6.5 Bennet和Bernstein不等式

通过考虑随机变量的方差, 可能推导出更紧地集中不等式, 下面介绍两种基于方差的集中不等式.

定理6.12 (Bennet不等式). *假设$X_1,\ldots,X_n$是n独立同分布的随机变量且满足$X_i-E[X_i]\le 1$. 设随机变量的均值为μ和方差为σ^2, 则有*

$$\Pr\left[\frac{1}{n}\sum_{i=1}^n (X_n-\mu)\ge\epsilon\right]\le\exp(-n\epsilon^2/(2\sigma^2+2\epsilon/3)).$$

Proof. 对任意$t>0$, 根据Chernoff方法有

$$\Pr\left[\frac{1}{n}\sum_{i=1}^n (X_n-\mu)\ge\epsilon\right]\le e^{-nt\epsilon}E\left[\exp\left(\sum_{i=1}^n t(X_i-\mu)\right)\right]=e^{-nt\epsilon}\left(E[e^{t(X_1-\mu)}]\right)^n$$

设$Y=X_1-\mu$, 利用公式$\ln z\le z-1$得到

$$\begin{aligned}\ln E[e^{t(X_1-\mu)}] &= \ln E[e^{tY}]\le E[e^{tY}]-1=t^2E\left[\frac{e^{tY}-tY-1}{t^2Y^2}Y^2\right]\\ &\le t^2E\left[\frac{e^t-t-1}{t^2}Y^2\right]=(e^t-t-1)\sigma^2\end{aligned}$$

这里利用$tY\le t$以及$(e^z-z-1)/z^2$是一个非单调递减的函数. 进一步有

$$e^t-t-1\le\frac{t^2}{2}\sum_{k=0}^\infty (t/3)^k=\frac{t^2}{2(1-t/3)}.$$

因此可得

$$\Pr\left[\frac{1}{n}\sum_{i=1}^n (X_n-\mu)\ge\epsilon\right]\le\exp\left(-nt\epsilon+\frac{nt^2\sigma^2}{2(1-t/3)}\right).$$

猜出$t=\epsilon/(\sigma^2+\epsilon/3)$, 带入完成证明. □

对于Bennet指数不等式, 设

$$\Pr\left[\frac{1}{n}\sum_{i=1}^n (X_n-\mu)\ge\epsilon\right]\le\exp(-n\epsilon^2/(2\sigma^2+2\epsilon/3))=\delta$$

其另外一种表述为: 至少以$1-\delta$的概率有

$$\frac{1}{n}\sum_{i=1}^n X_n\le\mu+\frac{2\ln 1/\delta}{3n}+\sqrt{\frac{2\sigma^2}{n}\ln\frac{1}{\delta}}.$$

当σ^2非常小, 或趋于0时, 得到更紧的收敛率$\bar{X}_n\le\mu+O(1/n)$.

下面考虑另一种基于方差的不等式, 与Bennet不等式不同之处在于约束随机变量的势:

定理6.13 (Bernstein不等式). *假设$X_1,\ldots,X_n$是n独立同分布的随机变量. 设随机变量的均值为μ, 以及方差为σ^2. 如果存在常数$b>0$, 使得对任意$m\ge 2$, 有$E[X_i^m]\le m!b^{m-2}\sigma^2/2$成立, 那么有*

$$\Pr\left[\frac{1}{n}\sum_{i=1}^n (X_n-\mu)\ge\epsilon\right]\le\exp\left(-\frac{n\epsilon^2}{2\sigma^2+2b\epsilon}\right).$$

Proof. 对任意$t > 0$, 根据Chernoff方法有

$$\Pr\left[\frac{1}{n}\sum_{i=1}^{n}(X_n-\mu)\geq\epsilon\right]\leq e^{-nt\epsilon}E\left[\exp\left(\sum_{i=1}^{n}(X_i-\mu)\right)\right]=e^{-nt\epsilon-n\mu t}\left(E[e^{tX_1}]\right)^n$$

利用公式$\ln z\leq z-1$有

$$\ln E[e^{tX_1}]\leq E[e^{tX}]-1=\sum_{m=1}^{\infty}E[X^m]\frac{t^m}{m!}\leq t\mu+\frac{t^2\sigma^2}{2}\sum_{m=2}^{\infty}(bt)^{m-2}=t\mu+\frac{t^2\sigma^2}{2(1-bt)}.$$

由此得到

$$\Pr\left[\frac{1}{n}\sum_{i=1}^{n}(X_n-\mu)\geq\epsilon\right]\leq=\exp\left(-nt\epsilon+\frac{t^2n\sigma^2}{2(1-bt)}\right)$$

取$t=\epsilon/(V+b\epsilon)$完成证明. □

习题. *给出Bernstein不等式的$1-\delta$表述.*

6.6 应用: 随机投影(Random Projection)

本节介绍不等式的一种应用: 随机投影. 假设$\mathbb{R}^d$维空间有n个点: $\bar{x}_1,\bar{x}_2,\cdots,\bar{x}_n$ (d非常大, 例如大约100万). 我们能否找到一个保距变换: $f:\mathbb{R}^d\to\mathbb{R}^k$ $(k\ll d)$, 使得以较大概率有

$$(1-\epsilon)\|\bar{x}_i-\bar{x}_j\|_2^2\leq\|f(\bar{x}_i)-f(\bar{x}_j)\|_2^2\leq(1+\epsilon)\|\bar{x}_i-\bar{x}_j\|_2^2.$$

随机投影广泛应用于高维数据的机器学习问题, 例如最近邻, k-近邻, 降维, 聚类等问题.

随机投影函数可以简单的表示为$f(\bar{x})=\bar{x}P/c$, 其中P为$d\times k$的一个随机矩阵, c为一常数(根据随机矩阵P唯一确定). 下面介绍三种常见的随机矩阵

- $P=(p_{ij})_{d\times k}\in\mathbb{R}^{d\times k}$, $p_{ij}\sim\mathcal{N}(0,1)$, 此时$c=\sqrt{k}$;
- $P=(p_{ij})_{d\times k}\in\{-1,1\}^{d\times k}$, $\Pr(p_{ij}=1)=\Pr(p_{ij}=-1)=1/2$, 此时$c=\sqrt{k}$; 【Rademacher随机变量】
- $P=(p_{ij})_{d\times k}\in\{-1,0,1\}^{d\times k}$, $\Pr(p_{ij}=0)=2/3$, $\Pr(p_{ij}=1)=\Pr(p_{ij}=-1)=1/6$, 此时$c=\sqrt{k/3}$. 【主要用于sparse投影, 减少计算量】

下面我们重点理论分析 Gaussian 随机变量, 其它随机变量可参考相关资料, 对 Gaussian 随机变量, 这里介绍著名的 Johnson–Lindenstrauss 引理, 简称 JL 引理.

引理6.5. *设$\bar{x}_1,\bar{x}_2,\cdots,\bar{x}_n\in\mathbb{R}^d$, 随机矩阵$P=(p_{ij})_{d\times k}\in\mathbb{R}^{d\times k}$, 其元素$p_{ij}\sim\mathcal{N}(0,1)$, 设*

$$\bar{y}_i=f(\bar{x}_i)=\bar{x}_iP/\sqrt{k},\qquad i\in[n]$$

将d维空间中n个点$\bar{x}_1,\bar{x}_2,\cdots,\bar{x}_n$通过随机矩阵$P$投影到$k$维空间$(k\ll d)$. 对任意$\epsilon\in(0,1/2)$, 如果$k\geq 8\log 2n/(\epsilon^2-\epsilon^3)$, 则至少以$1/2$的概率有

$$(1-\epsilon)\|\bar{x}_i-\bar{x}_j\|_2^2\leq\|\bar{y}_i-\bar{y}_j\|_2^2\leq(1+\epsilon)\|\bar{x}_i-\bar{x}_j\|_2^2\qquad(i,j\in[n]).$$

Proof. 下面分三步证明 J-L 引理.

第一步： 对任意非零 $\bar{x} = (x_1, x_2, \cdots, x_d) \in \mathbb{R}^d$, 我们需证明: $E[\|\bar{x}P\|_2^2] = \|\bar{x}\|_2^2$.

设$\bar{y} = \bar{x}P$, 根据$P = (p_{ij})_{d\times k}$ $(p_{ij} \sim \mathcal{N}(0,1))$, 我们有

$$E\left[\|\bar{y}\|_2^2\right] = E\left[\sum_{j=1}^{k}\left(\sum_{i=1}^{d}\frac{x_i p_{ij}}{\sqrt{k}}\right)^2\right] = \sum_{j=1}^{k}\frac{1}{k}E\left[\left(\sum_{i=1}^{d}x_i p_{ij}\right)^2\right] = \sum_{j=1}^{k}\frac{1}{k}\sum_{i=1}^{d}x_i^2 = \frac{1}{k}\sum_{j=1}^{k}\|\bar{x}\|_2^2 = \|\bar{x}\|_2^2.$$

即在期望的情况下, $\bar{x}$与$\bar{y}$具有相等的到原点的距离.

第二步： 证明: $\Pr[\|\bar{y}\|_2^2 \geq (1+\epsilon)\|\bar{x}\|_2^2] \leq \exp(-(\epsilon^2 - \epsilon^3)k/4)$.

为了简洁起见, 我们将矩阵表示为$P = (P^1, P^2, \ldots, P^k)$, 其中$P^i (i \in [d])$是一个$d \times 1$列向量的列向量. 因此我们有

$$\bar{y} = (\bar{x}P^1, \bar{x}P^2, \ldots, \bar{x}P^k)/\sqrt{k}.$$

设

$$v = k\frac{\|\bar{y}\|_2^2}{\|\bar{x}\|_2^2} = \frac{1}{\|\bar{x}\|_2^2}\sum_{j=1}^{k}(\bar{x}P^j)^2 = \sum_{j=1}^{k}\left(\frac{\bar{x}}{\|\bar{x}\|_2}P^j\right)^2 = \sum_{j=1}^{k}v_j^2$$

这里我们设$v_j = \bar{x}P^j/\|\bar{x}\|_2$, 由于$P^j$中每个元素服从$\mathcal{N}(0,1)$, 由高斯分布的性质有$v_j \sim \mathcal{N}(0,1)$. 对任意$t \in (0, 1/2)$, 根据 Chernoff 方法有

$$\Pr[\|\bar{y}\|_2^2 \geq (1+\epsilon)\|\bar{x}\|_2^2] = \Pr[v \geq (1+\epsilon)k] = \Pr[e^{tv} \geq e^{(1+\epsilon)kt}] \leq e^{-(1+\epsilon)kt}E[e^{tv}] = e^{-(1+\epsilon)kt}\left(E[e^{tv_1^2}]\right)^k.$$

进一步有

$$E[e^{tv_1^2}] = \int_{-\infty}^{+\infty}\frac{e^{tu^2}}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}du = \int_{-\infty}^{+\infty}\frac{e^{-\frac{u^2}{2}(1-2t)}}{\sqrt{2\pi}}du = \frac{1}{\sqrt{1-2t}}.$$

于是得到

$$\Pr\left[\|\bar{y}\|_2^2 \geq (1+\epsilon)\|\bar{x}\|_2^2\right] \leq \left(e^{-2(1+\epsilon)t}/(1-2t)\right)^{k/2}.$$

对上式右边t求最小解得$t_{\min} = \frac{\epsilon}{2(1+\epsilon)}$, 代入可得

$$\Pr[\|\bar{y}\|_2^2 \geq (1+\epsilon)\|\bar{x}\|_2^2] \leq [(1+\epsilon)e^{-\epsilon}]^{\frac{k}{2}}.$$

由$\epsilon \in (0, 1/2)$, 我们设$f(\epsilon) = \ln(1+\epsilon)$, 并有

$$f'(\epsilon) = \frac{1}{1+\epsilon}, f''(\epsilon) = -\frac{1}{(1+\epsilon)^2}, f'''(\epsilon) = \frac{2}{(1+\epsilon)^3} \leq 2.$$

根据泰勒中值定理有

$$f(\epsilon) = f(0) + f'(0)\epsilon + \frac{f''(0)\epsilon^2}{2!} + \frac{f'''(\xi)\epsilon^3}{3!} = \epsilon - \frac{\epsilon^2}{2} + \frac{\epsilon^2}{3} \leq -\frac{\epsilon^2 - \epsilon^3}{2}.$$

于是得到

$$\Pr[\|\bar{y}\|_2^2 \geq (1+\epsilon)\|\bar{x}\|_2^2] \leq e^{-k(\epsilon^2-\epsilon^3)/4}.$$

同理可证明$\Pr[\|\bar{y}\|_2^2 \leq (1-\epsilon)\|\bar{x}\|_2^2] \leq e^{-2k(\epsilon^2-\epsilon^3)/4}$.

第三步： 对于任意给定的$i \neq j$, 由第二步可知

$$\Pr[\|y_i - y_j\|_2^2 \geq (1+\epsilon)\|x_i - x_j\|_2^2] \leq e^{-k(\epsilon^2-\epsilon^3)/4}, \quad \Pr[\|y_i - y_j\|_2^2 \leq (1-\epsilon)\|x_i - x_j\|_2^2] \leq e^{-k(\epsilon^2-\epsilon^3)/4}.$$

由于$i \in [n], j \in [n]$, 因此共有$n(n-1)$对(i,j), 根据 Union 不等式有,

$$\Pr\left[\exists (i,j)\colon \|y_i - y_j\|_2^2 \geq (1+\epsilon)\|x_i - x_j\|_2^2 \quad \text{或} \quad \|y_i - y_j\|_2^2 \leq (1-\epsilon)\|x_i - x_j\|_2^2\right] \leq 2n^2 e^{-k(\epsilon^2-\epsilon^3)/4} \leq \frac{1}{2}.$$

根据$2n^2 e^{-k(\epsilon^2-\epsilon^3)/4} \leq 1/2$, 求解出$k \geq 8\log 2n/(\epsilon^2 - \epsilon^3)$. 引理得证. □