Ch 6 集中不等式 (Concentration)

回顾前一次课

机器学习根本问题: $P\left[\left|\frac{1}{n}\sum_{i=1}^{m} X_i - E(X_i)\right| > \epsilon\right]$<非常小?

Markov不等式: $P(X \geq \epsilon) \leq \frac{E(X)}{\epsilon}$

Chebyshev不等式: $P(|X-\mu| > \epsilon) \leq \frac{Var(X)}{\epsilon^2}$

单边Chebyshev不等式: $P(X-\mu \geq \epsilon) \leq \frac{\sigma^2}{\sigma^2+\epsilon^2}$

Hölder不等式: $E(|XY|) \leq \left(E(|X|^p)\right)^{\frac{1}{p}}\left(E(|Y|^q)\right)^{\frac{1}{q}}$

随机变量的矩生成函数(Moment Generating Function)

定义： 定义随机变量X的矩生成函数为

$$M_X(t) = E[e^{tX}].$$

定理： 随机变量X矩生成函数为$M_X(t)$, 对任意$n \geq 1$有

$$E[X^n] = M_X^{(n)}(0)$$

这里$M_X^{(n)}(0)$表示矩生成函数在$t = 0$的n阶导数，而$E[X^n]$被称为随机变量X的n阶矩 (moment).

矩生成函数性质

定理： 对随机变量X和Y，如果存在常数$\delta > 0$，使得当$t \in (-\delta, \delta)$时有$M_X(t) = M_Y(t)$成立，那么X与Y有相同的分布

定理： 若随机变量X与Y独立, 则有

$$M_{X+Y}(t) = M_X(t)M_Y(t)$$

Chernoff方法

给定任意随机变量X和任意$t > 0$和$\epsilon > 0$，利用Markov不等式有

$$P[X \geq \epsilon] = P[e^{tX} \geq e^{t\epsilon}] \leq e^{-t\epsilon}E[e^{tX}]$$

特别地, 有

$$P[X \geq \epsilon] \leq \min_{t>0}\{e^{-t\epsilon}E[e^{tX}]\}.$$

Chernoff方法

对任意$\epsilon > 0$和$t < 0$有

$$P[X \leq -\epsilon] = P[tX \geq -t\epsilon] \leq e^{t\epsilon}E[e^{tX}]$$

同理有

$$P[X \leq -\epsilon] \leq \min_{t<0}\{e^{t\epsilon}E[e^{tX}]\}.$$

上述方法称为Chernoff方法, 是证明集中不等式最重要的方法之一

二值Chernoff界

设随机变量$X_1, X_2, \cdots, X_n$相互独立且满足$X_i \sim \mathrm{Ber}(p_i)$，令$\mu = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p_i$. 对任意$\epsilon > 0$有

$$P\left[\sum_{i=1}^{n} X_i \geq (1+\epsilon)\mu\right] \leq \left(\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}\right)^{\mu}$$

对任意$0 < \epsilon < 1$有

$$P\left[\sum_{i=1}^{n} X_i \geq (1+\epsilon)\mu\right] \leq e^{-\mu\epsilon^2/3}$$

定理

设随机变量$X_1, X_2, \cdots, X_n$相互独立且满足$X_i \sim \mathrm{Ber}(p_i)$,令$\mu = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p_i$. 对任意$0 < \epsilon < 1$有

$$P\left[\sum_{i=1}^{n} X_i \le (1-\epsilon)\mu\right] \le \left(\frac{e^{-\epsilon}}{(1-\epsilon)^{1-\epsilon}}\right)^{\mu} \le e^{-\frac{\mu\epsilon^2}{2}}$$

Rademacher随机变量

若随机变量$X \in \{+1, -1\}$满足

$$P(X = +1) = P(X = -1) = 1/2$$

则称X为Rademacher随机变量

定理： 对n个独立的Rademacher随机变量$X_1, X_2, \cdots, X_n$, 有

$$P\left[\frac{1}{n}\sum_{i=1}^{n} X_i \geq \epsilon\right] \leq e^{-n\epsilon^2/2} \quad P\left[\frac{1}{n}\sum_{i=1}^{n} X_i \leq -\epsilon\right] \leq e^{-n\epsilon^2/2}$$

推论

对独立同分布的随机变量$X_1, X_2, \cdots, X_n$满足$P(X_1 = 0) = P(X_1 = 1) = 1/2$, 有

$$P\left[\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{2} \geq \epsilon\right] \leq e^{-2n\epsilon^2}$$

$$P\left[\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{2} \leq -\epsilon\right] \leq e^{-2n\epsilon^2}$$

有界的Chernoff不等式

研究有界的随机变量$X_i \in [a,b]$的Chernoff不等式

Chernoff 引理：设随机变量$X \in [0,1]$的期望$\mu = E[X]$. 对任意$t > 0$有

$$E[e^{tX}] \le \mathrm{e}^{t\mu + \frac{t^2}{8}}$$

推论: 随机变量$X \in [a,b]$的期望$\mu = E[X]$, 对任意$t > 0$有

$$E[e^{tX}] \le \mathrm{e}^{t\mu + \frac{t^2(b-a)^2}{8}}$$

Chernoff不等式

假设$X_1, \dots, X_n$是n独立的随机变量、且满足$X_i \in [a, b]$.
对任意$\epsilon > 0$有

$$P\left[\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E[X_i] \geq \epsilon\right] \leq e^{-2n\epsilon^2/(b-a)^2}$$

$$P\left[\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E[X_i] \leq -\epsilon\right] \leq e^{-2n\epsilon^2/(b-a)^2}$$