Generalization: performance on test data

L2: Basic Inequalities

1. Markov's inequality

$$Thm\ 1\ r.v.\ \ x > 0, Ex\ exists, \forall k > 0\ P(X \geq k) \leq Ex/x$$

2. Chebychev's inequality

$$Thm\ 2\ r, v, x, Ex, Var(x) = \sigma^2, \forall k > 0, P(|x - Ex| \geq k) \leq \sigma^2/k^2$$

3. Ex

$$r, v, x \sim N(0,1)\ Define\ function\ \Phi(u) = P(x \geq u)$$

Find elementary function f,g s.t $g(u) \leq \Phi(u) \leq f(u)$

$$f(u) = \frac{1}{u} e^{-\frac{u^2}{2}}, g(u) = \frac{1}{u + e^{-u}} e^{-\frac{u^2}{2}}$$

4. Chernoff Inequality

$$r.v\ x > 0\ Ex, Ex^2, Ex^3 \ldots Known, then\ \forall k > 0, P(x \geq k) \leq \min_{i \geq 1} Ex^i/k^i$$

Moment generating function: $M_x(t) = E[e^{tx}] = 1 + tEx + \frac{t^2}{2} Ex^2 + \cdots$

Then $P(x \geq k) \leq \inf_{t>0} e^{-tk} E[e^{tx}]$ Which one is better? (Not exercise)

5. Concentration Inequality

$i.i.d\ r.v. x_1, x_2 \ldots$ $\mathrm{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} x_i - Ex\right| \geq \varepsilon\right) = ?$

Central limit theorem (CLT)

6. Entropy, Relative entropy (kullback-ltibler Divergence)

$$Def\ r.v.\ \ x\ \ P = (p\_1 \ldots p\_n)\ Then\ H(x) = \sum_{i=1}^{n} p_i \log(1/p_i)$$

$Def\ (Related\ entropy, kl\ divergence):$

$$P = (p_1 \ldots p_n)\ Q = (q_1,,, q_n)\ Then\ D(P||Q) = \sum_{i=1}^{n} p_i \log\left(\frac{p_i}{q_i}\right), not\ symmetric$$

$log \sim bits\ ln \sim nats$

6. Chernoff Bound

1) $r.v.\ \ x, x_1 \ldots x_n, Bernoull, Ex = p, Then\ \ \forall \delta > 0, P\left(\frac{1}{n}\sum_{i=1}^{n} x_i - p \geq \delta\right) = ?\ (ex)$

$$\hat{x} = \sum_{i=1}^{n} x_i\,, then\ M_{\hat{x}}(t) = (1-p+pe^t)^n, Then\ P\big(\hat{x} \geq n(p+\delta)\big)$$

$$\leq \inf_{t>0} e^{-tn(p+\delta)}(1-p+pe^t)^n$$

$$And\ that\ is\ to\ minimize\ -t(p+\delta) + ln(1-p+pe^t), so \frac{pe^t}{1-p+pe^t} = p+\delta$$

$$Then\ e^t = \frac{(1-p)(p+\delta)}{p(1-p-\delta)}, and\ the\ value\ is \exp\{-t(p+\delta) + \ln{(1-p)/(1-p-\delta)}\}$$

$(ans) \exp\{-nD_B^{(e)}(p+\delta||p)\}$

$And\ if\ x\ is\ distributed\ on\ [0,1] and\ Ex = p\ while\ x\ is\ not\ bernoull,$

$then\ by\ Jessan's\ inequality\ we\ have\ E(e\text{^}tx) < \cdots\ Then\ we\ know\ it\ is\ better.$

$And\ if\ x\ is\ not\ identical, just\ independent\ the\ result\ is\ the\ same.$

Additive Chernoff bound:

$$\exp\{-nD_B^{(e)}(p+\delta||p)\} \leq e^{-2n\delta^2}$$

(ex)

7. Hoeffding's Inequality

$$x_1 \ldots x_n\ independent\ x_i \in [a_i, b_i], Ex_i = \mu_i$$

$$P\left(\frac{1}{n}\sum_{i=1}^{n} x_i - \frac{1}{n}\sum_{i=1}^{n} \mu_i \geq \varepsilon\right) \leq e^{-2n^2\varepsilon^2}/\sum_{i=1}^{n}(a_i - b_i)\text{^}2$$