Homework 1
李娜
15320170155232
2019/3/16

Hoeffding inequality

Introduction

Inequalities are useful for bounding quantities that might be hard to compute. Among these, Hoeffding inequality is a powerful technique and perhaps the most important inequality in learn theory.[1] In this report, what I mainly want to do is to repeat the proof of this inequality. Finally, I will talk about some applications of the inequality.

What's Hoeffding inequality

Hoeffding inequality was proposed and proved by Hoeffding(1963). The general format is as follows:
Let $X_1, \ldots, X_n$ be independent bounded random variables such that X_i falls in the interval $[a_i, b_i]$ with probability one. Denote their sum by $S_n = \sum_{i=1}^{n} X_i$. Then for any ε>0,we have

$$\mathbf{P}\{S_n - ES_n \geq \varepsilon\} \leq exp(\frac{-2\varepsilon^2}{\sum_{i=1}^{n}(b_i - a_i)^2})$$

and $$\mathbf{P}\{S_n - ES_n \leq -\varepsilon\} \leq exp(\frac{-2\varepsilon^2}{\sum_{i=1}^{n}(b_i - a_i)^2}).$$

Proof

To prove this inequality, we need to use Hoeffding Lemma, which is let X be a random variable with EX=0, $a \leq X \leq b$. Then for s >0, $E\{e^{sX}\} \leq e^{s^2(b-a)^2/8}$. So firstly, we will prove this Lemma.

- step 1

$f(t) = \exp(t)$ is a convex function. According to **Jensen's inequality**, if $f: R \to R$ is a convex function, then $f(E[x]) \leq E[f(x)]$.
so for $a \leq x \leq b$, we have,

$$e^{sx} \leq \frac{x-a}{b-a}e^{sb} + \frac{b-x}{b-a}e^{sa}$$

Applying EX=0,we get $Ee^{sx} \leq \frac{b}{b-a}e^{sa} - \frac{a}{b-a}e^{sb}$, define $v = -a/(b-a)$,we can easily know $1 - v = b/(b-a)$, $a = -v(b-a)$ and $b = (1-v)(b-a)$.

[1] John Duchi, CSS229 Supplement Lecture notes Hoeffding's inequality.

So, $\mathrm{E}e^{sx} \leq (1-v)e^{-sv(b-a)} + ve^{(1-v)s(b-a)}$

$$= (1-v+ve^{s(b-a)})e^{-sv(b-a)}$$

$$= e^{\log(1-v+ve^{s(b-a)})-sv(b-a)}$$

Define $u = \mathrm{s(b-a)}$ and $\Phi(u) = \log(1-v+ve^{u}) - uv$, we can get

$$\Phi(0) = 0 \; and \; \mathrm{E}e^{sx} \leq e^{\Phi(u)}.$$

Then take the derivative of Φ, the result is:

$$\Phi'(u) = -v + \frac{ve^{u}}{1-v+ve^{u}} \;, \quad \text{therefore} \quad \Phi'(0) = 0$$

And moreover, $\Phi''(u) = \frac{(1-v)e^{u}}{(1-v+ve^{u})^2} \leq \frac{1}{4}$ for $u > 0$

Thus, by Taylor series expansion, for some $\theta \in [0, u]$,

$$\Phi(u) = \Phi(0) + u\Phi'(0) + \frac{u^2}{2}\Phi''(u) \leq \frac{u^2}{8} = \frac{s^2(b-a)^2}{8}$$

So, $\mathrm{E}\{e^{sX}\} \leq e^{s^2(b-a)^2/8}$

Next, we are going to prove Hoeffding inequality.

- Step 2

Suppose s is an arbitrary positive number, for any random variable x, by **Markov's inequality**, we can get, $\mathrm{P(X > \varepsilon) = P}(e^{sX} > e^{s\varepsilon}) \leq \frac{Ee^{sX}}{e^{s\varepsilon}}$. This is the basic of **Chernoff bounds**.

Using Chernoff Bounds, we can find a $s > 0$ that minimizes the upper bound or makes upper bound small. So we have

$$\mathrm{P}\{S_n - ES_n \geq \varepsilon\} = P\{e^{s(S_n - ES_n)} \geq e^{s\varepsilon}\}$$

$$\leq e^{-s\varepsilon}E\{e^{s\sum_{i=1}^{n}(X_i - EX_i)}\}$$

$$= e^{-s\varepsilon}\prod_{i=1}^{n}E\{e^{s(X_i - EX_i)}\} \qquad \text{(because } X_i \text{ is independent)}$$

$$= e^{-s\varepsilon}\prod_{i=1}^{n}e^{s^2(b_i-a_i)^2/8} \qquad \text{(by Hoeffding Lemma)}$$

$$= e^{-s\varepsilon}e^{s^2\sum_{i=1}^{n}(b_i-a_i)^2/8}$$

$$= e^{\frac{-2\varepsilon^2}{\sum_{i=1}^{n}(b_i-a_i)^2}} \qquad \left(\mathrm{s} = \frac{4\varepsilon}{\sum_{i=1}^{n}(b_i-a_i)^2}\right)$$

The final step comes from

min $-\mathrm{s}\varepsilon + s^2\sum_{i=1}^{n}(b_i - a_i)^2/8$, take the derivative w.r.t s, we can get

$$-\varepsilon + \mathrm{s}\sum_{i=1}^{n}\frac{(b_i-a_i)^2}{4} = 0 \;\rightarrow\; \mathrm{s} = 4\varepsilon/\sum_{i=1}^{n}(b_i-a_i)^2$$

The second inequality is proved analogously.

Combining the two inequalities, we can get,

$$\mathrm{P}\{|S_n - ES_n| \geq \varepsilon\} \leq 2\exp\left(\frac{-2\varepsilon^2}{\sum_{i=1}^{n}(b_i-a_i)^2}\right)$$

Application

Hoeffding inequality is very important for error estimation methods. It can be applied to the learning problem, allowing us to make a prediction outside of Data set. Based on the inequality, a sample data set can be used to assess whether or not hypothesis(h) is close to the target function(f).

Reference

[1] John Duchi, CSS229 Supplement Lecture notes Hoeffding's inequality.

[2] Larry Wasserman, All of Statistics: A Concise Course in Statistical Inference, chapter 4.

[3] Luc Devroye, László Györfi and Gábor Lugosi, A Probabilistic Theory of pattern Recognition, chaper8.