Machine Learning Theory

FENG JIANTING

ContentsI High Dimensional Statistics

Part I. High Dimensional Statistics

1 Concentration Inequalities$\mathbb{P}X \geq t$

$$\mathbb{P}[X \geq t] \leq \frac{\mathbb{E}X}{t}, t > 0.$$

$\mathbb{E}X$

$t > 0$

$$\begin{aligned}\mathbb{P}[X \geq t] &= \int_{x \geq t} dP \\ &\leq \int_{x \geq t} \frac{x}{t} dP \\ &\leq \int_{\mathbb{R}} \frac{x}{t} dP \\ &= \frac{\mathbb{E}X}{t}\end{aligned}$$

$$\mathbb{P}[|X - \mu| \geq t] \leq \frac{X}{t^2}, t > 0.$$

$\mu = \mathbb{E}X$

$k \in \mathbb{N}_+$

$$\mathbb{P}[|X - \mu| \geq t] \leq \frac{\mathbb{E}\left[|X - \mu|^k\right]}{t^k}$$

$\mathbb{E}\left[|X-\mu|^k\right]$
$|X-\mu|^k$
$\varphi(\lambda)=\mathbb{E}\left[e^{\lambda(X-\mu)}\right]\lambda \leq |b| b>0 Y=(\lambda(X-\mu))$

$$\mathbb{P}\left[(X-\mu)\geq t\right]\mathbb{P}\left[e^{\lambda(X-\mu)}\geq e^{\lambda t}\right]\leq \frac{\mathbb{E}\left[e^{\lambda(X-\mu)}\right]}{e^{\lambda t}}$$

$$\mathbb{P}\left[(X-\mu)\geq t\right]\leq_{\lambda\in[0,b]}\left\{\mathbb{E}\left[e^{\lambda(X-\mu)}\right]-\lambda t\right\}$$

1.1 Sub-Gaussian variables and Hoeffding bounds

$X\sim N(\mu,\sigma^2)\mu\sigma^2$

$$\mathbb{E}\left[e^{\lambda(X-\mu)}\right]=e^{\frac{\lambda^2\sigma^2}{2}}$$

$\lambda\in\mathbb{R}$

$$_{\lambda\geq 0}\left\{\mathbb{E}\left[e^{\lambda(X-\mu)}\right]-\lambda t\right\}=_{\lambda\geq 0}\left\{\frac{\sigma^2\lambda^2}{2}-\lambda t\right\}=-\frac{t^2}{2\sigma^2}$$

$$\mathbb{P}\left[X-\mu\geq t\right]\leq e^{-\frac{t^2}{2\sigma^2}}$$

$X\mu=\mathbb{E}X\sigma>0$

$$\mathbb{E}\left[e^{\lambda(X-\mu)}\right]\leq e^{\sigma^2\lambda^2/2}$$

$\lambda\in\mathbb{R}$

σ

$$\mathbb{P}\left[|X-\mu|\geq t\right]\leq 2e^{-\frac{t^2}{2\sigma^2}}$$

$t\in\mathbb{R}$

$\varepsilon\{-1,1\}\sigma=1$

$X[a,b]X'X\lambda\in\mathbb{R}$

$$\mathbb{E}_X\left[e^{\lambda X}\right]=\mathbb{E}_X\left[e^{\lambda(X-\mathbb{E}_{X'}[X'])}\right]\leq\mathbb{E}_{X,X'}\left[e^{\lambda(X-X')}\right]$$

$e^{-\lambda x}\varepsilon\varepsilon(X-X')X-X'$

$$\mathbb{E}_{X,X'}\left[e^{\lambda(X-X')}\right]=\mathbb{E}_{X,X'}\left[\mathbb{E}_\varepsilon\left[e^{\lambda\varepsilon(X-X')}\right]\right]\leq\mathbb{E}_{X,X'}\left[e^{\frac{\lambda^2(X-X')^2}{2}}\right]$$

$|X - X'| \leq b - a$

$$\mathbb{E}_{X,X'}\left[e^{\frac{\lambda^2(X-X')^2}{2}}\right] \leq e^{\frac{\lambda^2(b-a)^2}{2}}$$

$\sigma = \frac{b-a}{2}$

$X_i, i = 1, \cdots, n X_i \mu_i \sigma_i t \geq 0$

$$\mathbb{P}\left[\sum_{i=1}^{n}(X_i - \mu_i) \geq t\right] \leq \left\{-\frac{t^2}{2\sum_{i=1}^{n}\sigma_i^2}\right\}$$

$$\begin{aligned}\mathbb{P}\left[\sum_{i=1}^{n}(X_i - \mu_i) \geq t\right] &\leq \left\{\frac{\lambda^2}{2}\sum_{i=1}^{n}\sigma_i^2 - \lambda t\right\} \\ &\leq \left\{-\frac{t^2}{2\sum_{i=1}^{n}\sigma_i^2}\right\}\end{aligned}$$

$\sigma \geq 0$

$$\mathbb{E}\left[e^{\lambda X}\right] \leq e^{\frac{\lambda^2\sigma^2}{2}} \lambda \in \mathbb{R}$$

$c \geq 0 Z \sim N(0, \tau)$

$$\mathbb{P}\left[|X| \geq s\right] \leq c\mathbb{P}\left[|Z| \geq s\right], s \geq 0$$

$\theta \geq 0$

$$\mathbb{E}\left[X^{2k}\right] \leq \frac{(2k)!}{2^k k!}\theta^{2k} k = 1, 2, \cdots$$

$\sigma \geq 0$

$$\mathbb{E}\left[e^{\frac{\lambda X^2}{2\sigma^2}}\right] \leq \frac{1}{\sqrt{1-\lambda}} \lambda \in [0, 1)$$

1.2 Sub-exponential variables and Bernstein bounds

$X \mu = \mathbb{E} X (v, \alpha)$

$$\mathbb{E}\left[e^{\lambda(X-\mu)}\right] \leq e^{\frac{v^2\lambda^2}{2}} |\lambda| < \frac{1}{\alpha}.$$

$(v, \alpha) = (\sigma, 0)$

$Z \sim N(0,1) X = Z^{@} \lambda < \frac{1}{2}$

$$\mathbb{E}\left[e^{\lambda(X-1)}\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\lambda(z^2-1)e^{-z^2/2}} dz$$
$$= \frac{e^{-\lambda}}{\sqrt{1-2\lambda}}$$

$\lambda > \frac{1}{2} |\lambda| < \frac{1}{4}$

$$\frac{e^{-\lambda}}{\sqrt{1-2\lambda}} \le e^{2\lambda^2} = e^{4\lambda^2/2}$$

$Z(v, \alpha) = (4, 2)$

$X(v, \alpha)$

$$\mathbb{P}[X - \mu \ge t] \le \begin{cases} e^{-\frac{t^2}{2v^2}} 0 \le t \le \frac{v^2}{\alpha}, \\ e^{-\frac{t}{2\alpha}} \ t > \frac{v^2}{\alpha}. \end{cases}$$

$$\mathbb{P}[X - \mu \ge t] \le \left(\mathbb{E}[e^{\lambda X}] - \lambda t \right) \le \left(\frac{v^2 \lambda^2}{2} - \lambda t \right)$$

$\lambda < \frac{1}{\alpha} g(\lambda, t) = \frac{v^2\lambda^2}{2} - \lambda t \lambda^* = \frac{t}{v^2}$

$\frac{t}{v^2} < \frac{1}{\alpha}{}_{\lambda \in [0, \alpha^{-1})} g(\lambda, t) = -\frac{t^2}{2v^2}$

$t \ge \frac{v^2}{\alpha} \lambda^* = \alpha^{-1}$

$$_{\lambda \in [0, \alpha^{-1})} g(\lambda, t) = -\frac{t}{\alpha^2} + \frac{v^2}{2\alpha^2} \le -\frac{t}{2\alpha}$$

$\frac{v^2}{\alpha} \le t$

$X^2 \mu = \mathbb{E} X \sigma^2 = Xb$

$$|\mathbb{E}\left[(X-\mu)^k\right]| \le \frac{1}{2} k! \sigma^2 b^{k-2} k \ge 2.$$

$$\mathbb{E}\left[e^{\lambda(X-\mu)}\right] = 1 + \frac{\lambda^2\sigma^2}{2} + \sum_{k=3}^{\infty} \lambda^k \frac{\mathbb{E}\left[(X-\mu)^k\right]}{k!}$$
$$\le 1 + \frac{\lambda^2\sigma^2}{2} + \frac{\lambda^2\sigma^2}{2} \sum_{k=3}^{\infty} (|\lambda| b)^{k-2},$$

$$\mathbb{E}\left[e^{\lambda(X-\mu)}\right] \le 1 + \frac{\lambda^2\sigma^2/2}{1 - b|\lambda|} \le e^{\frac{\lambda^2\sigma^2/2}{1-b|\lambda|}}$$

$$\mathbb{E}\left[e^{\lambda(X-\mu)}\right] \le e^{\frac{\lambda^2(\sqrt{2}\sigma)^2}{2}} |\lambda| < \frac{1}{2b}.$$

$(v, \alpha) = (\sqrt{2}\sigma, 2b)$

$$\mathbb{E}\left[e^{\lambda(X-\mu)}\right] \le e^{\frac{\lambda^2\sigma^2/2}{1-b|\lambda|}}$$

$|\lambda| < \frac{1}{b}$

$$\mathbb{P}\left[|X-\mu| \ge t\right] \le 2e^{\frac{t^2}{2(\sigma^2+bt)}} t \ge 0$$

$\lambda = \frac{t}{bt+\sigma^2} \in [0, b^{-1})$
$\{X_k\}_{k=1}^n$

$$\alpha_* = {}_{k\in[n]} \alpha_k v_* = \sqrt{\sum_{k=1}^n v_k^2}.$$

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^n (X_k - \mu_k) \ge t\right] \le \begin{cases} e^{-\frac{nt^2}{2(v_*^2/n)}}, 0 \le t \le \frac{v_*^2}{n\alpha_*}, \\ e^{-\frac{nt}{2\alpha_*}}, \quad t > \frac{v^2}{n\alpha_*} \end{cases}$$

$\chi^2 nY \sim \chi_n^2$

$$Y = \sum_{k=1}^n Z_k^2, Z_k \sim N(0,1)$$

$Z_k(2,4)Y(2\sqrt{n},4)$

$$\mathbb{P}\left[\left|\frac{1}{n}\sum_{k=1}^n Z_k^2 - 1\right| \ge 1\right] \le 2e^{-nt^2/8}, t \in (0,1)$$

χ^2

$N \ge 2\{u^1, \cdots, u^N\} \subset \mathbb{R}^d dF : \mathbb{R}^d \to \mathbb{R}^m m \ll d$

$$1 - \delta \le \frac{\|F(u^i) - F(u^j)\|_2^2}{\|u^i - u^j\|_2^2} \le 1 + \delta, u^i \ne u^j.$$

$X \in \mathbb{R}^{m\times d} N(0,1)$

$$\begin{aligned} F : \mathbb{R}^d &\to \mathbb{R}^m \\ u &\mapsto Xu/\sqrt{m} \end{aligned}$$

$Fx_i \in \mathbb{R}^d Xu \in \mathbb{R}^d \langle x_i, \frac{u}{\|u\|_2} \rangle \sim N(0,1)$

$$Y = \frac{\|Xu\|_2^2}{\|u\|_2^2} = \sum_{i=1}^{m} \langle x_i, u/\|u\| \rangle^2,$$

$\chi_m^2 m \chi^2$

$$\mathbb{P}\left[\left| \frac{\|Xu\|_2^2}{m\|u\|_2^2} - 1 \right| \geq \delta \right] \leq 2e^{-m\delta^2/8} \delta \in (0,1).$$

F

$$\mathbb{P}\left[\frac{\|F(u)\|_2^2}{|u\|_2^2} \notin [1-\delta, 1+\delta] \right] \leq 2e^{-m\delta^2/8} 0 \neq u \in \mathbb{R}^d.$$

$\binom{N}{2} u$

$$\mathbb{P}\left[\frac{\|F(u^i - u^j)\|_2^2}{\|u^i - u^j\|_2^2} \notin [1-\delta, 1+\delta] u^i \neq u^j \right] \leq 2\binom{N}{2} e^{-m\delta^2/8}$$

$\epsilon \in (0,1) \epsilon m \geq \frac{c}{\delta^2}$ N

X

ν, α

$$\mathbb{E}[e^{\lambda X}] \leq e^{\frac{\nu^2\lambda^2}{2}} |\lambda| < \frac{1}{\alpha}$$

$c_0 \mathbb{E}[e^{\lambda X}] < \infty \lambda < c_0$

$c_1, c_2 > 0$

$$\mathbb{P}^{|X| \geq t} \leq c_1 e^{-c_2 t} t > 0.$$

$\gamma = {}_{k \geq 2} \left[\frac{\mathbb{E}[X^k]}{k!} \right]^{1/k}$

$|X| \leq b$

1.3 Martingale-based methods $\{X_k\}_{k=1}^n f(X) = f(X_1, \cdots, X_n) f : \mathbb{R}^n \to \mathbb{R}$

$$Y_k = \mathbb{E}[f(X)|X_1, \cdots, X_k] k = 1, \cdots, n-1$$

$Y_0 = \mathbb{E}[f(X)] Y_n = f(X)$

$$f(X) - \mathbb{E}[f(X)] = Y_n - Y_0 = \sum_{k=1}^{n} \underbrace{Y_k - Y_{k-1}}_{D_k}$$

$\{Y_k\}_{k=1}^\infty \{F_k\}_{k=1}^\infty \{(Y_k, F_k)\}_{k=1}^\infty k \geq 1$

$$\mathbb{E}[|Y_k|] < \infty \mathbb{E}[Y_{k+1}|F_k] = Y_k.$$

$\{(D_k, F_k)\}_{k=1}^{\infty} k \geq 1$

$$\mathbb{E}[|D_k|] < \infty \mathbb{E}[D_{k+1}|F_k] = 0.$$

$\{(D_k, F_k)\}_{k=1}^{\infty} \mathbb{E}[e^{\lambda D_k}|F_{k-1}] \leq e^{\lambda^2 v_k^2/2} |\lambda| < \alpha_k^{-1}$

$\sum_{k=1}^{n} D_k (\sqrt{\sum_{k=1}^{n} v_k^2}, \alpha_*) \alpha_* = {}_{k=1,\cdots,n} \alpha_k$

$$\mathbb{P}\left[\left|\sum_{k=1}^{n} D_k\right| \geq t\right] \leq \begin{cases} 2e^{-\frac{t^2}{2\sum_{k=1}^{n}} v_k^2} 0 \leq t \leq \frac{\sum_{k=1}^{n} v_k^2}{\alpha_*} \\ 2e^{-\frac{t}{2\alpha_*}} \quad t > \frac{\sum_{k=1}^{n} v_k^2}{\alpha_*} \end{cases}$$

References