Machine Learning Exercise 2

何舜成

2012011515

1. Weaken Setting 1 (n binary independent and identically distributed random variables) and obtain the same concentration inequality

Modified setting: n independent binary random variables X_1, X_2,..., X_n, for each X_i, $E[X_i] = p_i$.

Proof:

$$Var[X_i] = E[X^2] - (E[x])^2 = p_i(1 - p_i)$$

n random variables are independent, therefore

$$Var[\bar{X}] = Var[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n^2}\sum_{i=1}^{n} Var[X_i] = \frac{1}{n^2}\sum_{i=1}^{n} p_i(1 - p_i)$$

Since

$$p_i(1 - p_i) \leq \frac{1}{4}, \forall i \in [n]$$

Apply Chebyshev ineq.

$$Pr[|\bar{X} - E[\bar{X}]| \geq \delta] \leq \frac{Var[\bar{X}]}{\delta^2} = \frac{\sum_{i=1}^{n} p_i(1 - p_i)}{n^2\delta^2} \leq \frac{1}{4n\delta^2} = O(n^{-1})$$

We get the same result from original setting.

2. Prove $D_e(p + \delta||p) \geq 2\delta^2, \forall\delta > 0, p \in [0, 1 - \delta]$.

Proof:

By definition of relative entropy

$$D_e(p+\delta||p) = (p+\delta)ln\frac{p+\delta}{p} + [1-(p+\delta)]ln\frac{1-(p+\delta)}{1-p}$$

Let $f(\delta) = D_e(p+\delta||p) - 2\delta^2$ and we get

$$\begin{aligned}
f(\delta) &= (p+\delta)ln\frac{p+\delta}{p} + [1-(p+\delta)]ln\frac{1-(p+\delta)}{1-p} - 2\delta^2 \\
f'(\delta) &= ln\frac{p+\delta}{p} - ln\frac{1-(p+\delta)}{1-p} - 4\delta \\
f''(\delta) &= \frac{1}{p+\delta} + \frac{1}{1-(p+\delta)} - 4
\end{aligned}$$

Since

$$f''(\delta) = \frac{1}{(p+\delta)[1-(p+\delta)]} - 4 \geq \frac{1}{1/4} - 4 \geq 0$$

And $f'(0) = 0$, therefore

$$f'(\delta) > 0, \forall \delta > 0$$

Since $f(0) = 0$, we get

$$f(\delta) > 0, \forall \delta > 0$$

Therefore we proved the inequality, and the ineq. is tight when $\delta \to 0$.

3. Prove Lemma 2. n **independent random variables** X_1**,** X_2**,...,** X_n**, where** X_i **takes its value from** $[a_i, b_i]$**.**

$$E[e^{t(X_1-E[X_1])}] \leq exp(\frac{t^2(b_1-a_1)^2}{8}), \forall t > 0$$

Proof:

Function e^{tx} is convex, therefore

$$e^{t(X_1-E[X_1])} \leq \frac{b_1-X_1}{b_1-a_1}e^{t(a_1-E[X_1])} + \frac{X_1-a_1}{b_1-a_1}e^{t(b_1-E[X_1])}$$

$$E[e^{t(X_1-E[X_1])}] \leq \frac{b_1-E[X_1]}{b_1-a_1}e^{t(a_1-E[X_1])} + \frac{E[X_1]-a_1}{b_1-a_1}e^{t(b_1-E[X_1])}$$

Substitute $E[X_1]$ with $c, c \in [a_1, b_1]$, and let

$$f(t) = ln(\frac{b-c}{b-a}e^{t(a-c)} + \frac{c-a}{b-a}e^{t(b-c)})$$

We get

$$\begin{aligned} f'(t) &= (b-c)(a-c)\frac{e^{t(a-c)} - e^{t(b-c)}}{(b-c)e^{t(a-c)} + (c-a)e^{t(b-c)}} \\ f''(t) &= (b-c)(c-a)(a-b)^2\frac{e^{t(a-c)}e^{t(b-c)}}{((b-c)e^{t(a-c)} + (c-a)e^{t(b-c)})^2} \end{aligned}$$

Since

$$\begin{aligned} f''(t) &\leq (b-c)(c-a)(a-b)^2\frac{e^{t(a-c)}e^{t(b-c)}}{4(b-c)e^{t(a-c)}(c-a)e^{t(b-c)}} \\ &= \frac{(b-a)^2}{4} \end{aligned}$$

Expand $f(t)$ to second order

$$\begin{aligned} f(t) &= f(0) + f'(0)t + \frac{f''(\xi)}{2}t^2 \\ &= 0 + 0 + \frac{f''(\xi)}{2}t^2 \\ &\leq \frac{t^2(b-a)^2}{8} \end{aligned}$$

Therefore

$$E[e^{t(X_1 - E[X_1])}] \leq e^{f(t)} \leq exp(\frac{t^2(b-a)^2}{8})$$