Almost sure convergence of sample version of Half Space depth

Rohan Shinde

M. Stat 2^{nd} Year Project

April 20, 2024

Half Space Depth

Half Space Depth (Sample Version)

For a given data cloud $\Omega_n = \{\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n\}$, the Tukey's half space depth at a point $\mathbf{x} \in \mathbb{R}^d$ with respect to given data cloud Ω_n is given by

$$\begin{aligned} \mathrm{HD}(\mathbf{x}, \Omega_n) &= \inf_{\substack{\alpha \\ ||\alpha||=1}} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\left\{\alpha'(\mathbf{X}_i - \mathbf{x}) \geq 0\right\} \\ &= 1 - \sup_{\substack{\alpha \\ ||\alpha||=1}} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\left\{\alpha'(\mathbf{X}_i - \mathbf{x}) \leq 0\right\} \end{aligned}$$

Half Space Depth

Half Space Depth (Population Version)

For a given $\mathbf{x} \sim F$, the Tukey's half space depth at a point $\mathbf{x} \in \mathbb{R}^d$ with respect to given distribution F is given by

$$\begin{aligned} \mathrm{HD}(\mathbf{x}, F) &= \inf_{\substack{\alpha \\ ||\alpha||=1}} \mathbb{P}\left(\alpha'(\mathbf{X}_i - \mathbf{x}) \geq 0\right) \\ &= 1 - \sup_{\substack{\alpha \\ ||\alpha||=1}} \mathbb{P}\left(\alpha'(\mathbf{X}_i - \mathbf{x}) \leq 0\right) \end{aligned}$$

Goal

Goal

Suppose $\mathbf{X}_1, \cdots, \mathbf{X}_n \overset{\text{iid}}{\sim} F$. Let $\text{HD}(\mathbf{x}, F_n)$ be the half space depth of $\mathbf{x}$ with respect to $\Omega_n = \{\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n\}$. Then as $n \to \infty$

$$\text{HD}(\mathbf{x}, F_n) \overset{\text{a.s}}{\to} \text{HD}(\mathbf{x}, F)$$

Goal

Goal

Suppose $\mathbf{X}_1, \cdots, \mathbf{X}_n \overset{\text{iid}}{\sim} F$. Let $\text{HD}(\mathbf{x}, F_n)$ be the half space depth of $\mathbf{x}$ with respect to $\Omega_n = \{\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n\}$. Then as $n \to \infty$

$$\text{HD}(\mathbf{x}, F_n) \overset{\text{a.s}}{\to} \text{HD}(\mathbf{x}, F)$$

$$\Big\Updownarrow$$

$$\Big|\text{HD}(\mathbf{x}, F_n) - \text{HD}(\mathbf{x}, F)\Big| \overset{\text{a.s}}{\to} 0 \quad \text{as } n \to \infty$$

Towards the goal

Towards the goal

Proposition 1

Suppose $\mathbf{X}_1, \cdots, \mathbf{X}_n \stackrel{\text{iid}}{\sim} F$ and $\mathbf{X} \sim F$ independently of $\mathbf{X}_1, \cdots, \mathbf{X}_n$. Then

$$\mathbb{P}\left(\sup_{\substack{\alpha \\ ||\alpha||=1}} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\left\{\alpha'(\mathbf{X}_i - \mathbf{x}) \geq 0\right\} - \mathbb{P}(\alpha'(\mathbf{X} - \mathbf{x}) > 0) \right| > \epsilon \right)$$

$$< K n^d e^{-2n\epsilon^2}$$

where K is a constant independent of n

Towards the goal

Proposition 1

Suppose $\mathbf{X}_1, \cdots, \mathbf{X}_n \overset{\text{iid}}{\sim} F$ and $\mathbf{X} \sim F$ independently of $\mathbf{X}_1, \cdots, \mathbf{X}_n$. Then

$$\mathbb{P}\left(\sup_{\substack{\alpha \\ ||\alpha||=1}} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\left\{ \alpha'(\mathbf{X}_i - \mathbf{x}) \geq 0 \right\} - \mathbb{P}(\alpha'(\mathbf{X} - \mathbf{x}) > 0) \right| > \epsilon \right) < K n^d e^{-2n\epsilon^2}$$

where K is a constant independent of n

Borel Cantelli Lemma

For a sequence of events $\{E_n\}_{n \geq 1}$ in some probability space,

$$\text{if} \quad \sum_{i=1}^{n} \mathbb{P}(E_n) < \infty \quad \text{then} \quad \mathbb{P}\left(\limsup_{n \to \infty} E_n \right) = 0$$

Proof of Goal

Proof of Goal

Note that

$$\left|\text{HD}(\mathbf{x}, F_n) - \text{HD}(\mathbf{x}, F)\right| = \Bigg|1 - \sup_{\substack{\alpha \\ ||\alpha||=1}} \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left\{\alpha'(\mathbf{X}_i - \mathbf{x}) \le 0\right\}$$

$$- 1 + \sup_{\substack{\alpha \\ ||\alpha||=1}} \mathbb{P}\left(\alpha'(\mathbf{X}_i - \mathbf{x}) \le 0\right)\Bigg|$$

$$= \Bigg|\sup_{\substack{\alpha \\ ||\alpha||=1}} \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left\{\alpha'(\mathbf{X}_i - \mathbf{x}) \le 0\right\} - \sup_{\substack{\alpha \\ ||\alpha||=1}} \mathbb{P}\left(\alpha'(\mathbf{X}_i - \mathbf{x}) \le 0\right)\Bigg|$$

But note that for any $\epsilon > 0$,

$$\left|\sup_{\substack{\alpha \\ ||\alpha||=1}} \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left\{\alpha'(\mathbf{X}_i - \mathbf{x}) \ge 0\right\} - \sup_{\substack{\alpha \\ ||\alpha||=1}} \mathbb{P}(\alpha'(\mathbf{X} - \mathbf{x}) > 0)\right| > \epsilon$$

$$\Rightarrow \sup_{\substack{\alpha \\ ||\alpha||=1}} \left|\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left\{\alpha'(\mathbf{X}_i - \mathbf{x}) \ge 0\right\} - \mathbb{P}(\alpha'(\mathbf{X} - \mathbf{x}) > 0)\right| > \epsilon$$

Proof of Goal (Contd.)

Thus using Proposition 1, along with this, we get that

$$\mathbb{P}\left(\left|\sup_{\substack{\alpha \\ ||\alpha||=1}} \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left\{\alpha'(\mathbf{X}_i - \mathbf{x}) \geq 0\right\} - \sup_{\substack{\alpha \\ ||\alpha||=1}} \mathbb{P}(\alpha'(\mathbf{X} - \mathbf{x}) > 0)\right| > \epsilon\right) < Kn^d e^{-2n\epsilon^2}$$

Now, $\dfrac{K(n+1)^d e^{-2(n+1)\epsilon^2}}{Kn^d e^{-2n\epsilon^2}} = e^{-2\epsilon^2}\left(1 + \dfrac{1}{n}\right)^d$ and hence

$$\limsup_{n\to\infty} \frac{K(n+1)^d e^{-2(n+1)\epsilon^2}}{Kn^d e^{-2n\epsilon^2}} = e^{-2\epsilon^2} < 1$$

Thus by ratio test $\displaystyle\sum_{n=1}^{\infty} Kn^d e^{-2n\epsilon^2} < \infty$

Proof of Goal (Contd.)

And hence

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\left| \sup_{\substack{\boldsymbol{\alpha} \\ ||\boldsymbol{\alpha}||=1}} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\left\{ \boldsymbol{\alpha}'(\mathbf{X}_i - \mathbf{x}) \geq 0 \right\} - \sup_{\substack{\boldsymbol{\alpha} \\ ||\boldsymbol{\alpha}||=1}} \mathbb{P}(\boldsymbol{\alpha}'(\mathbf{X} - \mathbf{x}) > 0) \right| > \epsilon \right) < \infty$$

By Borel Cantelli Lemma, we thus get that

$$\limsup_{n \to \infty} \left\{ \left| \sup_{\substack{\boldsymbol{\alpha} \\ ||\boldsymbol{\alpha}||=1}} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\left\{ \boldsymbol{\alpha}'(\mathbf{X}_i - \mathbf{x}) \geq 0 \right\} - \sup_{\substack{\boldsymbol{\alpha} \\ ||\boldsymbol{\alpha}||=1}} \mathbb{P}(\boldsymbol{\alpha}'(\mathbf{X} - \mathbf{x}) > 0) \right| > \epsilon \right\} = 0$$

Thus

$$\left| \mathrm{HD}(\mathbf{x}, F_n) - \mathrm{HD}(\mathbf{x}, F) \right| \overset{\text{a.s}}{\to} 0 \quad \text{as } n \to \infty$$

Proof of Proposition 1

Prerequisites

Shattering property

A class $\mathcal{D}$ is said to shatter a set of points F if it can pick out every possible subset (the empty subset and the whole of F included); that is, $\mathcal{D}$ shatters F if each of the subsets of F has the form $D \cap F$ for some D in $\mathcal{D}$

Shattering number

A binary function on a space $\mathcal{Z}$ is a function $f : \mathcal{Z} \to \{0, 1\}$. Let $\mathcal{F}$ be a class of binary functions on $\mathcal{Z}$. For any $z_1, \cdots, z_n$ define
$\mathcal{F}_{z_1,\cdots,z_n} = \{(f(z_1), \cdots, f(z_n)) : f \in \mathcal{F}\}$
$\mathcal{F}_{z_1,\ldots,z_n}$ is a finite collection of binary vectors and $|\mathcal{F}_{z_1,\ldots,n}| \leq 2^n$. The set $\mathcal{F}_{z_1,\ldots,n}$ is called the projection of $\mathcal{F}$ onto $z_1, \cdots, z_n$. The *growth function* or *shattering number* is defined by

$$s(\mathcal{F}, n) = \sup_{z_1,\cdots,z_n} |\mathcal{F}_{z_1,\cdots,z_n}|$$

Prerequisites

VC Dimension

VC dimension of a class of binary functions $\mathcal{F}$ is

$$\mathrm{VC}(\mathcal{F}) = \sup\{n : s(\mathcal{F}, n) = 2^n\}$$

Theorem(Vapnik and Chervonenkis)

Let $\mathcal{F}$ be a class of binary functions. For any $\epsilon > \sqrt{2/n}$,

$$\mathbb{P}\left(\sup_{f \in \mathcal{F}} |(P_n(f) - P(f)| > \epsilon\right) \leq 4s(\mathcal{F}, 2n)e^{-n\epsilon^2/8}$$

Sauer's lemma

For $n \geq \mathrm{VC}(\mathcal{F}) = d$

$$s(\mathcal{F}, n) \leq \sum_{i=0}^{\mathrm{VC}(\mathcal{F})} \binom{n}{i} \leq \left(\frac{ne}{\mathrm{VC}(\mathcal{F})}\right)^{\mathrm{VC}(\mathcal{F})} < n^{\mathrm{VC}(\mathcal{F})}$$

Proof of Proposition 1

From Hoeffding's lemma for i.i.d. random variables, for any fixed $\mathbf{x}$, $\boldsymbol{\alpha}$ and for every $\epsilon > 0$ we have

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left\{\boldsymbol{\alpha}'(\mathbf{X}_i - \mathbf{x}) \geq 0\right\} - \mathbb{P}(\boldsymbol{\alpha}'(\mathbf{X} - \mathbf{x}) > 0)\right| > \epsilon\right) < 2e^{-2n\epsilon^2}$$

Now, the set of halfspaces $\{\mathbf{X} : \boldsymbol{\alpha}'(\mathbf{X} - \mathbf{x}) > 0\}$ in $\mathbb{R}^d$ with varying $\boldsymbol{\alpha}$ has VC dimension $= d$.

Proof of proposition 1 (contd.)

Now note that

$$\mathbb{P}\left(\sup_{\substack{\boldsymbol{\alpha} \\ ||\boldsymbol{\alpha}||=1}} \left|\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left\{\boldsymbol{\alpha}'(\mathbf{X}_i - \mathbf{x}) \geq 0\right\} - \mathbb{P}(\boldsymbol{\alpha}'(\mathbf{X} - \mathbf{x}) > 0)\right| > \epsilon\right)$$
$$= \mathbb{P}\left(\sup_{f_{\boldsymbol{\alpha}} \in \mathcal{F}} |(P_n(f_{\boldsymbol{\alpha}}) - P(f_{\boldsymbol{\alpha}})| > \epsilon\right)$$

where $\mathcal{F} = \{\mathbf{X} \mapsto f_{\boldsymbol{\alpha}}(\mathbf{X}) := \mathbb{1}\{\boldsymbol{\alpha}'(\mathbf{X} - \mathbf{x}) > 0\} | \boldsymbol{\alpha} \in \mathbb{R}^d\}$
Using the theorem by Vapnik and Chervonenkis, we get that

$$\mathbb{P}\left(\sup_{f_{\boldsymbol{\alpha}} \in \mathcal{F}} |(P_n(f_{\boldsymbol{\alpha}}) - P(f_{\boldsymbol{\alpha}})| > \epsilon\right) \leq 4s(\mathcal{F}, 2n)e^{-n\epsilon^2/8}$$

Proof of Proposition 1(contd.)

But by Sauer's lemma, $s(\mathcal{F}, 2n) < (2n)^{\mathsf{VC}(\mathcal{F})} = 2^d n^d$. Thus

$$\mathbb{P}\left(\sup_{\substack{\boldsymbol{\alpha} \\ ||\boldsymbol{\alpha}||=1}} \left|\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left\{\alpha'(\mathbf{X}_i - \mathbf{x}) \geq 0\right\} - \mathbb{P}(\alpha'(\mathbf{X} - \mathbf{x}) > 0)\right| > \epsilon\right)$$
$$< Kn^d e^{-n\epsilon^2/8}$$

where K is constant independent of n.

Appendix

Halfspace VC dimension

Result

Let $\mathcal{F} = \{x \in \mathbb{R}^d : \text{sign}(\mathbf{w}^\top \mathbf{x}) = y \mid \mathbf{w} \in \mathbb{R}^d\}$ be the class of linear classifiers (or halfspaces) in $\mathbb{R}^d$. The VC dimension of $\mathcal{F}$ is d. If we allow for a bias term, then the VC dimension is $d + 1$.

To prove this, let's first describe a set S of d points that is shattered by the class. Consider the set of points $\mathbf{x}_i = \mathbf{e}_i$ for $i \in [d]$ where $\mathbf{e}_i$ is the ith standard basis vector that has 1 at coordinate i and 0 everywhere else. In order to show that these points can be shattered, for all labeling $y_1, \ldots, y_d \in \{-1, 1\}$ we need to show the existence of $f \in \mathcal{F}$ that realizes it. Consider labeling $y_1, \ldots, y_d \in \{-1, 1\}$ then choosing $\mathbf{w}$ as below suffices.

$$\mathbf{w} = \sum_{i=1}^{d} y_i \mathbf{e}_i$$

Halfspace VC dimension (Contd.)

Then we have, that for all $i \in [d]$, $\text{sign}(\mathbf{w}^\top \mathbf{x}_i) = y_i$. Thus it generates the labeling $(y_1, \ldots, y_d)$. Since we can do this for any labeling, these points can be shattered.

Now we need to show that no $d+1$ points can be shattered. In order to show this, let us consider any set of $d+1$ points $\mathbf{x}_1, \ldots, \mathbf{x}_{d+1}$. We know that no set of $d+1$ d-dimensional vectors can be linearly independent, thus there exists some $j \in [d+1]$ such that

$$\mathbf{x}_j = \sum_{i \neq j} \alpha_i \mathbf{x}_i,$$

such that at least one $\alpha_i \neq 0$.

Halfspace VC dimension (Contd.)

Suppose we consider the labeling where $y_j = -1$ and for all $i \neq j$ $y_i = \text{sign}(\alpha_i)$ if $\alpha_i \neq 0$ else $y_i = 1$. We will show that no $\mathbf{w}$ can achieve this labeling. Suppose there is a $\mathbf{w}$ that achieves this labeling then we have for all $i \neq j$ if $\alpha_i \neq 0$ then $\alpha_i(\mathbf{w}^\top \mathbf{x}_i) > 0$ since $\text{sign}(\mathbf{w}^\top \mathbf{x}_i) = y_i = \text{sign}(\alpha_i)$. This gives us,

$$\mathbf{w}^\top \mathbf{x}_j = \sum_{i \neq j} \alpha_i \mathbf{w}^\top \mathbf{x}_i > 0.$$

Thus $\mathbf{w}$ would label $\mathbf{x}_j$ incorrectly as positive when $y_j = -1$. This proves that no $d + 1$ points can be shattered.