Concentration inequalities of U-statistics for sampling without replacement

AiJianhang

2020.01.24

Contents

1. Concentration inequalities
2. Hoeffding’s inequality
3. An example in Machine learning

Concentration inequalities

- Chebyshev's inequality

If X is a random variable with variance:

$$Var(X) = E[(X - E(X))^2]$$

Then we have

$$\Pr\{|X - EX| > t\} \leq \frac{Var(X)}{t^2}$$

Concentration inequalities

- Law of large numbers

If $X_1, X_2, \ldots, X_n$ are i.i.d variables, and

$$\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n} \qquad E[\bar{X}_n] = \mu$$

Then we have

$$\Pr\{\lim_{n \to \infty} \bar{X}_n = \mu\} = 1$$

Concentration inequalities

- Generalization

If $X_1, X_2, \ldots, X_n$ are *independent* variables taking values in some set $\mathcal{X}$. Let $Z : \mathcal{X}^n \rightarrow \mathbb{R}$

$$Z = f(X_1, \ldots . X_n)$$

How large are typical deviation of Z from $E[Z]$?

In particular, we seek upper bound for, $t > 0$,

$$\Pr\{Z > E[Z] - t\} \qquad \Pr\{Z < E[Z] - t\}$$

Hoeffding's inequality

- Classcial version

If $\mathcal{X} = \{x_1, \ldots, x_N\}, X_1, X_2, \ldots, X_n$ are random variables sampled from it with replace, independent to each other. Then we have

$$\Pr\{\bar{X} - E[\bar{X}] \geq t\} \leq \exp\{-\frac{2nt^2}{(b-a)^2}\}$$

where $\bar{X} = \frac{X_1 + \cdots + X_n}{n}$

$$a = \min_{1 \leq i \leq N} x_i \qquad b = \max_{1 \leq i \leq N} x_i$$

Hoeffding's inequality

- U-statistics

Consider the function of $X_1, X_2, \ldots, X_n$,

$$U = \frac{1}{n^{(r)}} \sum_{n,r} g(X_{i_1}, \cdots, X_{i_r})$$

where $n^{(r)} = n(n-1)\cdots(n-r+1)$ and the sum $\sum_{n,r}$ is taken over all r-tuples $i_1, \cdots, i_r$ of distinct positive integers not exceeding n.

- U-statistics version

$$\Pr\{U - EU \geq t\} \leq \exp\{-\frac{-2kt^2}{(b-a)^2}\}$$

where $k = [n/r]$.

Hoeffding's inequality

U-statisitc order	Sampling with replacement	Sampling without replacement
r=1	√[1]	√
r>1	√[1]	?

Hoeffding's inequality

- Original version[1]

  Sample without replacement, all other conditions are the same

$$\Pr\{\bar{X} - E[\bar{X}] \geq t\} \leq \exp\{-\frac{2nt^2}{(b-a)^2}\}$$

- Improved version[2]

$$\Pr\{\bar{X} - E[\bar{X}] \geq t\} \leq \exp\{-\frac{2nt^2}{(1-f_n^*)(b-a)^2}\}$$

where $f_n^* = \frac{n-1}{N}$

Hoeffding's inequality

- Hoeffding-Serffling version[3]

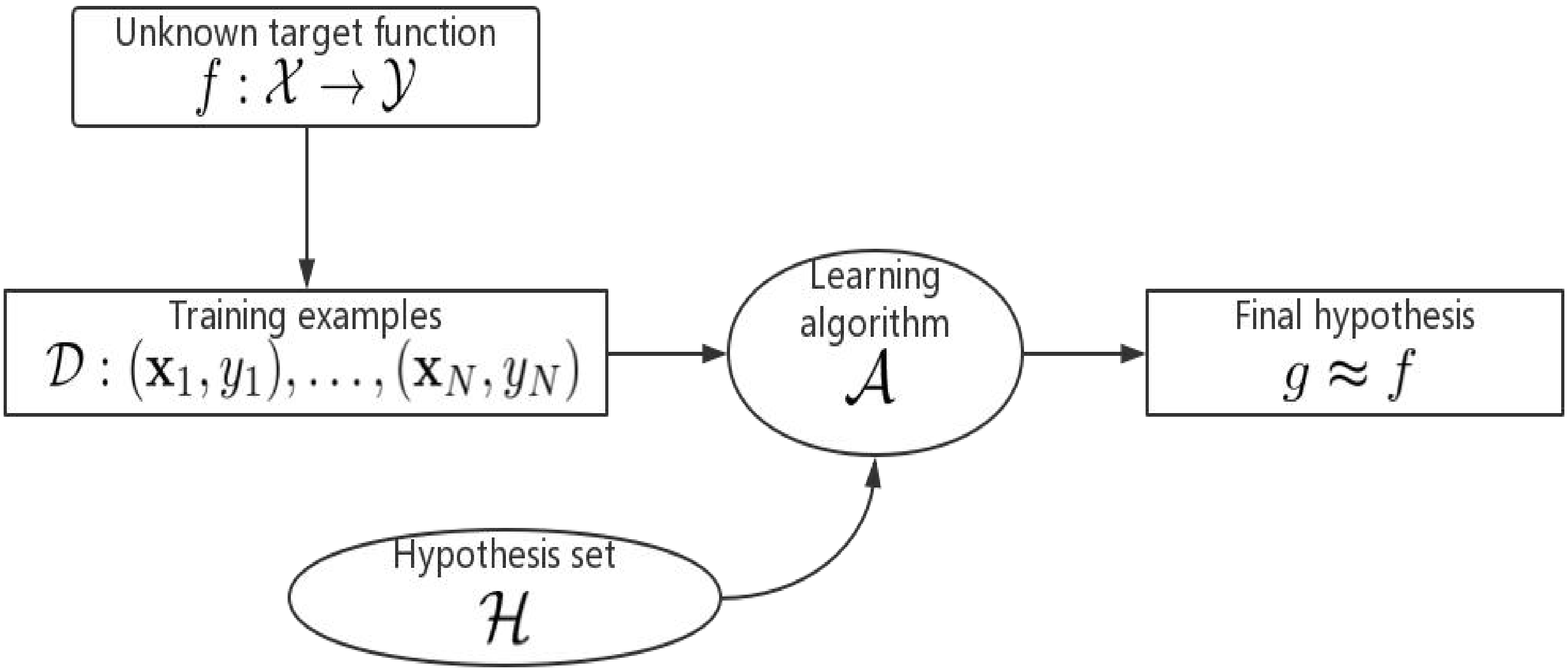
$$\Pr\Big(\max_{n\leq k\leq N-1} \frac{\sum_{j=1}^{k}(X_j-\mu)}{k} \geq t\Big) \leq \exp\Big\{-\frac{2nt^2}{(b-a)^2}\frac{1}{(1-\frac{n}{N})(1+\frac{1}{n})}\Big\}$$

<u>Question:</u>

Will this three version of bounds still hold when the U-statistics order bigger than 1?

<u>Still open</u>

An example in Machine learning



For some $h \in \mathcal{H}$

- $g \approx f$ inside $\mathcal{D}$: Yes
- $g \approx f$ outside $\mathcal{D}$: We hope so, how to guarantee it?

An example in Machine learning

Question

The probability of blue balls μ ?

Answer

Sample, approximate it with the fraction of blue balls ν

Sample size n, random variables X_i are independent,

$X_i = 1$,when it is blue;

$X_i = 0$,when it is white; $\quad i = 1, \ldots, n$

Then with Hoeffding's inequality, we have

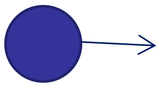
$$\Pr[|\nu - \mu| > t] \leq 2\exp(-2nt^2)$$

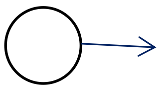
We say $\nu = \mu$ is *probably approximately correct(PAC)*

An example in Machine learning

Translate the “Marine balls” example to Machine learning

For some *fixed* $h \in \mathcal{H}$

 $h(\mathbf{x}) \neq f(\mathbf{x})$

 $h(\mathbf{x}) = f(\mathbf{x})$

Sample $X_i \longrightarrow \mathcal{D} : (\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)$

$\mu \longrightarrow$ Probability $h(\mathbf{x}) \neq f(\mathbf{x})$($E_{out}(h)$ errors outside sample)

$\nu \longrightarrow$ Fraction $h(\mathbf{x}_n) = y_n$ ($E_{in}(h)$ errors inside sample)

$$\Pr[|E_{out}(h) - E_{in}(h)| > t] \leq 2\exp(-2nt^2)$$

An example in Machine learning

Multiple h, $|\mathcal{H}| = M$

Bad sample: for some h, E_{in} and E_{out} is far away

$$\Pr[BAD\ \mathcal{D}]$$

$$= \Pr[\mathcal{D}\ is\ bad\ for\ h_1\ or \ldots\ or\ \mathcal{D}\ is\ bad\ for\ h_M]$$

$$\leq \Pr[\mathcal{D}\ is\ bad\ for\ h_1] + \cdots + \Pr[\mathcal{D}\ is\ bad\ for\ h_M]$$

$$\leq 2M \exp\left(-2t^2 n\right)$$

Now, when $|\mathcal{H}| = M$ is finite, n sample size is big enough,

for whatever g picked by $\mathcal{A}$, $E_{out}(g) \approx E_{in}(g)$.

If $\mathcal{A}$ find one g with $E_{in}(g) \approx 0$, *PAC* guarantee for $E_{out}(g) \approx 0$ ⟹ *Question answered*

An example in Machine learning

Question:

What if we lost the independent condition? What if we build more complicated models?

Other new version of concentration inequalites

Thank you