Gaussian Mean Width and Effective Dimension

Andrew Nobel

April, 2023

Multivariate Normal and Uniform Distribution on the Sphere

Fact: Let $Z \sim \mathcal{N}_n(0, I)$, and let $S^{n-1} = \{x \in \mathbb{R}^n : ||x|| = 1\}$ be the unit n-sphere

1. $Z \,/\, ||Z|| \;\sim$ Unif(S^{n-1})
2. $Z \,/\, ||Z||$ is independent of $||Z||$
3. $\mathbb{E}||Z|| \leq \sqrt{n}$
4. $\mathbb{E}||Z|| \,/\sqrt{n} \to 1$ as n tends to infinity
5. $\mathbb{P}(|\,||Z|| - \mathbb{E}||Z||\,| > t) \leq 2e^{-t^2/2}$

Rule of thumb: If $Z \sim \mathcal{N}_n(0, I)$ then $Z \approx \sqrt{n}$ Unif(S^{n-1})

Gaussian Mean Width

Width of a Set

Note: Each unit vector $\eta \in \mathbb{R}^n$ determines a direction, and a family of parallel hyperplanes that are each perpendicular to η

$$H = H(\eta, a) = \{x : \langle x, \eta \rangle = a\} \quad a \in \mathbb{R}$$

Definition: Let $K \subseteq \mathbb{R}^n$ be a bounded set. The width of K in direction η, denoted $w(K;\eta)$, is the minimum w for which there exist $a \leq b$ such that

1. K lies between $H(\eta, a)$ and $H(\eta, b)$
2. the difference $b - a \leq w$

Note: Best choice of constants a, b is

$$a = \inf_{u \in K} \langle \eta, u \rangle \quad \text{and} \quad b = \sup_{v \in K} \langle \eta, v \rangle$$

Gaussian Mean Width of a Set

Fact: Let $K - K = \{u - v : u, v \in K\}$. Width of set K in direction η is given by

$$w(K : \eta) = \sup_{x \in K-K} \langle \eta, x \rangle$$

Key idea: Study the size/dimensionality of K through its average width $\mathbb{E}w(K;\eta)$ when η is a randomly chosen direction.

Definition: The *Gaussian mean width* (GMW) of a bounded set $K \subseteq \mathbb{R}^n$ is given by

$$w(K) = \mathbb{E}w(K : V) = \mathbb{E}\left[\sup_{x \in K-K} \langle x, Z \rangle\right] \quad \text{where } Z \sim \mathcal{N}_n(0, I)$$

Aside: $w(K) \approx \sqrt{n}\,\mathbb{E}w(K : U)$ where $U \sim$ Unif(S^{n-1})

Properties of Gaussian Mean Width

Fact: Let $K \subseteq \mathbb{R}^n$ be bounded. Recall diam$(K) = \sup\{||u - v|| : u, v \in K\}$

1. $w(K) \geq 0$
2. $w(K) = 2\,\mathbb{E} \sup_{x \in K} \langle x, Z \rangle$
3. If $K \subseteq K'$ then $w(K) \leq w(K')$
4. For each $u \in \mathbb{R}^n$, $w(K) = w(K + u)$
5. If $A \in \mathbb{R}^{n \times n}$ is orthogonal, then $w(K) = w(AK)$, where $AK = \{Ax : x \in K\}$
6. $w(K) = \mathbb{E}\left[\sup_{x \in K - K} |\langle x, Z \rangle|\right]$
7. $w(K) = w(\text{cvx}(K))$
8. $\sqrt{2/\pi}\,\text{diam}(K) \leq w(K) \leq \sqrt{n}\,\text{diam}(K)$

Concentration of Gaussian Mean Width

Note: By Gaussian concentration, for each $t > 0$

$$\mathbb{P}\left(\left| w(K) - \sup_{x \in K-K} \langle x, Z \rangle \right| > t \right) \leq 2 \exp\left\{ \frac{-t^2}{2\,\mathsf{diam}(K)^2} \right\}$$

Cor: For each $s > 0$, with probability at least $1 - e^{-s^2/2}$,

$$\left| w(K) - \sup_{x \in K-K} \langle x, V \rangle \right| \leq s\,\mathsf{diam}(K)$$

Note: Replacing K by cvx(K) leaves mean width unchanged, makes calculation of $\sup_{K-K} \langle x, Z \rangle$ a convex optimization problem. Use this to estimate $w(K)$.

Examples

1. $K = S^{n-1} = \{x \in \mathbb{R}^n : ||x|| = 1\}$ unit sphere, $w(K) = 2\,\mathbb{E}||Z|| \le 2\sqrt{n}$

2. $K = B^n = \{x \in \mathbb{R}^n : ||x|| \le 1\}$ unit ball, $w(K) = 2\,\mathbb{E}||Z|| \le 2\sqrt{n}$

3. If $K \subseteq B^n \cap E$ where E is subspace of dimension d then $w(K) \le 2\sqrt{d}$

4. If $K \subseteq B^n$ is a finite set then $w(K) \le \sqrt{8 \log |K|}$

Interpretation of Gaussian Mean Width

Insight: Quantity $w(K)^2$ acts like an effective dimension of $K \subseteq B^n$

- effective dimension of K is bounded by ambient dimension n
- small change in K has small effect on effective dimension
- effective dimension of K equals effective dimension of cvx(K)

Generalized ℓ_p norms on $\mathbb{R}^n$

Definition: Let $x \in \mathbb{R}^n$

1. For $1 \leq p < \infty$ let $||x||_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$
2. For $p = \infty$ let $||x||_\infty = \max_{1 \leq i \leq n} |x_i|$
3. For $0 < p < 1$ let $||x||_p = \sum_{i=1}^n |x_i|^p$
4. For $p = 0$ let $||x||_0 = \sum_{i=1}^n \mathbb{I}(x_i \neq 0)$, number of non-zero coordinates of x

Note

- For all p we have $||x||_p = 0$ iff $x = 0$
- For $0 < p \leq \infty$, triangle inequality $||x + y||_p \leq ||x||_p + ||y||_p$ holds
- For $1 \leq p \leq \infty$ we have $||ax||_p = |a| \, ||x||_p$
- $|| \cdot ||_p$ is concave for $0 < p < 1$ and convex for $1 \leq p \leq \infty$

Example: Sparse Signals

Sparsity: Fix $s \leq n$ (usually $s << n$) and define the set

$$K = \{x \in \mathbb{R}^n : ||x||_2 \leq 1 \text{ and } ||x||_0 \leq s\} \subseteq B^n$$

of n vectors with length 1 and at most s non-zero components

Proposition: With K as defined above there are constants $c_1 \leq c_2$ not depending on n or s such that

$$c_1 \sqrt{s \log \frac{n}{s}} \leq w(K) \leq c_2 \sqrt{s \log \frac{n}{s}}$$

Note: Effective dimension $w(K)^2 \approx s \log \frac{n}{s}$ is approximately linear in sparsity s, has only logarithmic dependence on ambient dimension n