sum_i^k \sum_X x_i^2} \sqrt{\sum_i^k \sum_Y y_i^2} = N(1) \end{aligned}$$

Forster's Theorem Proof (Lower Bound)

We see $D \geq \frac{N^2}{k}$ as follows:

- First, $|M_{x,y}| \geq 1$ and $|\langle x, y \rangle| \leq 1$ for all x,y.
- Then

$$
\begin{aligned}
D &= \sum_X \sum_Y M_{x,y} \langle x, y \rangle \\
&\geq \sum_X \sum_Y (\langle x, y \rangle)^2 \\
&\geq \sum_Y \sum_X \langle y, (x \otimes x) y \rangle \\
&= \frac{N}{k} \sum_Y \langle y, y \rangle \\
&= \frac{N^2}{2}
\end{aligned}
$$

Alon-Boppa Consequences

- The Alon-Boppa theorem optimally describes limitations on spectral gaps.
- Given the second eigenvalue σ of a Δ regular graph,

$$\sigma \geq 2\sqrt{\Delta - 1} - o(1).$$

- Consequently, the best lower bound on sign rank yielded by spectral gap methods will roughly be $\frac{\sqrt{\Delta}}{2}$ when \$$\Delta \leq N^{o(1)}$.

Questions?