机器学习与数据科学

Machine Learning and Data Science

主讲: 李春光

www.pris.net.cn/teacher/lichunguang

模式识别与智能系统实验室

信息与通信工程学院 网络搜索教研中心

北京邮电大学

专题 六：支持向量机与统计学习理论
数学基础知识补充-II

- **内容提要**
  - 集中不等式(Concentration inequality)
    - Markov
    - Chebyshev
    - Hoeffding
  - 应用举例

大数定律与集中不等式

说说“靠谱儿”

- 大数定律:
  - 独立随机变量的算术平均值以很大概率趋近于其数学期望

$$\lim_{N\to\infty} \mathbf{P}\left(\left|\overline{\mu}_N - \mu\right| \geq \varepsilon\right) = 0,\ \forall \varepsilon > 0$$

- 集中不等式 (Concentration inequality)
  - 提供随机变量偏离某值(比如期望)的概率界
    - Markov / Chebyshev / Chernoff / Hoeffding / Bennett / Bernstein / McDiarmid 不等式

$$\mathbf{P}\left(\left|\overline{\mu}_N - \mu\right| < \varepsilon\right) > 1 - \delta$$

  - “以概率1-δ保证偏差小于ε”

常用的集中不等式

- Markov不等式
  - Gauss-Markov不等式

– **X 是个非负随机变量， a>0, 有：** $\mathbf{P}(X \geq a) \leq \frac{\mathbf{E}(X)}{a}$

其中E(X) 是X的数学期望

$$\mathbf{P}(X \geq a) = \mathbf{P}(f(X) \geq f(a)) \leq \frac{\mathbf{E}(f(X))}{f(a)}$$

其中f(.)是一个非负的严格递增单调函数

- Chebyshev不等式

– **X 是一个随机变量， a>0, 有：**

$$\mathbf{P}(|X - \mathbf{E}(X)| \geq a) \leq \frac{\mathbf{Var}(X)}{a^2}$$

常用的集中不等式

- Hoeffding不等式
  - 如果$X_1,\ldots, X_N$是独立随机变量，假设X_i有界，即$X_i \in [a_i, b_i]$, 令 $S_N = (X_1 + X_2 + \ldots + X_N)/N$, 则

$$\mathbf{P}\left(\left|S_N - \mathbf{E}\left(S_N\right)\right| \geq t\right) \leq 2\exp\left(-\frac{2t^2N^2}{\sum_{i=1}^{N}\left(b_i - a_i\right)^2}\right)$$

  - 例: 假设w固定，损失函数L取值范围在[0,1]，则有

$$\mathbf{P}\left\{\left|R(w) - R_{emp}(w)\right| \geq \varepsilon\right\} \leq 2\exp\left(-2\varepsilon^2 N\right)$$

Hoeffding不等式

- 如果$X_1,\ldots, X_N$是独立随机变量，假设X_i有界，即$X_i \in [a_i, b_i]$, 令 $S_N = (X_1 + X_2 + \ldots + X_N)/N$, 则

$$\mathbf{P}\left(\left|S_N - \mathbf{E}(S_N)\right| \geq t\right) \leq 2\exp\left(-\frac{2t^2N^2}{\sum_{i=1}^{N}(b_i - a_i)^2}\right)$$

  – 例：假设**w**固定，损失函数**L**取值范围在**[0,1]**，则有

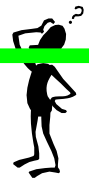
$$\mathbf{P}\left\{\left|R_{emp}(\mathbf{w}) - R(\mathbf{w})\right| \geq \varepsilon\right\} \leq 2\exp\left(-2\varepsilon^2 N\right)$$

专题 六：支持向量机与统计学习理论
数学基础知识补充-II

- **内容提要**
  - 集中不等式**(Concentration inequality)**
    - Markov
    - Chebyshev
    - Hoeffding
  - 应用举例

集中不等式的应用：例1

- 例1: 尝试给出大数定律的收敛速度



  - Markov不等式

$$\mathbf{P}\left(\left|\bar{\mu}_N-\mu\right|<\varepsilon\right)>1-\frac{\mathbf{E}\left(\left|\bar{\mu}_N-\mu\right|\right)}{\varepsilon} \Rightarrow 1-\delta$$

No convergence rate ...

  - Chebyshev不等式

$$\mathbf{P}\left(\left|\bar{\mu}_N-\mu\right|<\varepsilon\right)>1-\frac{\sigma^2}{N\varepsilon^2} \Rightarrow \delta=\frac{\sigma^2}{N\varepsilon^2} \Rightarrow \varepsilon=\frac{\sigma}{\sqrt{N\delta}}$$

Not tight enough...

  - Hoeffding不等式

$$\mathbf{P}\left(\left|\bar{\mu}_N-\mu\right|\geq\varepsilon\right)\leq 2\exp\left(-\frac{2\varepsilon^2 N}{\frac{1}{N}\sum_{i=1}^{N}\left(b_i-a_i\right)^2}\right)=2\exp\left(-\frac{2\varepsilon^2 N}{c^2}\right)$$

其中 $c^2=\frac{1}{N}\sum_{i=1}^{N}\left(b_i-a_i\right)^2$

$$\Rightarrow \delta=2\exp\left(-\frac{2\varepsilon^2 N}{c^2}\right) \Rightarrow \varepsilon=c\sqrt{\frac{1}{2N}\log\frac{2}{\delta}}$$

Tighter bound!

集中不等式的应用：例2

- 例2 (计算可靠估计所需要的样本数量N)
  - 要求以概率1- δ (比如0.9999) 使得所估计事件A发生的概率偏离其准确概率小于 ε (比如 0.01)
- **方法1:**
  - 使用**Chebyshev**不等式

$$\mathbf{P}\left(\left|\overline{\mu}_N-\mu\right|<\varepsilon\right)>1-\frac{\sigma^2}{N\varepsilon^2} \Rightarrow \delta=\frac{\sigma^2}{N\varepsilon^2} \Rightarrow N=\frac{\sigma^2}{\delta\varepsilon^2}$$

    - 估计方差:
      - 如果无任何信息，则假设均匀分布，即p=0.5，此时方差为0.25（最大可能方差）

$$N=\frac{\sigma^2}{\delta\varepsilon^2}=\frac{0.25}{0.0001\times(0.01)^2}=2.5\times10^7$$

      - 假设已知p<0.2，则最大可能方差为0.16

$$N=1.6\times10^7$$

集中不等式的应用：例2

- 例2: (可靠地估算所需要的样本数量N)
  - 要求以概率**1- δ (**比如**0.9999)** 使得所估计事件**A**发生的概率偏离其准确概率小于 **ε (**比如 **0.01)**
- **方法2**:
  - 使用**Hoeffding**不等式

$$\mathbf{P}\left(\left|\bar{\mu}_N - \mu\right| \ge \varepsilon\right) \le 2\exp\left(-\frac{2\varepsilon^2 N}{\frac{1}{N}\sum_{i=1}^{N}\left(b_i - a_i\right)^2}\right) = 2\exp\left(-\frac{2\varepsilon^2 N}{c^2}\right)$$

其中 $c^2 = \frac{1}{N}\sum_{i=1}^{N}\left(b_i - a_i\right)^2$

➡ $$\delta = 2\exp\left(-\frac{2\varepsilon^2 N}{c^2}\right)$$

➡ $$N = \frac{c^2}{2\varepsilon^2}\log\frac{2}{\delta} = \frac{1}{2\times(0.01)^2}\log\left(\frac{2}{0.0001}\right) = 49,517$$