HW 3

1. Tail Bounds

1.1. Known: bias $\varepsilon = 0.002$ failure rate $\delta = 5\%$

① According to Chebyshev. Inequality

$$\delta \le \frac{\sigma^2}{m\varepsilon^2} \Rightarrow m \le \frac{\sigma^2}{\delta\varepsilon^2}$$

$$\sigma^2 \le 0.07 \times (1-0.07) = 0.0651$$

$$\Rightarrow m \le \frac{0.0651}{0.05 \times 0.002^2} = 325500$$

for 41 colors: $41m = 13.3$ million.

② According to Hoeffding Inequality

$$\delta \le 2\exp\left[-\frac{2m\varepsilon^2}{c^2}\right]$$ where $c=1$ for bernoulli variable.

$$m \le \frac{c^2}{2\varepsilon^2} \lg\frac{2}{\delta} = 461109$$

for 41 colors: $41m = 18.9$ million

③ CLT. $Z(1-\delta) = 1.96$

$$\frac{Z(1-\delta)\sigma}{\sqrt{m}} \ge \varepsilon$$

$$m \le \frac{Z^2(1-\delta)\sigma^2}{\varepsilon^2} = 62532$$

for 41 colors: $41m = 2.56$ million

Assuming Google tracking 10 million for testing per day

According Chebshev. 31.92 hr.

Hoeffding: 45.36 hr

CLT: 6.14 hr

④ According to chebshev,

$$\varepsilon \le \sqrt{\frac{\sigma^2}{m\delta}}$$

So we can evenly test each color

And if $\mu_i + \varepsilon < \mu_{max} - \varepsilon$

then color i can be removed from the test. with probility $1-\delta$

1.2.

Estimate bound better than bound: happens when click rate | color i $\neq$ click rate | color j.

Estimate bound worse than bound happens when samples are not evenly distributed

1.3. $f_n(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \alpha_i \phi_i(x)$

$$E(f_n(x)) = E\left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \alpha_i \phi_i(x)\right]$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} E[\alpha_i \phi_i(x)]$$

Since $E(\alpha_i) = 0$ and Slutsky theorem

$$= 0$$

$$Var(f_n(x)) = Var\left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \alpha_i \phi_i(x)\right]$$

$$= \frac{1}{n} \sum_{i=1}^{n} Var(\alpha_i \phi_i(x))$$

Since $\alpha_i \in U[-1, 1)$

$$Var(\alpha_i) = \int_{-1}^{1} (x - \mu^0)^2 p(x) dx = \frac{1}{3}$$

$$= \frac{1}{3} \sum_{i=1}^{n} \phi_i^2(x)$$

According to CLT

$$f_n(x_i) \sim \mathcal{N}\left(0, \frac{1}{3} \sum_{i=1}^{n} \phi_i^2(x_i)\right)$$

Covariance

$$E[f_n(x_i) f_n(x_j)]$$

$$= E\left[\frac{1}{n} \sum_{k=1}^{n} \alpha_k \phi_k(x_i) \sum_{l=1}^{n} \alpha_l \phi_l(x_j)\right]$$

$$= \frac{1}{n} \sum_{k=1}^{n} \sum_{l=1}^{n} E[\alpha_k \alpha_l \phi_k(x_i) \phi_l(x_j)]$$

for $k \neq l$ $\quad \alpha_k \perp\!\!\!\perp \alpha_l$ $\quad E(\alpha_k \alpha_l) = E(\alpha_k) E(\alpha_l) = 0$

$$= \frac{1}{n} \sum_{k=1}^{n} E[\alpha_k^2 \phi_k(x_i) \phi_k(x_j)]$$

where $E[\alpha_k^2] = \frac{1}{3}$

$$= \frac{1}{3n} \sum_{k=1}^{n} \phi_k(x_i) \phi_k(x_j)$$

Given $|\phi(x)| \leq C$

$$\leq \frac{C^2}{3}$$

2.1 When data is highly noisy, boxcar kernel is preferred because knn can easily pick up the wrong label depending on the size of k. Otherwise, knn might be preferred.

2.2.1 Bayes error comes from noise of the data, which is not preventable

2.2.2 $E[(y-\hat{f}(x))^2]$

$= E[(f(x)+\epsilon-\hat{f}(x))^2]$

$= E(\epsilon^2) + E[(f(x)-\hat{f}(x))^2]$

$+ 2E[\epsilon(f(x)-\hat{f}(x))]$

where $E(\epsilon^2) = Var(\epsilon) + \cancel{E(\epsilon)}^2$ $\quad 0$

$= \sigma^2$

Also $2E[\epsilon(f(x)-\hat{f}(x))]$

$= 2\cancel{E(\epsilon)}\,E(f(x)-\hat{f}(x)) = 0$ $\quad 0$

So. $E[(y-\hat{f}(x))^2]$

$= \sigma^2 + E[(f(x)-\hat{f}(x))^2]$

where $E[(f(x)-\hat{f}(x))^2]$

$= E[(f(x)-E(\hat{f}(x))+E(\hat{f}(x))-\hat{f}(x))^2]$

$= E[(f(x)-E(\hat{f}(x)))^2] + E[(E(\hat{f}(x))-\hat{f}(x))^2]$

$+E[(f(x)-E[\hat{f}(x)])(E[\hat{f}(x)]-\hat{f}(x))]$

where $E[(f(x) - E[\hat{f}(x)])(E[\hat{f}(x)] - \hat{f}(x))]$

$$= E[f(x)E(\hat{f}(x)) - E(\hat{f}(x))^2 - f(x)\hat{f}(x) + E(\hat{f}(x))\hat{f}(x)]$$

$$= \cancel{f(x)E[\hat{f}(x)]} - \cancel{E[\hat{f}(x)]^2} - \cancel{f(x)E[\hat{f}(x)]} + \cancel{E[\hat{f}(x)]^2}$$

$$= 0$$

and let $z = f(x) - E(\hat{f}(x))$

$$Var(z) = E[(z - E(z))^2]$$

where $E(z) = f(x) - E(\hat{f}(x)) = z$

$\therefore Var(z) = 0$

$\Rightarrow E(z^2) = E(z)^2$

$$\Rightarrow E[(f(x) - E(\hat{f}(x)))^2]$$

$$= E[f(x) - E(\hat{f}(x))]^2$$

Therefore $E[(y - \hat{f}(x))^2]$

$$= \underbrace{\sigma^2}_{\text{Inevitable error}} + \underbrace{E[f(x) - E[\hat{f}(x)]]^2}_{\text{bias}^2} + \underbrace{E[(\hat{f}(x) - E[\hat{f}(x)])^2]}_{\text{Variance}}$$

(Bayes Error)

2.2.3 when k or h increases, bias grows and variance decreases. Vice versa.